

Modeling Time Varying Systems Using Hidden Control Neural Architecture

Esther Levin

Multi-layered neural networks have recently been proposed for non(cid:173)linear prediction and system modeling. Although proven successful for modeling time invariant nonlinear systems, the inability of neural networks to characterize temporal variability has so far been an obstacle in applying them to complicated non stationary signals, such as speech.

In this paper we present a network architecture, called "Hidden Control Neural Network" (HCNN), for modeling signals generated by nonlinear dynamical systems with restricted time variability. The approach taken here is to allow the mapping that is implemented by a multi layered neural network to change with time as a function of an additional control input signal. This network is trained using an algorithm that is based on "back-propagation" and segmentation algorithms for estimating the unknown control together with the network's parameters. The HCNN approach was applied to several tasks including modeling of time-varying nonlinear systems and speaker-independent recognition of connected digits, yielding a word accuracy of 99.1 %.

Using Genetic Algorithms to Improve Pattern Classification Performance

Eric Chang, Richard P. Lippmann

Genetic algorithms were used to select and create features and to select reference exemplar patterns for machine vision and speech pattern classification tasks. For a complex speech recognition task, genetic algorithms required no more computation time than traditional approaches to feature selection but reduced the number of input features required by a factor of five (from 153 to 33 features). On a difficult artificial machine-vision task, genetic algorithms were able to create new features (polynomial functions of the original features) which reduced classification error rates from 19% to almost 0%. Neural net and k nearest neighbor (KNN) classifiers were unable to provide such low error rates using only the original features. Genetic algorithms were also used to reduce the number of reference exemplar patterns for a KNN classifier. On a 338 training pattern vowel-recognition problem with 10 classes, genetic algorithms reduced the number of stored exemplars from 338 to 43 without significantly increasing classification error rate. In all applications, genetic algorithms were easy to apply and found good solutions in many fewer trials than would be required by exhaustive search. Run times were long, but not unreasonable. These results suggest that genetic algorithms are becoming practical for pattern classification problems as faster serial and parallel computers are developed.

Generalization Dynamics in LMS Trained Linear Networks

Yves Chauvin

For a simple linear case, a mathematical analysis of the training and generalization (validation) performance of networks trained by gradient descent on a Least Mean Square cost function is provided as a function of the learning parameters and of the statistics of the training data base. The analysis predicts that generalization error dynamics are very dependent on a priori initial weights. In particular, the generalization error might sometimes weave within a computable range during extended training. In some cases, the analysis provides bounds on the optimal number of training cycles for minimal validation error. For a speech labeling task, predicted weaving effects were qualitatively tested and observed by computer simulations in networks trained by the linear and non-linear back-propagation algorithm.

Optimal Filtering in the Salamander Retina

Fred Rieke, W. Owen, William Bialek

The dark-adapted visual system can count photons with a reliability limited by thermal noise in the rod photoreceptors - the processing circuit

uity between the rod cells and the brain is essentially noiseless and in fact may be close to optimal. Here we design an optimal signal processor which estimates the time-varying light intensity at the retina based on the rod signals. We show that the first stage of optimal signal processing involves passing the rod cell output through a linear filter with characteristics determined entirely by the rod signal and noise spectra. This filter is very general; in fact it is the first stage in any visual signal processing task at 10¹⁰ photon flux. We identify the output of this first-stage filter with the intracellular voltage response of the bipolar cell the first anatomical stage in retinal signal processing. From recent data on tiger salamander photoreceptors we extract the relevant spectra and make parameter-free, quantitative predictions of the bipolar cell response to a dim, diffuse flash. Agreement with experiment is essentially perfect. As far as we know this is the first successful predictive theory for neural dynamics.

Oscillation Onset in Neural Delayed Feedback

André Longtin

This paper studies dynamical aspects of neural systems with delayed negative feedback modelled by nonlinear delay-differential equations. These systems undergo a Hopf bifurcation from a stable fixed point to a stable limit cycle oscillation as certain parameters are varied. It is shown that their frequency of oscillation is robust to parameter variations and noisy fluctuations, a property that makes these systems good candidates for pacemakers. The onset of oscillation is postponed by both additive and parametric noise in the sense that the state variable spends more time near the fixed point than it would in the absence of noise. This is also the case when noise affects the delayed variable, i.e. when the system has a faulty memory. Finally, it is shown that a distribution of delays (rather than a fixed delay) also stabilizes the fixed point solution.

Relaxation Networks for Large Supervised Learning Problems

Joshua Alspector, Robert Allen, Anthony Jayakumar, Torsten Zeppenfeld, Ronny Meir

Feedback connections are required so that the teacher signal on the output neurons can modify weights during supervised learning. Relaxation methods are needed for learning static patterns with full-time feedback connections. Feedback network learning techniques have not achieved wide popularity because of the still greater computational efficiency of back-propagation. We show by simulation that relaxation networks of the kind we are implementing in VLSI are capable of learning large problems just like back-propagation networks. A microchip incorporates deterministic mean-field theory learning as well as stochastic Boltzmann learning. A multiple-chip electronic system implementing these networks will make high-speed parallel learning in them feasible in the future.

Neural Networks Structured for Control Application to Aircraft Landing

Charles Schley, Yves Chauvin, Van Henkle, Richard Golden

We present a generic neural network architecture capable of controlling non-linear plants. The network is composed of dynamic, parallel, linear maps gated by non-linear switches. Using a recurrent form of the back-propagation algorithm, control is achieved by optimizing the control gains and task-adapted switch parameters. A mean quadratic cost function computed across a nominal plant trajectory is minimized along with performance constraint penalties. The approach is demonstrated for a control task consisting of landing a commercial aircraft in difficult wind conditions. We show that the network yields excellent performance while remaining within acceptable damping response constraints.

Closed-Form Inversion of Backpropagation Networks: Theory and Optimization Issues

Michael Rossen

We describe a closed-form technique for mapping the output of a trained backpropagation network into input activity space. The mapping is an inverse mapping in the sense that, when the image of the mapping in input activity space is propagated forward through the normal network dynamics, it reproduces the output used to generate that image. When more than one such inverse mappings exist, our inverse mapping is special in that it has no projection onto the nullspace of the activation flow operator for the entire network. An important by-product of our calculation, when more than one inverse mappings exist, is an orthogonal basis set of a significant portion of the activation flow operator nullspace. This basis set can be used to obtain an alternate inverse mapping that is optimized for a particular real-world application.

Dynamics of Generalization in Linear Perceptrons

Anders Krogh, John Hertz

We study the evolution of the generalization ability of a simple linear perceptron with N inputs which learns to imitate a "teacher perceptron". The system is trained on $p = \alpha N$ binary example inputs and the generalization ability measured by testing for agreement with the teacher on all $2N$ possible binary input patterns. The dynamics may be solved analytically and exhibits a phase transition from imperfect to perfect generalization at $\alpha = 1$. Except at this point the generalization ability approaches its asymptotic value exponentially, with critical slowing down near the transition; the relaxation time is $\propto (1 - \alpha)^{-2}$. Right at the critical point, the approach to perfect generalization follows a power law $\propto t^{-1/2}$. In the presence of noise, the generalization ability is degraded by an amount $\propto (\alpha - 1)^{-1}$ just above $\alpha = 1$.

Translating Locative Prepositions

Paul Munro, Mary Tabasko

A network was trained by back propagation to map locative expressions of the form "noun-preposition-noun" to a semantic representation, as in Cosic and Munro (1988). The network's performance was analyzed over several simulations with training sets in both English and German. Translation of prepositions was attempted by presenting a locative expression to a network trained in one language to generate a semantic representation; the semantic representation was then presented to the network trained in the other language to generate the appropriate preposition.

Adaptive Spline Networks

Jerome Friedman

A network based on splines is described. It automatically adapts the number of units, unit parameters, and the architecture of the network for each application.

An Analog VLSI Chip for Finding Edges from Zero-crossings

Wyeth Bair, Christof Koch

We have designed and tested a one-dimensional 64 pixel, analog CMOS VLSI chip which localizes intensity edges in real-time. This device exploits on-chip photoreceptors and the natural filtering properties of resistive networks to implement a scheme similar to and motivated by the Difference of Gaussians (DOG) operator proposed by Marr and Hildreth (1980). Our chip computes the zero-crossings associated with the difference of two exponential weighting functions. If the derivative across this zero-crossing is above a threshold, an edge is reported. Simulations indicate that this technique will extend well to two dimensions.

Comparison of three classification techniques: CART, C4.5 and Multi-Layer Perceptrons

A. C. Tsoi, R. Pearson

In this paper, after some introductory remarks into the classification problem as considered in various research communities, and some discussions concerning some of the reasons for ascertaining the performances of the three chosen algorithms, viz., CART (Classification and Regression Tree), C4.5 (one of the more recent versions of a popular induction tree technique known as ID3), and a multi-layer perceptron (MLP), it is proposed to compare the performances of these algorithms under two criteria: classification and generalisation. It is found that, in general, the MLP has better classification and generalisation accuracies compared with the other two algorithms.

ART2/BP architecture for adaptive estimation of dynamic processes

Einar Sørheim

The goal has been to construct a supervised artificial neural network that learns incrementally an unknown mapping. As a result a network consisting of a combination of ART2 and backpropagation is proposed and is called an "ART2/BP" network. The ART2 network is used to build and focus a supervised backpropagation network. The ART2/BP network has the advantage of being able to dynamically expand itself in response to input patterns containing new information. Simulation results show that the ART2/BP network outperforms a classical maximum likelihood method for the estimation of a discrete dynamic and nonlinear transfer function.

Self-organization of Hebbian Synapses in Hippocampal Neurons

Thomas Brown, Zachary Mainen, Anthony Zador, Brenda Claiborne

We are exploring the significance of biological complexity for neuronal computation. Here we demonstrate that Hebbian synapses in realistically-modeled hippocampal pyramidal cells may give rise to two novel forms of self-organization in response to structured synaptic input. First, on the basis of the electrotonic relationships between synaptic contacts, a cell may become tuned to a small subset of its input space. Second, the same mechanisms may produce clusters of potentiated synapses across the space of the dendrites. The latter type of self-organization may be functionally significant in the presence of nonlinear dendritic conductances.

