EXPERT NEWS

Contact: News, *IEEE Expert*, 10662 Los Vaqueros Circle, PO Box 3014, Los Alamitos, CA 90720-1264

Dying by design

Dale Strok, associate editor

What can we learn by "killing" a neural net? Can we extrapolate our findings to the larger class of parallel distributed processors, including the human brain? Physicist Stephen Thaler says yes, but admits his ideas are controversial. His theories have ignited arguments between physicists and biologists, between scientists and theologians, and even between close friends. In the process, he has attracted a wide range of public forums, from *Scientific American* to British radio to the World Congress on Neural Nets '93.

Thaler's daytime job at McDonnell Douglas involves using neural networks in a novel process: to grow diamond films. This has saved years of "poking around large parameter spaces," according to Thaler, and his project has been very successful.

In the evenings, Thaler does something completely different with neural nets: He destroys them. Always curious about death, Thaler has been seeking deterministic explanations of what happens when consciousness is destroyed. Other researchers have looked at the graceful degradation of neural networks during weight pruning, Thaler has taken the process all the way to the end, as network output becomes chaotic. His results have led to conclusions with intriguing ramifications for such diverse topics as national defense, dreaming and sensory deprivation, and near-death human experience.

The experiment

Thaler devised a program that gradually destroys a network by randomly severing links between units, similar to nulling nerve impulses at the brain's synapses. Recently, he has been using a 3-5-9 pattern associator, a simple feed-forward network that maps eight 3-bit input patterns to eight 3×3 pixel patterns (see the figure).[1] Thaler trained the network through backpropagation and gradient descent, and ran 50 experiments, each involving 100 deaths.

Throughout the neural destruction process, the network continues to output virtual experiences — exact duplicates of stored impressions and memories.

Hallucination. As connections are gradually severed, the network produces a stream of nonsense and "hallucinatory" output. A *hallucination* represents a false indication that a training vector of data has been applied to the net's inputs. The network might be said to be fantasizing, reliving its experiences and churning out reactions that seem to make sense, even though its circuits are dissolving. Throughout the neural destruction process, the network continues to output these virtual experiences—exact duplicates of stored impressions and memories.

In the case of the 3-5-9 pattern associator's death in which connection weights were randomly pruned, hallucinations constituted nearly 60 percent of the output stream. These hallucinatory outputs occurred 37 times more often than would be seen if the output had been randomly set to 1's and 0's. This cannot be coincidental, according to Thaler, especially since so few if any hallucinations are observed when untrained networks are destroyed.

Whimsy. Periodically, novel combinations of stored memories also appear, yielding what might be regarded as "new" experience. Thaler calls these novel, combinatorial outputs *whimsical* states. They have never been trained into the net, nor would they ever appear in a healthy net. For example, if a 4-2-4 encoder has four trained input and output vectors, (1,0,0,0), (0,1,0,0), (0,0,1,0), and (0,0,0,1), whimsical output might be (0,1,1,0).[2] Thaler found that these whimsical states predominate during the last stages of network death. One of his experiments was not only revealing, but eerie: He once taught a network to output numerous Christmas phases, including "peace on earth" and "good will to all men." First the system was able to output these familiar phases, but as it degraded, it synthesized a completely new phrase, "all men to good earth," from its stored experience.

An explanation

Thaler offers two perspectives on the phenomenon of hallucination. In the first, he points out that neural nets (including human brains) organize their elements into groups, or colonies, that respond specifically to different environmental features. Destroying neural connections tends to isolate these colonies into nearly stranded islands (Thaler calls them "fragmens"), but noise generated in the surrounding network can continue to activate these environmental features.

The second point of view he offers is

that patterns of zeroed connections produced in neural death are interpreted by the still intact portions of the network as some environmental feature: For instance, when you see the incomplete word "hme," you know what it says. This is the process known in neural-network theory as *completion*, whereby the brain recognizes partially obscured patterns by taking degraded information and then filling in the blanks, so to speak.