Connection Topology and Dynamics in Lateral Inhibition Networks

C.M Marcus, F. Waugh, R. Westervelt

We show analytically how the stability of two-dimensional lateral inhibition neural networks depends on the local connection topology. For various network topologies, we calculate the critical time delay for the onset of oscillation in continuous-time networks and present analytic phase diagrams characterizing the dynamics of discrete-time networks.

A Delay-Line Based Motion Detection Chip

Tim Horiuchi, John Lazzaro, Andrew Moore, Christof Koch

Inspired by a visual motion detection model for the rabbit retina and by a computational architecture used for early audition in the barn owl, we have designed a chip that employs a correlation model to report the one-dimensional field motion of a scene in real time. Using subthreshold analog VLSI techniques, we have fabricated and successfully tested a 8000 transistor chip using a standard MOSIS process.

Back Propagation is Sensitive to Initial Conditions

John Kolen, Jordan Pollack

functions with

Applications of Neural Networks in Video Signal Processing

John Pearson, Clay D. Spence, Ronald Sverdlow

Although color TV is an established technology, there are a number of longstanding problems for which neural networks may be suited. Impulse noise is such a problem, and a modular neural network approach is presented in this paper. The training and analysis was done on conventional computers, while real-time simulations were performed on a massively parallel computer called the Princeton Engine. The network approach was compared to a conventional alternative, a median filter. Real-time simulations and quantitative analysis is demonstrated the technical superiority of the neural system. Ongoing work is investigating the complexity and cost of implementing this system in hardware.

Grouping Contours by Iterated Pairing Network

Amnon Shashua, Shimon Ullman

Shimon Ullman

Simulation of the Neocognitron on a CCD Parallel Processing Architecture

Michael Chuang, Alice Chiang

The neocognitron is a neural network for pattern recognition and feature extraction. An analog CCD parallel processing architecture developed at Lincoln Laboratory is particularly well suited to the computational requirements of shared-weight networks such as the neocognitron, and implementation of the neocognitron using the CCD architecture was simulated. A modification to the neocognitron training procedure, which improves network performance under the limited arithmetic precision that would be imposed by the CCD architecture, is presented.

Order Reduction for Dynamical Systems Describing the Behavior of Complex Neurons
Thomas Kepler, L. Abbott, Eve Marder

We have devised a scheme to reduce the complexity of dynamical systems belonging to a class that includes most biophysically realistic neural models. The reduction is based on transformations of variables and perturbation expansions and it preserves a high level of fidelity to the original system. The techniques are illustrated by reductions of the Hodgkin-Huxley system and an augmented Hodgkin-Huxley system.

Note on Learning Rate Schedules for Stochastic Optimization

Christian Darken, John Moody

We present and compare learning rate schedules for stochastic gradient descent, a general algorithm which includes LMS, on-line backpropagation and k-means clustering as special cases. We introduce "search-then-converge" type schedules which outperform the classical constant and "running average" (lft) schedules both in speed of convergence and quality of solution.

Remarks on Interpolation and Recognition Using Neural Nets

Eduardo Sontag

We consider different types of single-hidden-layer feedforward nets: with or without direct input to output connections, and using either threshold or sigmoidal activation functions. The main results show that direct connections in threshold nets double the recognition but not the interpolation power, while using sigmoids rather than thresholds allows (at least) doubling both. Various results are also given on VC dimension and other measures of recognition capabilities.

Rapidly Adapting Artificial Neural Networks for Autonomous Navigation

Dean Pomerleau

The ALVINN (Autonomous Land Vehicle In a Neural Network) project addresses the problem of training artificial neural networks in real time to perform difficult perception tasks. ALVINN is a back-propagation network that uses inputs from a video camera and an imaging laser rangefinder to drive the CMU Navla

b, a modified Chevy van. This paper describes training techniques which allow ALVINN to learn in under 5 minutes to autonomously control the Navlab by watching a human driver's response to new situations. Using these techniques, ALVINN has been trained to drive in a variety of circumstances including single-lane paved and unpaved roads, multilane lined and unlined roads, and obstacle-ridden on- and off-road environments, at speeds of up to 20 miles per hour.

A Recurrent Neural Network for Word Identification from Continuous Phoneme Strings

Robert Allen, Candace Kamm

A neural network architecture was designed for locating word boundaries and identifying words from phoneme sequences. This architecture was tested in three sets of studies. First, a highly redundant corpus with a restricted vocabulary was generated and the network was trained with a limited number of phonemic variations for the words in the corpus. Tests of network performance on a transfer set yielded a very low error rate. In a second study, a network was trained to identify words from expert transcriptions of speech. On a transfer test, error rate for correct simultaneous identification of words and word boundaries was 18%. The third study used the output of a phoneme classifier as the input to the word and word boundary identification network. The error rate on a transfer test set was 49% for this task. Overall, these studies provide a first step at identifying words in connected discourse with a neural network.

Adjoint-Functions and Temporal Learning Algorithms in Neural Networks

N. Toomarian, J. Barhen

The development of learning algorithms is generally based upon the minimization of an energy function. It is a fundamental requirement to compute the gradient of this energy function with respect to the various parameters of the neural architecture, e.g., synaptic weights, neural gain, etc.

In principle, this requires solving a system of nonlinear equations for each parameter of the model, which is computationally very expensive. A new methodology for neural learning of time-dependent nonlinear mappings is presented. It exploits the concept of adjoint operators to enable a fast global computation of the network's response to perturbations in all the systems parameters. The importance of the time boundary conditions of the adjoint functions is discussed. An algorithm is presented in which the adjoint sensitivity equations are solved simultaneously (i.e., forward in time) along with the nonlinear dynamics of the neural networks. This methodology makes real-time applications and hardware implementation of temporal learning feasible.

Language Induction by Phase Transition in Dynamical Recognizers

Jordan Pollack

A higher order recurrent neural network architecture learns to recognize and generate languages after being "trained" on categorized exemplars. Studying these networks from the perspective of dynamical systems yields two interesting discoveries: First, a longitudinal examination of the learning process illustrates a new form of mechanical inference: Induction by phase transition. A small weight adjustment causes a "bifurcation" in the limit behavior of the network. This phase transition corresponds to the onset of the network's capacity for generalizing to arbitrary-length strings.

Second, a study of the automata resulting from the acquisition of previously published languages indicates that while the architecture is NOT guaranteed to find a minimal finite automata consistent with the given exemplars, which is an NP-Hard problem, the architecture does appear capable of generating non-regular languages by exploiting fractal and chaotic dynamics. I end the paper with a hypothesis relating linguistic generative capacity to the behavioral regimes of non-linear dynamical systems.

.

Chaitin-Kolmogorov Complexity and Generalization in Neural Networks

Barak Pearlmutter, Ronald Rosenfeld

We present a unified framework for a number of different ways of failing to generalize properly. During learning, sources of random information contaminate the network, effectively augmenting the training data with random information. The complexity of the function computed is therefore increased, and generalization is degraded. We analyze replicated networks, in which a number of identical networks are independently trained on the same data and their results averaged. We conclude that replication almost always results in a decrease in the expected complexity of the network, and that replication therefore increases expected generalization. Simulations confirming the effect are also presented.

VLSI Implementations of Learning and Memory Systems: A Review

Mark Holler

A large number of VLSI implementations of neural network models have been reported. The diversity of these implementations is noteworthy. This paper attempts to put a group of representative VLSI implementations in perspective by comparing and contrasting them. Design trade-offs are discussed and some suggestions for the direction of future implementation efforts are made.

Neural Dynamics of Motion Segmentation and Grouping

Ennio Mingolla

A neural network model of motion segmentation by visual cortex is described. The model clarifies how preprocessing of motion signals by a Motion Oriented Contrast Filter (MOC Filter) is joined to long-range cooperative motion mechanisms in a motion Cooperative Competitive Loop (CC Loop) to control phenomena such as induced motion, motion capture, and motion aftereffects. The total model system is a motion Boundary Contour System (BCS) that is computed in parallel with a static BCS before both systems cooperate to generate a boundary representation for three dimensional visual form perception. The present investigations clarify how the static BCS can be modified for use in motion segmentation problems, notably for analyzing how ambiguous local movements (the aperture problem) on a complex moving shape are suppressed and actively reorganized into a coherent global motion signal.

Exploratory Feature Extraction in Speech Signals

Nathan Intrator

A novel unsupervised neural network for dimensionality reduction which seeks directions emphasizing multimodality is presented, and its connection to exploratory projection pursuit methods is discussed. This leads to a new statistical insight to the synaptic modification equations governing learning in Bienenstock, Cooper, and Munro (BCM) neurons (1982). The importance of a dimensionality reduction principle based solely on distinguishing features, is demonstrated using a linguistically motivated phoneme recognition experiment, and compared with feature extraction using back-propagation network.

Constructing Hidden Units using Examples and Queries

Eric Baum, Kevin Lang

While the network learning problem for 2-layer threshold nets is NP-hard when learning from examples alone (as with backpropagation), (Baum, 91) has now proved that a learner can employ queries to evade the hidden unit credit assignment problem and PAC-load nets with up to four hidden units in polynomial time. Empirical tests show that the method can also learn far more complicated functions such as randomly generated networks with 200 hidden units. The algorithm easily approximates Wieland's 2-spirals function using a single layer of 50 hidden units, and requires only 30 minutes of CPU time to learn 200-bi

t parity to 99.7% accuracy.

Analog Neural Networks as Decoders

Ruth Erlanson, Yaser Abu-Mostafa

Analog neural networks with feedback can be used to implement 1((cid:173) Winner-Take-All (KWT) networks. In turn, KWT networks can be used as decoders of a class of nonlinear error-correcting codes. By in(cid:173) terconnecting such KWT A networks, we can construct decoders capable of decoding more powerful codes. We consider several families of inter(cid:173) connected KWT networks, analyze their performance in terms of coding theory metrics, and consider the feasibility of embedding such networks in VLSI technologies.

A Reinforcement Learning Variant for Control Scheduling

Aloke Guha

We present an algorithm based on reinforcement and state recurrence learning techniques to solve control scheduling problems. In particular, we have devised a simple learning scheme called "handicapped learning", in which the weights of the associative search element are reinforced, either positively or negatively, such that the system is forced to move towards the desired setpoint in the shortest possible trajectory. To improve the learning rate, a variable reinforcement scheme is employed: negative reinforcement values are varied depending on whether the failure occurs in handicapped or normal mode of operation. Furthermore, to realize a simulated annealing scheme for accelerated learning, if the system visits the same failed state successively, the negative reinforcement value is increased. In examples studied, these learning schemes have demonstrated high learning rates, and therefore may prove useful for in-situ learning.

Kohonen Networks and Clustering: Comparative Performance in Color Clustering

Wesley Snyder, Daniel Nissman, David Van den Bout, Griff Bilbro

The problem of color clustering is defined and shown to be a problem of assigning a large number (hundreds of thousands) of 3-vectors to a small number (256) of clusters. Finding those clusters in such a way that they best represent a full color image using only 256 distinct colors is a burdensome computational problem. In this paper, the problem is solved using "classical" techniques -- k-means clustering, vector quantization (which turns out to be the same thing in this application), competitive learning, and Kohonen self-organizing feature maps. Quality of the result is judged subjectively by how much the pseudo-color result resembles the true color image, by RMS quantization error, and by run time. The Kohonen map provides the best solution.

Reconfigurable Neural Net Chip with 32K Connections

H. P. Graf, R. Janow, D. Henderson, R. Lee

We describe a CMOS neural net chip with a reconfigurable network architecture. It contains 32,768 binary, programmable connections arranged in 256 'building block' neurons. Several 'building blocks' can be connected to form long neurons with up to 1024 binary connections or to form neurons with analog connections. Single- or multi-layer networks can be implemented with this chip. We have integrated this chip into a board system together with a digital signal processor and fast memory. This system is currently in use for image processing applications in which the chip extracts features such as edges and corners from binary and gray-level images.

e-Entropy and the Complexity of Feedforward Neural Networks

Robert C. Williamson

We develop a new feedforward neural network representation of Lipschitz functions from $[0, p] \times \mathbb{R}^n$ into $[0, 1]$ based on the level sets of the function. We show that

Extensions of a Theory of Networks for Approximation and Learning: Outliers and Negative Examples

Federico Girosi, Tomaso Poggio, Bruno Caprile

Learning an input-output mapping from a set of examples can be regarded as synthesizing an approximation of a multi-dimensional function. From this point of view, this form of learning is closely related to regularization theory, and we have previously shown (Poggio and Girosi, 1990a, 1990b) the equivalence between regularized least squares and a class of three-layer networks that we call regularization networks. In this note, we extend the theory by introducing ways of

Stochastic Neurodynamics

J.D. Cowan

The main point of this paper is that stochastic neural networks have a mathematical structure that corresponds quite closely with that of quantum field theory. Neural network Liouvillians and Lagrangians can be derived, just as can spin Hamiltonians and Lagrangians in QFT. It remains to show the efficacy of such a description.

Phonetic Classification and Recognition Using the Multi-Layer Perceptron

Hong Leung, James Glass, Michael Phillips, Victor W. Zue

In this paper, we will describe several extensions to our earlier work, utilizing a segment-based approach. We will formulate our segmental framework and report our study on the use of multi-layer perceptrons for detection and classification of phonemes. We will also examine the outputs of the network, and compare the network performance with other classifiers. Our investigation is performed within a set of experiments that attempts to recognize 38 vowels and consonants in American English independent of speaker. When evaluated on the TIM IT database, our system achieves an accuracy of 56%.

Learning Trajectory and Force Control of an Artificial Muscle Arm by Parallel-hierarchical Neural Network Model

Masazumi Katayama, Mitsuo Kawato

We propose a new parallel-hierarchical neural network model to enable motor learning for simultaneous control of both trajectory and force. By integrating Hogan's control method and our previous neural network control model using a feedback-error-learning scheme. Furthermore, two hierarchical control laws which apply to the model, are derived by using the Moore-Penrose pseudo-inverse matrix. One is related to the minimum muscle-tension-change trajectory and the other is related to the minimum motor-command-change trajectory. The human arm is redundant at the dynamics level since joint torque is generated by agonist and antagonist muscles. Therefore, acquisition of the inverse model is an ill-posed problem. However, the combination of these control laws and feedback-error-learning resolve the ill-posed problem. Finally, the efficiency of the parallel-hierarchical neural network model is shown by learning experiments using an artificial muscle arm and computer simulations.

Basis-Function Trees as a Generalization of Local Variable Selection Methods for Function Approximation

Terence Sanger

Local variable selection has proven to be a powerful technique for approximating functions in high-dimensional spaces. It is used in several statistical methods, including CART, ID3, C4, MARS, and others (see the bibliography for references to these algorithms). In this paper I present a tree-structured network which is a generalization of these techniques. The network provides a framework for understanding the behavior of such algorithms and for modifying them to suit particular applications.

A Short-Term Memory Architecture for the Learning of Morphophonemic Rules

Michael Gasser, Chan-Do Lee

Despite its successes, Rumelhart and McClelland's (1986) well-known approach to the learning of morphophonemic rules suffers from two deficiencies: (1) It performs the artificial task of associating forms with forms rather than perception or production. (2) It is not constrained in ways that human learners are. This paper describes a model which addresses both objections. Using a simple recurrent architecture which takes both forms and "meanings" as inputs, the model learns to generate verbs in one or another "tense", given arbitrary meanings, and to recognize the tenses of verbs. Furthermore, it fails to learn reversal processes unknown in human language.

Associative Memory in a Network of 'Biological' Neurons

Wulfram Gerstner

The Hopfield network (Hopfield, 1982,1984) provides a simple model of an associative memory in a neuronal structure. This model, however, is based on highly artificial assumptions, especially the use of formal-two state neurons (Hopfield, 1982) or graded-response neurons (Hopfield, 1984). What happens if we replace the formal neurons by 'real' biological neurons? We address this question in two steps. First, we show that a simple model of a neuron can capture all relevant features of neuron spiking, i. e., a wide range of spiking frequencies and a realistic distribution of interspike intervals. Second, we construct an associative memory by linking these neurons together. The analytical solution for a large and fully connected network shows that the Hopfield solution is valid only for neurons with a short refractory period. If the refractory period is longer than a critical duration i.e., the solutions are qualitatively different. The associative character of the solutions, however, is preserved.

Oriented Non-Radial Basis Functions for Image Coding and Analysis

Avijit Saha, Jim Christian, Dun-Sung Tang, Wu Chuan-Lin

We introduce oriented non-radial basis function networks (ONRBF) as a generalization of Radial Basis Function networks (RBF)- wherein the Euclidean distance metric in the exponent of the Gaussian is replaced by a more general polynomial. This permits the definition of more general regions and in particular - hyper-ellipses with orientations. In the case of hyper-surface estimation this scheme requires a smaller number of hidden units and alleviates the "curse of dimensionality" associated kernel type approximators. In the case of an image, the hidden units correspond to features in the image and the parameters associated with each unit correspond to the rotation, scaling and translation properties of that particular "feature". In the context of the ONBF scheme, this means that an image can be represented by a small number of features. Since, transformation of an image by rotation, scaling and translation correspond to identical transformations of the individual features, the ONBF scheme can be used to considerable advantage for the purposes of image recognition and analysis.

Evaluation of Adaptive Mixtures of Competing Experts

Steven Nowlan, Geoffrey E. Hinton

We compare the performance of the modular architecture, composed of competing expert networks, suggested by Jacobs, Jordan, Nowlan and Hinton (1991) to the performance of a single back-propagation network on a complex, but low-dimensional, vowel recognition task. Simulations reveal that this system is capable of uncovering interesting decompositions in a complex task. The type of decomposition is strongly influenced by the nature of the input to the gating network that decides which expert to use for each case. The modular architecture also exhibits consistently better generalization on many variations of the task.

Spoken Letter Recognition

Mark Fanty, Ronald Cole

Through the use of neural network classifiers and careful feature selection, we have achieved high-accuracy speaker-independent spoken letter recognition. For isolated letters, a broad-category segmentation is performed. Location of segment boundaries allows us to measure features at specific locations in the signal such as vowel onset, where important information resides. Letter classification is performed with a feed-forward neural network. Recognition accuracy on a test set of 30 speakers was 96%. Neural network classifiers are also used for pitch tracking and broad-category segmentation of letter strings. Our research has been extended to recognition of names spelled with pauses between the letters. When searching a database of 50,000 names, we achieved 95% first choice name retrieval. Work has begun on a continuous letter classifier which does frame-by-frame phonetic classification of spoken letters.

Design and Implementation of a High Speed CMAC Neural Network Using Programmable CMOS Logic Cell Arrays

W. Miller, Brian Box, Erich Whitney, James Glynn

A high speed implementation of the CMAC neural network was designed using dedicated CMOS logic. This technology was then used to implement two general purpose CMAC associative memory boards for the VME bus. Each board implements up to 8 independent CMAC networks with a total of one million adjustable weights. Each CMAC network can be configured to have from 1 to 512 integer inputs and from 1 to 8 integer outputs. Response times for typical CMAC networks are well below 1 millisecond, making the networks sufficiently fast for most robot control problems, and many pattern recognition and signal processing problems.

Shaping the State Space Landscape in Recurrent Networks

Patrice Simard, Jean Rayz, Bernard Victorri

Bernard Victorri ELSAP Universite de Caen 14032 Caen Cedex France

EMPATH: Face, Emotion, and Gender Recognition Using Holons

Garrison Cottrell, Janet Metcalfe

The dimensionality of a set of 160 1:~ :a:~s ~.10 . female subjects IS reduced network The extracted features do not correspond to in previous face recognition systems (KaR-nae, 19~i)y' f~tures we call holons. The holons are fV~ t~! \ d' tances between facial elements.

Interaction Among Ocularity, Retinotopy and On-center/Off-center Pathways During Development

Shigeru Tanaka

The development of projections from the retinas to the cortex is mathematically analyzed according to the previously proposed thermodynamic formulation of the self-organization of neural networks. Three types of submodality included in the visual afferent pathways are assumed in two models: model (A), in which the ocularity and retinotopy are considered separately, and model (B), in which on-center/off-center pathways are considered in addition to ocularity and retinotopy. Model (A) shows striped ocular dominance spatial patterns and, in ocular dominance histograms, reveals a dip in the binocular bin. Model (B) displays spatially modulated irregular patterns and shows single-peak behavior in the histograms. When we compare the simulated results with the observed results, it is evident that the ocular dominance spatial patterns and histograms for models (A) and (B) agree very closely with those seen in monkeys and cats.