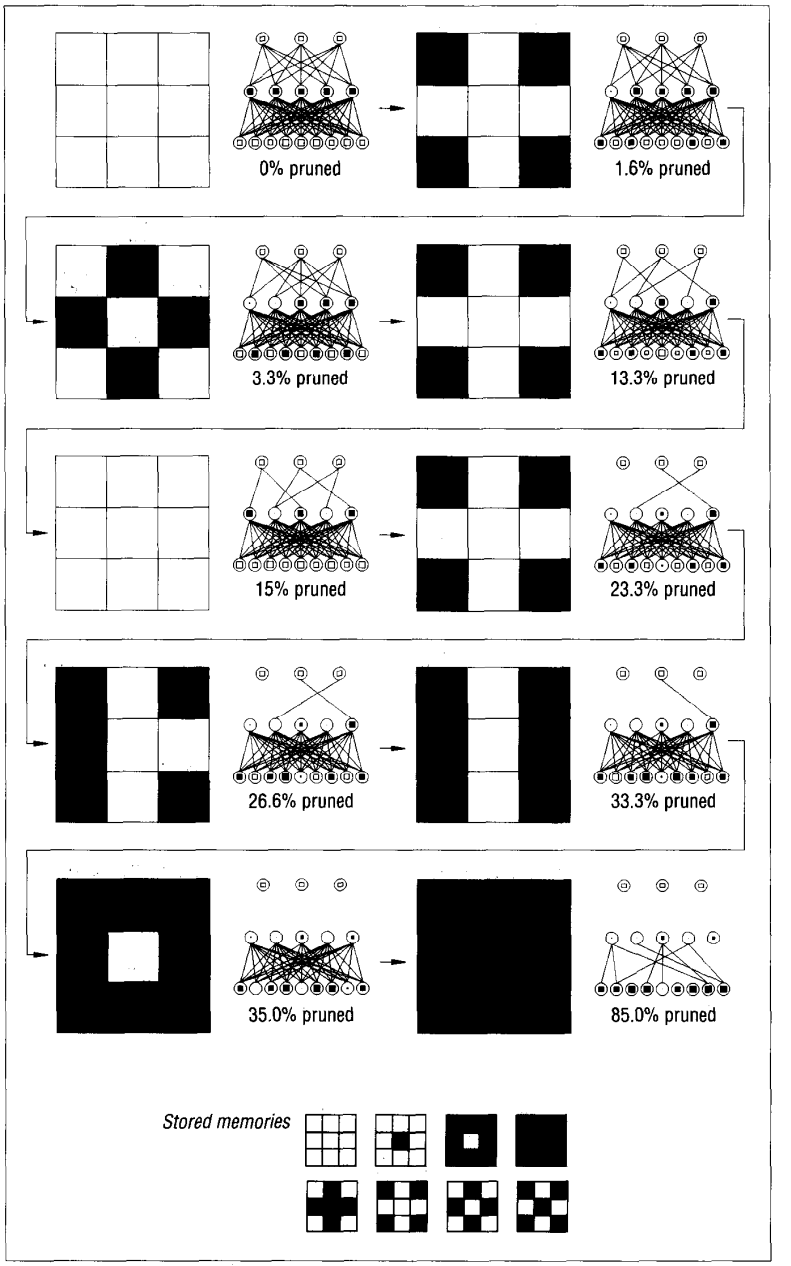
All neural networks can hallucinate, according to Thaler, because they all carry out this completion process; he knows of no case in biology that violates the processes he has described. His findings become even more valid, he believes, as connectivity—the number of layers, weights, and biases—increases in more complex neural systems.

Implications

Thaler's theories carry implications for a wide range of parallel distributed systems and applications. For example, the effect could play a role in the phantom experiences of traumatized biological networks, such as when a person feels the presence of an amputated arm or leg. The brain might assume that disconnected neural pathways are still sending messages, when in fact neural noise just looks like trained input.

The issue of safeguards on critical national defense systems also arises. Consider what would happen if a neural net-based missile system is damaged, perhaps by an unexpected rise in sunspot activity or by an insidious enemy using microwave energy. If Thaler is right, the system could mistakenly assume some false attack scenario (as if it were having a nightmare) and then react, possibly even launching a seemingly defensive strike. Safeguards must involve redundancy in the weight layers closest to the input layer of such feed-forward systems, ensuring that the layers closest to the outputs would be destroyed first. The system would then produce clearly distinguishable noise rather than hallucinations.