Phase-coupling in Two-Dimensional Networks of Interacting Oscillators

Ernst Niebur, Daniel Kammen, Christof Koch, Daniel Ruderman, Heinz Schuster

Coherent oscillatory activity in large networks of biological or artificial neural units may be a useful mechanism for coding information pertaining to a single perceptual object or for detailing regularities within a data set. We consider the dynamics of a large array of simple coupled oscill

ators under a variety of connection schemes. Of particular interest is the rapid and robust phase-locking that results from a "sparse" scheme where each oscillator is strongly coupled to a tiny, randomly selected subset of its neighbors.

An Analog VLSI Splining Network

Daniel Schwartz, Vijay Samalam

We have produced a VLSI circuit capable of learning to approximate arbitrary smooth of a single variable using a technique closely related to splines. The circuit effectively has 512 knots space on a uniform grid and has full support for learning. The circuit also can be used to approximate multi-variable functions as sum of splines.

Designing Linear Threshold Based Neural Network Pattern Classifiers

Terrence L. Fine

The three problems that concern us are identifying a natural domain of pattern classification applications of feed forward neural networks, selecting an appropriate feedforward network architecture, and assessing the tradeoff between network complexity, training set size, and statistical reliability as measured by the probability of incorrect classification. We close with some suggestions, for improving the bounds that come from Vapnik-Chervonkis theory, that can narrow, but not close, the chasm between theory and practice.

Qualitative structure from motion

Daphna Weinshall

Exact structure from motion is an ill-posed computation and therefore very sensitive to noise. In this work I describe how a qualitative shape representation, based on the sign of the Gaussian curvature, can be computed directly from motion disparities, without the computation of an exact depth map or the directions of surface normals. I show that humans can judge the curvature sense of three points undergoing 3D motion from two, three and four views with success rate significantly above chance. A simple RBF net has been trained to perform the same task.

Learning to See Rotation and Dilation with a Hebb Rule

Martin Sereno, Margaret Sereno

Previous work (M.I. Sereno, 1989; cf. M.E. Sereno, 1987) showed that a feedforward network with area VI-like input-layer units and a Hebb rule can develop area MT-like second layer units that solve the aperture problem for pattern motion. The present study extends this earlier work to more complex motions. Saito et al. (1986) showed that neurons with large receptive fields in macaque visual area MST are sensitive to different senses of rotation and dilation, irrespective of the receptive field location of the movement singularity. A network with an MT-like second layer was trained and tested on combinations of rotating, dilating, and translating patterns. Third-layer units learn to detect specific senses of rotation or dilation in a position-independent fashion, despite having position-dependent direction selectivity within their receptive fields.

A Comparative Study of the Practical Characteristics of Neural Network and Conventional Pattern Classifiers

Kenney Ng, Richard P. Lippmann

Seven different pattern classifiers were implemented on a serial computer and compared using artificial and speech recognition tasks. Two neural network (radial basis function and high order polynomial GMDH network) and five conventional classifiers (Gaussian mixture, linear tree, K nearest neighbor, KD-tree, and condensed K nearest neighbor) were evaluated. Classifiers were chosen to be representative of different approaches to pattern classification and to complement and extend those evaluated in a previous study (Lee and Lippmann, 1989). This and the previous study both demonstrate that classification error rates

can be equivalent across different classifiers when they are powerful enough to form minimum error decision regions, when they are properly tuned, and when sufficient training data is available. Practical characteristics such as training time, classification time, and memory requirements, however, can differ by orders of magnitude. These results suggest that the selection of a classifier for a particular task should be guided not so much by small differences in error rate, but by practical considerations concerning memory usage, computational resources, ease of implementation, and restrictions on training and classification times.

Real-time autonomous robot navigation using VLSI neural networks

Lionel Tarassenko, Michael Brownlow, Gillian Marshall, Jan Tombs, Alan Murray

We describe a real time robot navigation system based on three VLSI neural network modules. These are a resistive grid for path planning, a nearest-neighbour classifier for localization using range data from a time-of-flight infrared sensor and a sensory-motor associative network for dynamic obstacle avoidance.

CAM Storage of Analog Patterns and Continuous Sequences with $3N^2$ Weights

Bill Baird, Frank Eeckman

A simple architecture and algorithm for analytically guaranteed associative memory storage of analog patterns, continuous sequences, and chaotic attractors in the same network is described. A matrix inversion determines network weights, given prototype patterns to be stored. There are N units of capacity in an N node network with $3N^2$ weights. It costs one unit per static attractor, two per Fourier component of each sequence, and four per chaotic attractor. There are no spurious attractors, and there is a Lyapunov function in a special coordinate system which governs the approach of transient states to stored trajectories. Unsupervised or supervised incremental learning algorithms for pattern classification, such as competitive learning or bootstrap Widrow-Hoff can easily be implemented. The architecture can be "folded" into a recurrent network with higher order weights that can be used as a model of cortex that stores oscillatory and chaotic attractors by a Hebb rule. Hierarchical sensory-motor control networks may be constructed of interconnected "cortical patches" of these network modules. Network performance is being investigated by application to the problem of real time handwritten digit recognition.

Generalization Properties of Radial Basis Functions

Sherif Botros, Christopher Atkeson

We examine the ability of radial basis functions (RBFs) to generalize. We compare the performance of several types of RBFs. We use the inverse dynamics of an idealized two-joint arm as a test case. We find that without a proper choice of a norm for the inputs, RBFs have poor generalization properties. A simple global scaling of the input variables greatly improves performance. We suggest some efficient methods to approximate this distance metric.

Learning Time-varying Concepts

Anthony Kuh, Thomas Petsche, Ronald Rivest

This work extends computational learning theory to situations in which concepts vary over time, e.g., system identification of a time-varying plant. We have extended formal definitions of concepts and learning to provide a framework in which an algorithm can track a concept as it evolves over time. Given this framework and focusing on memory-based algorithms, we have derived some PAC-style sample complexity results that determine, for example, when tracking is feasible.

We have also used a similar framework and focused on incremental tracking algorithms for which we have derived some bounds on the mistake or error rates for some specific concept classes.

A Theory for Neural Networks with Time Delays

Bert de Vries, José Príncipe

We present a new neural network model for processing of temporal patterns. This model, the gamma neural model, is as general as a convolution delay model with arbitrary weight kernels $w(t)$. We show that the gamma model can be formulated as a (partially prewired) additive model. A temporal hebbian learning rule is derived and we establish links to related existing models for temporal processing.

Natural Dolphin Echo Recognition Using an Integrator Gateway Network

Herbert Roitblat, Patrick Moore, Paul Nachtigall, Ralph Penner

We have been studying the performance of a bottlenosed dolphin on a delayed matching-to-sample task to gain insight into the processes and mechanisms that the animal uses during echolocation. The dolphin recognizes targets by emitting natural sonar signals and listening to the echoes that return. This paper describes a novel neural network architecture, called an integrator gateway network, that we have developed to account for this performance. The integrator gateway network combines information from multiple echoes to classify targets with about 90% accuracy. In contrast, a standard backpropagation network performed with only about 63% accuracy.

Compact EEPROM-based Weight Functions

A. Kramer, C. Sin, R. Chu, P. Ko

We are focusing on the development of a highly compact neural net weight function based on the use of EEPROM devices. These devices have already proven useful for analog weight storage, but existing designs rely on the use of conventional voltage multiplication as the weight function, requiring additional transistors per synapse. A parasitic capacitance between the floating gate and the drain of the EEPROM structure leads to an unusual J-V characteristic which can be used to advantage in designing a compact synapse. This novel behavior is well characterized by a model we have developed. A single-device circuit results in a 1-quadrant synapse function which is nonlinear, though monotonic. A simple extension employing 2 EEPROMs results in a 2 quadrant function which is much more linear. This approach offers the potential for more than a ten-fold increase in the density of neural net implementations.

Discovering Viewpoint-Invariant Relationships That Characterize Objects

Richard Zemel, Geoffrey E. Hinton

Using an unsupervised learning procedure, a network is trained on an ensemble of images of the same two-dimensional object at different positions, orientations and sizes. Each half of the network "sees" one fragment of the object, and tries to produce as output a set of 4 parameters that have high mutual information with the 4 parameters output by the other half of the network. Given the ensemble of training patterns, the 4 parameters on which the two halves of the network can agree are the position, orientation, and size of the whole object, or some recoding of them. After training, the network can reject instances of other shapes by using the fact that the predictions made by its two halves disagree. If two competing networks are trained on an unlabelled mixture of images of two objects, they cluster the training cases on the basis of the objects' shapes, independently of the position, orientation, and size.

Reinforcement Learning in Markovian and Non-Markovian Environments

Jürgen Schmidhuber

This work addresses three problems with reinforcement learning and adaptive neuro-control: 1. Non-Markovian interfaces between learner and environment. 2. On-line learning based on system realization. 3. Vector-valued adaptive critics. An algorithm is described which is based on system realization and on two interacting fully recurrent continually running networks which may learn in parallel. Problems with parallel

learning are attacked by 'adaptive randomness'. It is also described how interacting model/controller systems can be combined with vector-valued 'adaptive critics' (previous critics have been scalar).

Second Order Properties of Error Surfaces: Learning Time and Generalization

Yann LeCun, Ido Kanter, Sara Solla

The learning time of a simple neural network model is obtained through an analytic computation of the eigenvalue spectrum for the Hessian matrix, which describes the second order properties of the cost function in the space of coupling coefficients. The form of the eigenvalue distribution suggests new techniques for accelerating the learning process, and provides a theoretical justification for the choice of centered versus biased state variables.

Connectionist Music Composition Based on Melodic and Stylistic Constraints

Michael C. Mozer, Todd Soukup

We describe a recurrent connectionist network, called CONCERT, that uses a set of melodies written in a given style to compose new melodies in that style. CONCERT is an extension of a traditional algorithmic composition technique in which transition tables specify the probability of the next note as a function of previous context. A central ingredient of CONCERT is the use of a psychologically-grounded representation of pitch.

Transforming Neural-Net Output Levels to Probability Distributions

John Denker, Yann LeCun

(1) The outputs of a typical multi-output classification network do not satisfy the axioms of probability; probabilities should be positive and sum to one. This problem can be solved by treating the trained network as a preprocessor that produces a feature vector that can be further processed, for instance by classical statistical estimation techniques. (2) We present a method for computing the first two moments of the probability distribution indicating the range of outputs that are consistent with the input and the training data. It is particularly useful to combine these two ideas: we implement the ideas of section 1 using Parzen windows, where the shape and relative size of each window is computed using the ideas of section 2. This allows us to make contact between important theoretical ideas (e.g. the ensemble formalism) and practical techniques (e.g. back-prop). Our results also shed new light on and generalize the well-known "soft max" scheme.

Can neural networks do better than the Vapnik-Chervonenkis bounds?

David Cohn, Gerald Tesauro

We describe a series of careful numerical experiments which measure the average generalization capability of neural networks trained on a variety of simple functions. These experiments are designed to test whether average generalization performance can surpass the worst-case bounds obtained from formal learning theory using the Vapnik-Chervonenkis dimension (Blumer et al., 1989). We indeed find that, in some cases, the average generalization is significantly better than the VC bound: the approach to perfect performance is exponential in the number of examples m , rather than the $\ln m$ result of the bound. In other cases, we do find the $\ln m$ behavior of the VC bound, and in these cases, the numerical prefactor is closely related to prefactor contained in the bound.

Discovering Discrete Distributed Representations with Iterative Competitive Learning

Michael C. Mozer

Competitive learning is an unsupervised algorithm that classifies input patterns into mutually exclusive clusters. In a neural net framework, each cluster is represented by a processing unit that competes with others in a winner-take-all pool for an input pattern. I present a simple extension to the algorithm that allows it to construct discrete, distributed

representations. Discrete representations are useful because they are relatively easy to analyze and their information content can readily be measured. Distributed representations are useful because they explicitly encode similarity. The basic idea is to apply competitive learning iteratively to an input pattern, and after each stage to subtract from the input pattern the component that was captured in the representation at that stage. This component is simply the weight vector of the winning unit of the competitive pool. The subtraction procedure forces competitive pools at different stages to encode different aspects of the input. The algorithm is essentially the same as a traditional data compression technique known as multistep vector quantization, although the neural network perspective suggests potentially powerful extensions to that approach.

A Neural Expert System with Automated Extraction of Fuzzy If-Then Rules and Its Application to Medical Diagnosis

Yoichi Hayashi

This paper proposes a fuzzy neural expert system (FNES) with the following two functions: (1) Generalization of the information derived from the training data and embodiment of knowledge in the form of the fuzzy neural network; (2) Extraction of fuzzy If-Then rules with linguistic relative importance of each proposition in an antecedent (If-part) from a trained neural network. This paper also gives a method to extract automatically fuzzy If-Then rules from the trained neural network. To prove the effectiveness and validity of the proposed fuzzy neural expert system, a fuzzy neural expert system for medical diagnosis has been developed.

Learning by Combining Memorization and Gradient Descent

John Platt

We have created a radial basis function network that allocates a new computational unit whenever an unusual pattern is presented to the network. The network learns by allocating new units and adjusting the parameters of existing units. If the network performs poorly on a presented pattern, then a new unit is allocated which memorizes the response to the presented pattern. If the network performs well on a presented pattern, then the network parameters are updated using standard LMS gradient descent. For predicting the Mackey Glass chaotic time series, our network learns much faster than do those using back-propagation and uses a comparable number of synapses.

A Neural Network Approach for Three-Dimensional Object Recognition

Volker Tresp

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

Distributed Recursive Structure Processing

Geraldine Legendre, Yoshiro Miyata, Paul Smolensky

Harmonic grammar (Legendre, et al., 1990) is a connectionist theory of linguistic well-formedness based on the assumption that the well-formedness of a sentence can be measured by the harmony (negative energy) of the corresponding connectionist state. Assuming a lower-level connectionist network that obeys a few general connectionist principles but is otherwise unspecified, we construct a higher-level network with an equivalent harmony function that captures the most linguistically relevant global aspects of the lower level network. In this paper, we extend the tensor product representation (Smolensky 1990) to fully recursive representations of recursively structured objects like sentences in the lower-level network. We show theoretically and with an example the power of the new technique for parallel distributed structure processing.

Dynamics of Learning in Recurrent Feature-Discovery Networks

Todd Leen

The self-organization of recurrent feature-discovery networks is studied from the perspective of dynamical systems. Bifurcation theory reveals parameter regimes in which multiple equilibria or limit cycles coexist with the equilibrium at which the networks perform principal component analysis.

Navigating through Temporal Difference

Peter Dayan

Barto, Sutton and Watkins [2] introduced a grid task as a didactic example of temporal difference planning and asynchronous dynamical programming. This paper considers the effects of changing the coding of the input stimulus, and demonstrates that the self-supervised learning of a particular form of hidden unit representation improves performance.

On the Circuit Complexity of Neural Networks

V. P. Roychowdhury, K. Y. Siu, A. Orlitsky, T. Kailath

We introduce a geometric approach for investigating the power of threshold circuits. Viewing n -variable boolean functions as vectors in \mathbb{R}^2 , we invoke tools from linear algebra and linear programming to derive new results on the realizability of boolean functions using threshold gates. Using this approach, one can obtain: (1) upper-bounds on the number of spurious memories in Hopfield networks, and on the number of functions implementable by a depth- d threshold circuit; (2) a lower bound on the number of orthogonal input functions required to implement a threshold function; (3) a necessary condition for an arbitrary set of input functions to implement a threshold function; (4) a lower bound on the error introduced in approximating boolean functions using sparse polynomials; (5) a limit on the effectiveness of the only known lower-bound technique (based on computing correlations among boolean functions) for the depth of threshold circuits implementing boolean functions, and (6) a constructive proof that every boolean function f of n input variables is a threshold function of polynomially many input functions, none of which is significantly correlated with f . Some of these results lead to generalizations of key results concerning threshold circuit complexity, particularly those that are based on the so-called spectral or Harmonic analysis approach. Moreover, our geometric approach yields simple proofs, based on elementary results from linear algebra, for many of these earlier results.

An Attractor Neural Network Model of Recall and Recognition

Eytan Ruppin, Yehezkel Yeshurun

This work presents an Attractor Neural Network (ANN) model of Recall and Recognition. It is shown that an ANN model can qualitatively account for a wide range of experimental psychological data pertaining to these two main aspects of memory access. Certain psychological phenomena are accounted for, including the effects of list-length, word frequency, presentation time, context shift, and aging. Thereafter, the probabilities of successful Recall and Recognition are estimated, in order to possibly enable further quantitative examination of the model.

Training Knowledge-Based Neural Networks to Recognize Genes in DNA Sequences

Michiel Noordewier, Geoffrey Towell, Jude Shavlik

We describe the application of a hybrid symbolic/connectionist machine learning algorithm to the task of recognizing important genetic sequences. The symbolic portion of the KBANN system utilizes inference rules that provide a rough correct method for recognizing a class of DNA sequences known as eukaryotic splice-junctions. We then map this "domain theory" into a neural network and provide training examples. Using the samples, the neural network's learning algorithm adjusts the domain theory so that it properly classifies these DNA sequences. Our procedure constitutes a general method for incorporating preexisting knowledge

ge into artificial neural networks. We present an experiment in molecular genetics that demonstrates the value of doing so.

ALCOVE: A Connectionist Model of Human Category Learning

John Kruschke

ALCOVE is a connectionist model of human category learning that fits a broad spectrum of human learning data. Its architecture is based on well(cid:173) established psychological theory, and is related to networks using radial basis functions. From the perspective of cognitive psychology, ALCOVE can be construed as a combination of exemplar-based representation and error(cid:173) driven learning. From the perspective of connectionism, it can be seen as incorporating constraints into back-propagation networks appropriate for modelling human learning.

Multi-Layer Perceptrons with B-Spline Receptive Field Functions

Stephen Lane, Marshall Flax, David Handelman, Jack Gelfand

Multi-layer perceptrons are often slow to learn nonlinear functions with complex local structure due to the global nature of their function approximations. It is shown that standard multi-layer perceptrons are actually a special case of a more general network formulation that incorporates B-splines into the node computations. This allows novel spline network architectures to be developed that can combine the generalization capabilities and scaling properties of global multi-layer feedforward networks with the computational efficiency and learning speed of local computational paradigms. Simulation results are presented for the well known spiral problem of Weiland and of Lang and Witbrock to show the effectiveness of the Spline Net approach.

Bumptrees for Efficient Function, Constraint and Classification Learning

Stephen Omohundro

A new class of data structures called "bumptrees" is described. These structures are useful for efficiently implementing a number of neural network related operations. An empirical comparison with radial basis functions is presented on a robot arm mapping learning task. Applications to density estimation, classification, and constraint representation and learning are also outlined.

Planning with an Adaptive World Model

Sebastian Thrun, Knut Möller, Alexander Linden

We present a new connectionist planning method [TML90]. By interaction with an unknown environment, a world model is progressively constructed(cid:173) using gradient descent. For deriving optimal actions with respect to future reinforcement, planning is applied in two steps: an experience net(cid:173) work proposes a plan which is subsequently optimized by gradient descent with a chain of world models, so that an optimal reinforcement may be obtained when it is actually run. The appropriateness of this method is demonstrated by a robotics application and a pole balancing task.

Simple Spin Models for the Development of Ocular Dominance Columns and Iso-Orientation Patches

J.D. Cowan, A. Friedman

Simple classical spin models well-known to physicists as the ANNNI and Heisenberg XY Models, in which long-range interactions occur in a pattern given by the Mexican Hat operator, can generate many of the structural properties characteristic of the ocular dominance columns and iso-orientation patches seen in cat and primate visual cortex.

A Multiscale Adaptive Network Model of Motion Computation in Primates

H. Wang, Bimal Mathur, Christof Koch

We demonstrate a multiscale adaptive network model of motion computation in primate area MT. The model consists of two stages: (1) local velocities are

e measured across multiple spatio-temporal channels, and (2) the optical flow field is computed by a network of direction(cid:173) selective neurons at multiple spatial resolutions. This model embeds the computational efficiency of Multigrid algorithms within a parallel network as well as adaptively computes the most reliable estimate of the flow field across different spatial scales. Our model neurons show the same nonclassical receptive field properties as Allman's type I MT neurons. Since local velocities are measured across multiple channels, various channels often provide conflicting measurements to the network. We have incorporated a veto scheme for conflict resolution. This mechanism provides a novel explanation for the spatial frequency dependency of the psychophysical phenomenon called Motion Capture.