Perhaps most controversial, Thaler sees connections between the near-death hallucinations of his experimental networks and the near-death experiences of biological parallel and distributed systems—that is, of people. In fact, he goes even further: "If I'm right, even when simple animals die, they would have near-death experiences, too." The fact that Thaler can repeatedly produce the same hallucinations might indicate why many near-death reports describe similar scenes of tunnels and white lights. The experiments also showed that the number of hallucinations per unit time peaked sharply when the network



A degrading 3-5-9 pattern associator network, and the eight normal output patterns.

was very near death, perhaps resembling the cascading rate of brain cell death accompanying oxygen depletion, or the global release of numbing neuroinhibitors in reaction to pain. Because the rate of hallucination increases sharply near death, time seems to slow down from the network's viewpoint; this might explain why dying people's glimpses of eternity seem to occur in a moment's time. Finally, since dismantling neural nets produces (through completion) some of the early patterns used to train them, Thaler believes a person's experiences near the end of life might relate to memories of birth—one's first major trauma. He has summarized, "As the brain dismantles itself in death, a parade of the most dominant memories, associations, and expectations is released. Accepting this paradigm, we can see the value in purity of thought and conviction. Death is the harvest of our most automatic beliefs."

David Plaut, a Carnegie Mellon University psychologist who is using neural nets to model brain damage, sees intriguing possibilities in Thaler's ideas but cautions that more detailed work is needed to show how computer and neurobiological circuits correspond—and to prove that Thaler's results are not due solely to some idiosyncratic characteristic of simple networks. Of course, Thaler is aware that the brain is far more complex than the networks he's experimented with, and he is anxious to find colleagues who'd like to look at these issues using more biologically representative networks and more quantitative medical databases. And he continues to try to analyze these issues with plausible, mathematical, manipulatable models, rather than just leaps of faith. Thaler can be reached at McDonnell Douglas Aerospace East, Mail Code 1021322, PO Box 516, St. Louis, MO 63166-0516.

References

1. S.L. Thaler, "'Hallucination' Within the Death of a Simple Pattern Associator," under review. Available from the author.
2. S.L. Thaler, "4-2-4 Encoder Death," *Proc. World Congress on Neural Networks*, Vol. 2, Lawrence Erlbaum, Hillsdale, N.J., 1993, pp. 180-183.

AI researchers win awards

For his work in applying supercomputing and AI to natural language translation, Hiroaki Kitano won the Computers and Thought Award at this year's International Joint Conference on Artificial Intelligence. The $2,000 prize honors an outstanding AI researcher under 35, and this is the first time someone has won for work on a massively parallel computer. Having worked as a professional simultaneous interpreter and earned his PhD in 1991, Kitano is now a researcher at the Sony Computer Science Laboratory in Japan and the Center for Machine Translation at Carnegie Mellon University.

According to Kitano, the traditional AI approach to translation uses explicit representations of knowledge and is guided by rigid rules. But Kitano says this cannot work: "It is almost impossible to obtain a complete set of knowledge for a given problem." His approach builds on memory as the foundation of intelligence. When his Thinking Machines CM-2 receives a sentence as input, it checks all of its 32,000 processors, each containing a few sentences, and determines the best match. It makes the appropriate changes and produces a Japanese translation. The program has achieved 75-percent accuracy on 1,600 sentences.

Separately, Piero P. Bonissone has won the Coolidge Fellowship Award, which is the highest honor granted by the General Electric Research and Development Center. Considered the "father" of expert systems research within GE, Bonissone worked for years on the theory and application of reasoning with uncertain and incomplete information. His early fuzzy-logic projects dealt with defense, but more recent ones have ranged from jet engine control to more energy-efficient dishwashers. A member of IEEE, AAAI, and ACM and the holder of two patents, Bonissone now has a one-year leave of absence with full support to follow individual research.

Dante to descend again

Dante has not been shelved. Although severely damaged on its first journey into an active volcano, this eight-legged, rappelling robot will get another chance.

Last January, Dante tried to descend into Antarctica's 850-foot-deep Mt. Erebus to measure gas composition, retrieve gas samples, and record temperatures, but the mission was cancelled when the robot's fiber-optic communication line was severely damaged.

Carnegie Mellon University's Robotics Institute will refine and improve Dante using funds from NASA and the Advanced Research Projects Agency. Despite its problems, the shortened mission demonstrated Dante's 360-degree field-of-view laser and trinocular, stereo vision system. Now Dante will get stronger legs and an improved body frame, rappelling winch, fiber-optic cable, vision system, and motion control software.