Spherical Units as Dynamic Consequential Regions: Implications for Attention, Competition and Categorization

Stephen Hanson, Mark Gluck

to construct dynamic

Speech Recognition Using Connectionist Approaches

Khalid Choukri

This paper is a summary of SPRINT project aims and results. The project focus on the use of neuro-computing techniques to tackle various problems that remain unsolved in speech recognition. First results concern the use of feed(cid:173)forward nets for phonetic units classification, isolated word recognition, and speaker adaptation.

Feedback Synapse to Cone and Light Adaptation

Josef Skrzypek

Light adaptation (LA) allows cone vision to remain functional between twilight and the brightest time of day even though, at any one time, their intensity-response (I-R) characteristic is limited to 3 log units of the stimulating light. One mechanism underlying LA, was localized in the outer segment of an isolated cone (1,2). We found that by adding annular illumination, an I-R characteristic of a cone can be shifted along the intensity domain. Neural network involving feedback synapse from horizontal cells to cones is involved to be in register with ambient light level of the periphery. An equivalent electrical circuit with three different transmembrane channels leakage, photocurrent and feedback was used to model static behavior of a cone. SPICE simulation showed that interactions between feedback synapse and the light sensitive conductance in the outer segment can shift the I-R curves along the intensity domain, provided that phototransduction mechanism is not saturated during maximally hyperpolarized light response.

Direct memory access using two cues: Finding the intersection of sets in a connectionist model

Janet Wiles, Michael Humphreys, John Bain, Simon Dennis

For lack of alternative models, search and decision processes have provided the dominant paradigm for human memory access using two or more cues, despite evidence against search as an access process (Humphreys, Wiles & Bain, 1990). We present an alternative process to search, based on calculating the intersection of sets of targets activated by two or more cues. Two methods of computing the intersection are presented, one using information about the possible targets, the other constraining the cue-target strengths in the memory matrix. Analysis using orthogonal vectors to represent the cues and targets demonstrates the competence of both processes, and simulations using sparse distributed representations demonstrate the performance of the latter process for tasks involving 2 and 3 cues.

Flight Control in the Dragonfly: A Neurobiological Simulation

William Faller, Marvin Luttges

Neural network simulations of the dragonfly flight neurocontrol system have been developed to understand how this insect uses complex, unsteady aerodynamics. The simulation networks account for the ganglionic spatial distribution of cells as well as the physiologic operating range and the stochastic cellular firing history of each neuron. In addition the motor neuron firing patterns, "flight command sequences", were utilized. Simulation training was targeted against both the cellular and flight motor neuron firing patterns. The trained networks accurately resynthesized the intraganglionic cellular firing patterns. These in turn controlled the motor neuron firing patterns that drive wing musculature during flight. Such networks provide both neurobiological analysis tools and first generation controls for the use of "unsteady" aerodynamics.

A Framework for the Cooperation of Learning Algorithms

Léon Bottou, Patrick Gallinari

We introduce a framework for training architectures composed of several modules. This framework, which uses a statistical formulation of learning systems, provides a unique formalism for describing many classical connectionist algorithms as well as complex systems where several algorithms interact. It allows to design hybrid systems which combine the advantages of connectionist algorithms as well as other learning algorithms.

Continuous Speech Recognition by Linked Predictive Neural Networks

Joe Tebelskis, Alex Waibel, Bojan Petek, Otto Schmidbauer

We present a large vocabulary, continuous speech recognition system based on Linked Predictive Neural Networks (LPNN's). The system uses neural networks as predictors of speech frames, yielding distortion measures which are used by the One Stage DTW algorithm to perform continuous speech recognition. The system, already deployed in a Speech to Speech Translation system, currently achieves 95%, 58%, and 39% word accuracy on tasks with perplexity 5, 111, and 402 respectively, outperforming several simple HMMs that we tested. We also found that the accuracy and speed of the LPNN can be slightly improved by the judicious use of hidden control inputs. We conclude by discussing the strengths and weaknesses of the predictive approach.

A B-P ANN Commodity Trader

Joseph Collard

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

Statistical Mechanics of Temporal Association in Neural Networks

Andreas Herz, Zhaoping Li, J. van Hemmen

We study the representation of static patterns and temporal associations in neural networks with a broad distribution of signal delays. For a certain class of such systems, a simple intuitive understanding of the spatio-temporal computation becomes possible with the help of a novel Lyapunov functional. It allows a quantitative study of the asymptotic network behavior through a statistical mechanical analysis. We present analytic calculations of both retrieval quality and storage capacity and compare them with simulation results.

Cholinergic Modulation May Enhance Cortical Associative Memory Function

Michael Hasselmo, Brooke Anderson, James Bower

Combining neuropharmacological experiments with computational modeling, we have shown that cholinergic modulation may enhance associative memory function in piriform (olfactory) cortex. We have shown that the acetylcholine analogue carbachol selectively suppresses synaptic transmission between cel

ls within piriform cortex, while leaving input connections unaffected. When tested in a computational model of piriform cortex, this selective suppression, applied during learning, enhances associative memory performance.

Neural Network Application to Diagnostics and Control of Vehicle Control Systems
Kenneth Marko

Diagnosis of faults in complex, real-time control systems is a complicated task that has resisted solution by traditional methods. We have shown that neural networks can be successfully employed to diagnose faults in digitally controlled powertrain systems. This paper discusses the means we use to develop the appropriate databases for training and testing in order to select the optimum network architectures and to provide reasonable estimates of the classification accuracy of these networks on new samples of data. Recent work applying neural nets to adaptive control of an active suspension system is presented.

Evolution and Learning in Neural Networks: The Number and Distribution of Learning Trials Affect the Rate of Evolution

Ron Keesing, David Stork

and

A Lagrangian Approach to Fixed Points

Eric Mjolsness, Willard Miranker

We present a new way to derive dissipative, optimizing dynamics from the Lagrangian formulation of mechanics. It can be used to obtain both standard and novel neural net dynamics for optimization problems. To demonstrate this we derive standard descent dynamics as well as nonstandard variants that introduce a computational attention mechanism.

Neural Network Implementation of Admission Control

Rodolfo Milito, Isabelle Guyon, Sara Solla

A feedforward layered network implements a mapping required to control an unknown stochastic nonlinear dynamical system. Training is based on a novel approach that combines stochastic approximation ideas with backpropagation. The method is applied to control admission into a queueing system operating in a time-varying environment.

Computing with Arrays of Bell-Shaped and Sigmoid Functions

Pierre Baldi

We consider feed-forward neural networks with one non-linear hidden layer and linear output units. The transfer function in the hidden layer are either bell-shaped or sigmoid. In the bell-shaped case, we show how Bernstein polynomials on one hand and the theory of the heat equation on the other are relevant for understanding the properties of the corresponding networks. In particular, these techniques yield simple proofs of universal approximation properties, i.e. of the fact that any reasonable function can be approximated to any degree of precision by a linear combination of bell-shaped functions. In addition, in this framework the problem of learning is equivalent to the problem of reversing the time course of a diffusion process. The results obtained in the bell-shaped case can then be applied to the case of sigmoid transfer functions in the hidden layer, yielding similar universality results. A conjecture related to the problem of generalization is briefly examined.

VLSI Implementation of TInMANN

Matt Melton, Tan Phan, Doug Reeves, Dave Van den Bout

A massively parallel, all-digital, stochastic architecture - TInMANN - is described which performs competitive and Kohonen types of learning. A VLSI design is shown for a TInMANN neuron which fits within a small, inexpensive MOSIS TinyChip frame, yet which can be used to build larger networks of several hundred neurons. The neuron operates at a speed of 15 MHz which allows the network to pro

cess 290,000 training examples per second. Use of level sensitive scan logic provides the chip with 100% fault coverage, permitting very reliable neural systems to be built.

SEXNET: A Neural Network Identifies Sex From Human Faces

B.A. Golomb, D.T. Lawrence, T.J. Sejnowski

Sex identification in animals has biological importance. Humans are good at making this determination visually, but machines have not matched this ability. A neural network was trained to discriminate sex in human faces, and performed as well as humans on a set of 90 exemplars. Images sampled at 30x30 were compressed using a 900x40x900 fully-connected back-propagation network; activities of hidden units served as input to a back-propagation "SexNet" trained to produce values of 1 for male and 0 for female faces. The network's average error rate of 8.1% compared favorably to humans, who averaged 11.6%. Some SexNet errors mimicked those of humans.

From Speech Recognition to Spoken Language Understanding: The Development of the MIT SUMMIT and VOYAGER Systems

Victor Zue, James Glass, David Goodine, Lynette Hirschman, Hong Leung, Michael P. Hillips, Joseph Polifroni, Stephanie Seneff

Spoken language is one of the most natural, efficient, flexible, and economical means of communication among humans. As computers play an ever increasing role in our lives, it is important that we address the issue of providing a graceful human-machine interface through spoken language. In this paper, we will describe our recent efforts in moving beyond the scope of speech recognition into the realm of spoken-language understanding. Specifically, we report on the development of an urban navigation and exploration system called VOYAGER, an application which we have used as a basis for performing research in spoken-language understanding.

Generalization by Weight-Elimination with Application to Forecasting

Andreas Weigend, David Rumelhart, Bernardo Huberman

Inspired by the information theoretic idea of minimum description length, we add a term to the back propagation cost function that penalizes network complexity. We give the details of the procedure, called weight-elimination, describe its dynamics, and clarify the meaning of the parameters involved. From a Bayesian perspective, the complexity term can be usefully interpreted as an assumption about prior distribution of the weights. We use this procedure to predict the sunspot time series and the notoriously noisy series of currency exchange rates.

A Novel Approach to Prediction of the 3-Dimensional Structures of Protein Backbones by Neural Networks

Henrik Fredholm, Henrik Bohr, Jakob Bohr, Søren Brunak, Rodney Cotterill, Benny Lautrup, Steffen Petersen

Three-dimensional (3D) structures of protein backbones have been predicted using neural networks. A feed forward neural network was trained on a class of functionally, but not structurally, homologous proteins, using backpropagation learning. The network generated tertiary structure information in the form of binary distance constraints for the C α atoms in the protein backbone. The binary distance between two C α atoms was 0 if the distance between them was less than a certain threshold distance, and 1 otherwise. The distance constraints predicted by the trained neural network were utilized to generate a folded conformation of the protein backbone, using a steepest descent minimization approach.

Back Propagation Implementation on the Adaptive Solutions CNAPS Neurocomputer Chip

Hal McCartor

The Adaptive Solutions CNAPS architecture chip is a general purpose neurocomputer

ter chip. It has 64 processors, each with 4 K bytes of local memory, running at 25 megahertz. It is capable of implementing most current neural network algorithms with on chip learning. This paper discusses the implementation of the Back Propagation algorithm on an array of these chips and shows performance figures from a clock accurate hardware simulator. An eight chip configuration on one board can update 2.3 billion connections per second in learning mode and process 9.6 billion connections per second in feed forward mode.

Asymptotic slowing down of the nearest-neighbor classifier

Robert Snapp, Demetri Psaltis, Santosh Venkatesh

If patterns are drawn from an n-dimensional feature space according to a probability distribution that obeys a weak smoothness criterion, we show that the probability that a random input pattern is misclassified by a nearest-neighbor classifier using M random reference patterns asymptotically satisfies

Connectionist Approaches to the Use of Markov Models for Speech Recognition

Hervé Bourlard, Nelson Morgan, Chuck Wooters

Previous work has shown the ability of Multilayer Perceptrons (MLPs) to estimate emission probabilities for Hidden Markov Models (HMMs). The advantages of a speech recognition system incorporating both MLPs and HMMs are the best discrimination and the ability to incorporate multiple sources of evidence (features, temporal context) without restrictive assumptions of distributions or statistical independence. This paper presents results on the speaker-dependent portion of DARPA's English language Resource Management database. Results support the previously reported utility of MLP probability estimation for continuous speech recognition. An additional approach we are pursuing is to use MLPs as nonlinear predictors for autoregressive HMMs. While this is shown to be more compatible with the HMM formalism, it still suffers from several limitations. This approach is generalized to take account of time correlation between successive observations, without any restrictive assumptions about the driving noise.

A Second-Order Translation, Rotation and Scale Invariant Neural Network

Shelly Goggin, Kristina Johnson, Karl Gustafson

A second-order architecture is presented here for translation, rotation and scale invariant processing of 2-D images mapped to n input units. This new architecture has a complexity of $O(n)$ weights as opposed to the $O(n^3)$ weights usually required for a third-order, rotation invariant architecture. The reduction in complexity is due to the use of discrete frequency information. Simulations show favorable comparisons to other neural network architectures.

Signal Processing by Multiplexing and Demultiplexing in Neurons

David Tam

to

Lg Depth Estimation and Ripple Fire Characterization Using Artificial Neural Networks

John Perry, Douglas Baumgardt

This study has demonstrated how artificial neural networks (ANNs) can be used to characterize seismic sources using high-frequency regional seismic data. We have taken the novel approach of using ANNs as a research tool for obtaining seismic source information, specifically depth of focus for earthquakes and ripple-fire characteristics for economic blasts, rather than as just a feature classifier between earthquake and explosion populations. Overall, we have found that ANNs have potential applications to seismic event characterization and identification, beyond just as a feature classifier. In future studies, these techniques should be applied to actual data of regional seismic events recorded at the new regional seismic arrays. The results of this study indicates that an ANN should

ould be evaluated as part of an operational seismic event identification system .

Adaptive Range Coding
Bruce Rosen, James Goodwin, Jacques Vidal
these

RecNorm: Simultaneous Normalisation and Classification applied to Speech Recognition
John Bridle, Stephen J. Cox
A particular form of neural network is described, which has terminals for acoustic patterns, class labels and speaker parameters. A method of training this network to "tune in" the speaker parameters to a particular speaker is outlined, based on a trick for converting a supervised network to an unsupervised mode. We describe experiments using this approach in isolated word recognition based on whole-word hidden Markov models. The results indicate an improvement over speaker-independent performance and, for unlabelled data, a performance close to that achieved on labelled data.

A four neuron circuit accounts for change sensitive inhibition in salamander retina
Jeffrey Teeters, Frank Eeckman, Frank Werblin
In salamander retina, the response of On-Off ganglion cells to a central flash is reduced by movement in the receptive field surround. Through computer simulation of a 2-D model which takes into account their anatomical and physiological properties, we show that interactions between four neuron types (two bipolar and two amacrine) may be responsible for the generation and lateral conductance of this change sensitive inhibition. The model shows that the four neuron circuit can account for previously observed movement sensitive reductions in ganglion cell sensitivity and allows visualization and prediction of the spatio-temporal pattern of activity in change sensitive retinal cells.

A Method for the Efficient Design of Boltzmann Machines for Classification Problems
Ajay Gupta, Wolfgang Maass
We introduce a method for the efficient design of a Boltzmann machine (or a Hopfield net) that computes an arbitrary given Boolean function f . This method is based on an efficient simulation of acyclic circuits with threshold gates by Boltzmann machines. As a consequence we can show that various concrete Boolean functions f that are relevant for classification problems can be computed by scalable Boltzmann machines that are guaranteed to converge to their global maximum configuration with high probability after constantly many steps.

How Receptive Field Parameters Affect Neural Learning
Bartlett Mel, Stephen Omohundro
We identify the three principle factors affecting the performance of learning by networks with localized units: unit noise, sample density, and the structure of the target function. We then analyze the effect of unit receptive field parameters on these factors and use this analysis to propose a new learning algorithm which dynamically alters receptive field properties during learning.

Proximity Effect Corrections in Electron Beam Lithography Using a Neural Network
Robert Frye, Kevin Cummings, Edward Rietman
We have used a neural network to compute corrections for images written by electron beams to eliminate the proximity effects caused by electron scattering. Iterative methods are effective, but require prohibitively long computation time. We have instead trained a neural network to perform equivalent corrections, resulting in a significant speed-up. We have examined hardware implementations using both analog and digital electronic networks. Both had an acceptably small error

r of 0.5% compared to the iterative results. Additionally, we verified that the neural network correctly generalized the solution of the problem to include patterns not contained in its training set. We have experimentally verified this approach on a Cambridge Instruments EBMF 10.5 exposure system.

Exploiting Syllable Structure in a Connectionist Phonology Model

David Touretzky, Deirdre Wheeler

In a previous paper (Touretzky & Wheeler, 1990a) we showed how adding a clustering operation to a connectionist phonology model produced a parallel processing account of certain "iterative" phenomena. In this paper we show how the addition of a second structuring primitive, syllabification, greatly increases the power of the model. We present examples from a non-Indo-European language that appear to require rule ordering to at least a depth of four. By adding syllabification circuitry to structure the model's perception of the input string, we are able to handle these examples with only two derivational steps. We conclude that in phonology, derivation can be largely replaced by structuring.

Integrated Modeling and Control Based on Reinforcement Learning and Dynamic Programming

Richard S. Sutton

This is a summary of results with Dyna, a class of architectures for intelligent systems based on approximating dynamic programming methods. Dyna architectures integrate trial-and-error (reinforcement) learning and execution-time planning into a single process operating alternately on the world and on a learned forward model of the world. We describe and show results for two Dyna architectures, Dyna-AHC and Dyna-Q. Using a navigation task, results are shown for a simple Dyna-AHC system which simultaneously learns by trial and error, learns a world model, and plans optimal routes using the evolving world model. We show that Dyna-Q architectures (based on Watkins's Q-learning) are easy to adapt for use in changing environments.

A Connectionist Learning Control Architecture for Navigation

Jonathan R. Bachrach

A novel learning control architecture is used for navigation. A sophisticated test-bed is used to simulate a cylindrical robot with a sonar belt in a planar environment. The task is short-range homing in the presence of obstacles. The robot receives no global information and assumes no comprehensive world model. Instead the robot receives only sensory information which is inherently limited. A connectionist architecture is presented which incorporates a large amount of a priori knowledge in the form of hard-wired networks, architectural constraints, and initial weights. Instead of hard-wiring static potential fields from object models, myarchitecture learns sensor-based potential fields, automatically adjusting them to avoid local minima and to produce efficient homing trajectories. It does this without object models using only sensory information. This research demonstrates the use of a large modular architecture on a difficult task.

On Stochastic Complexity and Admissible Models for Neural Network Classifiers

Padhraic Smyth

Given some training data how should we choose a particular network classifier from a family of networks of different complexities? In this paper we discuss how the application of stochastic complexity theory to classifier design problems can provide some insights into this problem. In particular we introduce the notion of admissible models whereby the complexity of models under consideration is affected by (among other factors) the class entropy, the amount of training data, and our prior belief. In particular we discuss the implications of these results with respect to neural architectures and demonstrate the approach on real data from a medical diagnosis task.

Analog Computation at a Critical Point: A Novel Function for Neuronal Oscillations?

Leonid Kruglyak, William Bialek

We show that a simple spin system biased at its critical point can encode spatial characteristics of external signals, such as the dimensions of "objects" in the visual field. In the temporal correlation functions of individual spins. Qualitative arguments suggest that regularly firing neurons should be described by a planar spin of unit length. and such XY models exhibit critical dynamics over a broad range of parameters. We show how to extract these spins from spike trains and then measure the interaction Hamiltonian using simulations of small clusters of cells. Static correlations among spike trains obtained from simulations of large arrays of cells are in agreement with the predictions from these Hamiltonians, and dynamic correlations display the predicted encoding of spatial information. We suggest that this novel representation of object dimensions in temporal correlations may be relevant to recent experiments on oscillatory neural firing in the visual cortex.

Integrated Segmentation and Recognition of Hand-Printed Numerals

James Keeler, David Rumelhart, Wee Leow

Neural network algorithms have proven useful for recognition of individual, segmented characters. However, their recognition accuracy has been limited by the accuracy of the underlying segmentation algorithm.

Conventional, rule-based segmentation algorithms encounter difficulty if the characters are touching, broken, or noisy. The problem in these situations is that often one cannot properly segment a character until it is recognized yet one cannot properly recognize a character until it is segmented. We present here a neural network algorithm that simultaneously segments and recognizes in an integrated system. This algorithm has several novel features: it uses a supervised learning algorithm (backpropagation), but is able to take position-independent information as targets and self-organize the activities of the units in a competitive fashion to infer the positional information. We demonstrate this ability with overlapping hand-printed numerals.