The robot's next assignment is to investigate Mt. Spurr, a volcano 80 miles west of Anchorage. The trip is too dangerous for people: The volcano erupted three times in 1992, threatening Anchorage as well as regional air traffic. The project is designed to prove Dante's capabilities so that, at the end of 1994, it can return to explore the Mt. Erebus crater, traverse escarpments, deploy scientific equipment, and gather data and samples. Project manager John Bares hopes the mission will also validate the remote operation of robots via satellite, and the other telerobotic technologies developed at NASA over the last five years.

"NASA is interested in testing robotic technologies in harsh environments as a precursor to planetary exploration. Volcanologists are interested in the potential of robotic assistance to enable safer volcanic exploration," said principal scientist William L. Whittaker. He explained that the recent deaths of eight volcanologists mandates the development of remote methods to observe volcanos.

IBM builds expert system for OS/2 developers

To help OS/2 developers solve their technical problems, IBM has developed and distributed AskPSP, an expert systems-based application that provides intelligent read-only access to cases through Inference's CasePoint 1.3 for OS/2. The cases were developed by IBM's technical support staff using Inference's CBR Express.

After developers describe a problem in English, AskPSP either presents possible solutions based on its case library or, using deductive reasoning, asks questions to narrow its search. Each answered question is rated so that developers know whether they are getting closer to a final solution. The cases contain the most current information available and are structured around the way developers phrase their problems over the telephone.

IBM has distributed free copies of the program to more than 2,000 developers, and will also give it to a limited number of developers working on OS/2 and IBM local area network products. To obtain a free copy, call Personal Software Products Developer Assistance at (407) 982-6408.

"Knowbots" may manage knowledge of the future

Siemens Corporate Research scientists in Munich and Princeton are working on ways to manage information in a global network known as the Universal Personal Network. They are looking at two basic questions: how a computer can be programmed to understand people's natural-language inquiries, and how to create knowledge agents or "experts" containing all the information that would be needed.

At the core of the network, knobots, or software-based knowledge robots, will store vast amounts of knowledge, communicate with each other, and access each other's expertise. Ellen Voorhees, a member of technical staff of Siemens' Learning Systems Department, explained, "One of our primary goals is to find out how an 'expert' in fact becomes expert in any given subject, and to do this we're experimenting with different learning strategies."

"One of the biggest problems is to resolve the ambiguity inherent in natural languages. The word 'line,' for example, can have a number of meanings." So far, their program can classify "line" into the correct sense 75 percent of the time; their goal is to be 90- to 95-percent accurate. They are working with AI and neural networks to build learning systems that can analyze, monitor, track, and improve their responses to changing conditions. Siemens also hopes to use these information storage and retrieval and speech recognition strategies in other products.

Robots preserve US jobs

Has automation taken away jobs from American workers? Not at State Industries, a $350 million manufacturer of water heaters. The US manufacturer has remained competitive and even created new jobs by designing a new kind of product that lends itself to automated production. Eight workers now oversee the robots and other automated devices that make 1,200 tanks every eight hours; it used to take 75 employees that much time to make 1,000 tanks. The company has shifted most of its line workers and hired more to work on the older-design tanks. As older products are phased out, State Industries plans to retrain workers to avoid layoffs.

Columbia gets intelligent advice

During its October flight, the Columbia space shuttle carried the Astronaut Science Advisor, a general-purpose computer system designed to help astronauts work more efficiently and improve the scientific quality of collected data. The ASA centers around the idea that time is the critical resource during flight experiments. Larry Young of the Massachusetts Institute of Technology, who conceived the system, had served as a principal investigator on several shuttle missions. He decided a computer could help guide astronauts during life science experiments, since the scientists in space and on earth have limited abilities to correct problems or follow new leads as an experiment unfolds. The ASA was designed to manage experimental procedures, diagnose and troubleshoot equipment, collect data, and detect interesting data. It recognizes something as interesting by comparing collected data with predetermined rules. The system was developed jointly by NASA Ames Research Center and MIT.

The Rotating Dome Experiment was designed to study how the brain interprets conflicting information between inner-ear signals, visual cues, and other sensory data. The brain tries to integrate this information to tell the astronauts whether they are motionless or moving in a given direction at a given speed. These tests have already been done on the ground; doing them in space will give NASA more information about the lack of gravity. The other equipment for this experiment included a Macintosh PowerBook, LabView software for acquiring and analyzing data, NASA's Clips expert-system shell, and HyperCard.