Speech Recognition Using Demi-Syllable Neural Prediction Model

Ken-ichi Iso, Takao Watanabe

The Neural Prediction Model is the speech recognition model based on pattern prediction by multilayer perceptrons. Its effectiveness was confirmed by the speaker-independent digit recognition experiments.

This paper presents an improvement in the model and its application to large vocabulary speech recognition, based on subword units. The improvement involves an introduction of "backward prediction," which further improves the prediction accuracy of the original model with only "forward prediction". In application of the model to speaker-dependent large vocabulary speech recognition, the demi-syllable unit is used as a subword recognition unit. Experimental results indicated a 95.2% recognition accuracy for a 5000 word test set and the effectiveness was confirmed for the proposed model improvement and the demi-syllable subword units.

A Recurrent Neural Network Model of Velocity Storage in the Vestibulo-Ocular Reflex

Thomas Anastasio

A three-layered neural network model was used to explore the organization of the vestibulo-ocular reflex (VOR). The dynamic model was trained using recurrent back-propagation to produce compensatory, long duration eye muscle motoneuron outputs in response to short duration vestibular afferent head velocity inputs. The network learned to produce this response prolongation, known as velocity storage, by developing complex, lateral inhibitory interactions among the interneurons. These had the low basal

ine, long time constant, rectified and skewed responses that are characteristic of real VOR inter(cid:173)neurons. The model suggests that all of these features are interrelated and result from lateral inhibition.

The Devil and the Network: What Sparsity Implies to Robustness and Memory
Sanjay Biswas, Santosh Venkatesh

Robustness is a commonly bruited property of neural networks; in particu(cid:173)lar, a folk theorem in neural computation asserts that neural networks-in contexts with large interconnectivity-continue to function efficiently, al(cid:173)beit with some degradation, in the presence of component damage or loss. A second folk theorem in such contexts asserts that dense interconnectiv(cid:173)ity between neural elements is a sine qua non for the efficient usage of resources. These premises are formally examined in this communication in a setting that invokes the notion of the "devil" l in the network as an agent that produces sparsity by snipping connections.

Development and Spatial Structure of Cortical Feature Maps: A Model Study
Klaus Obermayer, Helge Ritter, Klaus Schulten

K. Schulten Beckman -Insti t u te University of Illinois Urbana, IL 61801

Connectionist Implementation of a Theory of Generalization

Roger Shepard, Sheila Kannappan

Empirically, generalization between a training and a test stimulus falls off in close approximation to an exponential decay function of distance b etween the two stimuli in the "stimulus space" obtained by multidimensional sc aling. Math(cid:173)ematically, this result is derivable from the assumption that an individual takes the training stimulus to belong to a "conseque ntial" region that includes that stimulus but is otherwise of unknown loca tion, size, and shape in the stimulus space (Shepard, 1987). As the individua l gains additional information about the consequential region-by finding other stimuli to be consequential or nol-the theory predicts the shape of the gene ralization function to change toward the function relating actual probabil ity of the consequence to location in the stimulus space. This paper describes a natural connectionist implementation of the theory, and illustrates how impl ications of the theory for generalization, discrimination, and classification l earning can be explored by connectionist simulation.

Discrete Affine Wavelet Transforms For Anaylsis And Synthesis Of Feedfoward Neur al Networks

Y. Pati, P. Krishnaprasad

In this paper we show that discrete affine wavelet transforms can provide a too l for the analysis and synthesis of standard feedforward neural net(cid:173)works. It is shown that wavelet frames for L2(IR) can be constructed based upon si gmoids. The spatia-spectral localization property of wavelets can be exploited in defining the topology and determining the weights of a feedforward network. Training a network constructed using the synthe(cid:173)sis procedure described here involves minimization of a convex cost func(cid:173)tional and therefore avoids pitfalls inherent in standard backpropagation algorithms. Extension of t hese methods to L2(IRN) is also discussed.

Convergence of a Neural Network Classifier

John Baras, Anthony LaVigna

In this paper, we prove that the vectors in the LVQ learning algorit hm converge. We do this by showing that the learning algorithm perfor ms stochastic approximation. Convergence is then obtained by identifying the appropriate conditions on the learning rate and on the underlyin g statistics of the classification problem. We also present a modificati on to the learning algorithm which we argue results in convergence of t he LVQ error to the Bayesian optimal error as the appropriate parameters beco me large.

Optimal Sampling of Natural Images: A Design Principle for the Visual System

William Bialek, Daniel Ruderman, A. Zee

We formulate the problem of optimizing the sampling of natural images using an array of linear filters. Optimization of information capacity is constrained by the noise levels of the individual channels and by a penalty for the construction of long-range interconnections in the array. At low signal-to-noise ratios the optimal filter characteristics correspond to bound states of a Schrodinger equation in which the signal spectrum plays the role of the potential. The resulting optimal filters are remarkably similar to those observed in the mammalian visual cortex and the retinal ganglion cells of lower vertebrates. The observed scale invariance of natural images plays an essential role in this construction.

Time Trials on Second-Order and Variable-Learning-Rate Algorithms

Richard Rohwer

The performance of seven minimization algorithms are compared on five neural network problems. These include a variable-step-size algorithm, conjugate gradient, and several methods with explicit analytic or numerical approximations to the Hessian.

A competitive modular connectionist architecture

Robert Jacobs, Michael Jordan

We describe a multi-network, or modular, connectionist architecture that captures that fact that many tasks have structure at a level of granularity intermediate to that assumed by local and global function approximation schemes. The main innovation of the architecture is that it combines associative and competitive learning in order to learn task decompositions. A task decomposition is discovered by forcing the networks comprising the architecture to compete to learn the training patterns. As a result of the competition, different networks learn different training patterns and, thus, learn to partition the input space. The performance of the architecture on a "what" and "where" vision task and on a multi-payload robotics task are presented.

A Model of Distributed Sensorimotor Control in the Cockroach Escape Turn

R.D. Beer, G. Kacmarcik, R.E. Ritzmann, H.J. Chiel

In response to a puff of wind, the American cockroach turns away and runs. The circuit underlying the initial turn of this escape response consists of three populations of individually identifiable nerve cells and appears to employ distributed representations in its operation. We have reconstructed several neuronal and behavioral properties of this system using simplified neural network models and the backpropagation learning algorithm constrained by known structural characteristics of the circuitry. In order to test and refine the model, we have also compared the model's responses to various lesions with the insect's responses to similar lesions.

A VLSI Neural Network for Color Constancy

Andrew Moore, John Allman, Geoffrey Fox, Rodney Goodman

A system for color correction has been designed, built, and tested successfully; the essential components are three custom chips built using submicron threshold analog CMOS VLSI. The system, based on Land's Retinex theory of color constancy, produces colors similar in many respects to those produced by the visual system. Resistive grids implemented in analog VLSI perform the smoothing operation central to the algorithm at video rates. With the electronic system, the strengths and weaknesses of the algorithm are explored.

Stereopsis by a Neural Network Which Learns the Constraints

Alireza Khotanzad, Ying-Wung Lee

This paper presents a neural network (NN) approach to the problem of stereopsis. The correspondence problem (finding the correct matches between the pixels of

the epipolar lines of the stereo pair from amongst all the possible matches) is posed as a non-iterative many-to-one mapping. A two-layer feed forward NN architecture is developed to learn and code this nonlinear and complex mapping using the back-propagation learning rule and a training set. The important aspect of this technique is that none of the typical constraints such as uniqueness and continuity are explicitly imposed. All the applicable constraints are learned and internally coded by the NN enabling it to be more flexible and more accurate than the existing methods. The approach is successfully tested on several random dot stereograms. It is shown that the net can generalize its learned mapping to cases outside its training set. Advantages over the Marr-Poggio Algorithm are discussed and it is shown that the NN performance is superior.

The Tempo 2 Algorithm: Adjusting Time-Delays By Supervised Learning

Ulrich Bodenhausen, Alex Waibel

In this work we describe a new method that adjusts time-delays and the widths of time-windows in artificial neural networks automatically. The input of the units are weighted by a gaussian input-window over time which allows the learning rules for the delays and widths to be derived in the same way as it is used for the weights. Our results on a phoneme classification task compare well with results obtained with the TDNN by Waibel et al., which was manually optimized for the same task.

Learning Theory and Experiments with Competitive Networks

Griff Bilbro, David van den Bout

We apply the theory of Tishby, Levin, and Solla (TLS) to two problems. First we analyze an elementary problem for which we find the predictions consistent with conventional statistical results. Second we numerically examine the more realistic problem of training a competitive net to learn a probability density from samples. We find TLS useful for predicting average training behavior.

Further Studies of a Model for the Development and Regeneration of Eye-Brain Maps

Jack Cowan, A. Friedman

We describe a computational model of the development and regeneration of specific eye-brain circuits. The model comprises a self-organizing map-forming network which uses local Hebb rules, constrained by (genetically determined) molecular markers. Various simulations of the development and regeneration of eye-brain maps in fish and frogs are described, in particular successful simulations of experiments by Schmidt-Cicerone-Easter; Meyer; and Yoon.

The Recurrent Cascade-Correlation Architecture

Scott Fahlman

Recurrent Cascade-Correlation (RCCC) is a recurrent version of the Cascade Correlation learning architecture of Fahlman and Lebiere [Fahlman, 1990]. RCCC can learn from examples to map a sequence of inputs into a desired sequence of outputs. New hidden units with recurrent connections are added to the network as needed during training. In effect, the network builds up a finite-state machine tailored specifically for the current problem. RCCC retains the advantages of Cascade-Correlation: fast learning, good generalization, automatic construction of a near-minimal multi-layered network, and incremental training.

Sequential Adaptation of Radial Basis Function Neural Networks and its Application to Time-series Prediction

V. Kadiramanathan, M. Niranjana, F. Fallside

We develop a sequential adaptation algorithm for radial basis function (RBF) neural networks of Gaussian nodes, based on the method of successive F-P projections. This method makes use of each observation efficiently in that the n

etwork mapping function so obtained is consistent with that information and is also optimal in the least L_2 -norm sense. The RBF network with the F-Projection adaptation algorithm was used for predicting a chaotic time-series. We compare its performance to an adaptation scheme based on the method of stochastic approximation, and show that the F-Projections algorithm converges to the underlying model much faster.