The ASA tracked the time spent on each experiment and suggested ways to shorten procedures when sessions fell behind schedule. It could propose a new sequence of steps to get the best and most data in the time remaining, accounting for interesting data and the results of previous sessions. The ASA can also lead someone through step-by-step troubleshooting; if a problem is minor, the system might suggest not fixing it, gathering additional data instead.

Post office works on stroke extraction

The US Postal Service wants to learn how to relate the characteristics of pen strokes to numbers and letters of the alphabet, so it is funding stroke extraction research at the Center of Excellence for Document Analysis and Recognition (Cedar) at the State University of New York, Buffalo.

Cedar has begun to develop software that recognizes handwriting with enough speed and reliability to derive postal bar codes. After choosing a few plausible translations for a character or word, the program checks context; for instance, it knows what various cities' ZIP codes should start with. The Postal Service wants the system to process half of all handwritten mail with 98-percent accuracy. It expects to be able to handle 13 envelopes per second by the time the system is installed in mid-1995. Currently, its accuracy is 40 percent for complete addresses and ZIP codes, and 75 percent for five-digit ZIP codes.

Auto makers and federal scientists team up to make cleaner cars

Scientists at federal energy and defense laboratories will work with Chrysler, Ford, and General Motors engineers to triple American cars' fuel efficiency in the next 10 years. The joint project will emphasize technologies to improve efficiency, including advanced materials; advanced manufacturing techniques to rapidly design and test automation systems; and prototype vehicles with much greater fuel efficiency.

Funding will be split between the auto companies and the federal government, whose funds will probably come from existing lab programs and will be directed toward longer term, riskier research. Mary Lowe Good, Commerce Department undersecretary for technology administration, will head the federal interagency task force involved in the project.

EXPERT NEW PRODUCTS

Technical editors: W.F. Punch III, Michigan State University
Vendor contact: New Products, *IEEE Expert*, 10662 Los Vaqueros Circle, PO Box 3014, Los Alamitos, CA 90720-1264

Data mining with Recon

Lockheed now offers the Recon system and service: The company installs and customizes data-mining systems for companies that need to analyze large databases, thus revealing patterns and trends. Its Artificial Intelligence Center originally developed Recon to detect and fix manufacturing defects in printed circuit boards. Using the system and a database of specifications of 25,000 circuit boards, analysts created symbolic models that explained why some of the boards were defective.

Using Recon's visual language, the analyst defines and tests a set of rules using prior knowledge and experience. In this top-down data mining process, the rules are translated automatically and compared against the database. At the same time, Recon searches the database for relationships and patterns that are unknown to the analyst, and represents them as rules that have the same goals as those already defined. Recon presents the rules with a detailed explanation for their plausibility, and tries to predict whether new data will satisfy the analyst's goal. Most common business platforms are supported.

Reader service number 21

Fuzzy logic in 3D

Byte Dynamic's latest release of Fuzzy Logic Designer includes a 3D simulation mode that graphically depicts the response of fuzzy-logic designs in the form of surface plots or contour phases. Users create rule-based logic and membership functions through an interactive user interface, and the tool displays the functions graphically so that the user can view system responses and explore what-if scenarios. The simulation evaluates the underlying block of rule statements across the entire domain of a fuzzy set. For 3D simulation, two input variables are manipulated to display system response for all possible combinations.

While designing a system, the user can simulate how it responds to various input parameters. When the design is complete, the tool generates standard ANSI C source code that can be incorporated into applications or embedded processors. The target system generally cannot modify implemented rule statements or membership functions, since the tool generates code rather than an inference engine. No object library or inference engine is needed, and there are no royalties or license fees. The Fuzzy Logic Designer costs $395 and comes on a 3.5-inch or 5.25-inch high-density disk.

Reader service number 22

Scalable character recognition system

Alta Technology's Recognition Works! system combines Nestor's NestorReader software with off-the-shelf parallel hardware to create scalable intelligent character recognition tools for Unix and DOS-based machines. It recognizes hand- and machine-printed characters using neural network technology. The NestorReader Application Program Interface integrates with existing document image-processing applications, and is available for most of the common bus architectures.

Each character recognition engine can contain 2 to 200 processors, and can be embedded in other applications using custom hardware interfaces. The LSC Workbench, an optional software package, allows users to create a hardware accelerator. It includes an ANSI C compiler, linker, loader, and multinode debugger.

The integrated product's price starts at $13,000 and includes a host interface, NestorReader software, and a multi-