Comparing the Performance of Connectionist and Statistical Classifiers on an Ima
ge Segmentation Problem
Sheri Gish, W. Blanz
In this study, we test the suitability of a connection(cid:173)
****

Analog Circuits for Constrained Optimization
John Platt
This paper explores whether analog circuitry can adequately per(cid:173)
form constrained optimization. Constrained optimization circuits are desig
ned using the differential multiplier method. These cir(cid:173)cuits fu
lfill time-varying constraints correctly. Example circuits in(cid:173)clude
a quadratic programming circuit and a constrained flip-flop.
****

Training Stochastic Model Recognition Algorithms as Networks can Lead to Maximum
Mutual Information Estimation of Parameters
John Bridle
One of the attractions of neural network approaches to pattern recogniti
on is the use of a discrimination-based training method. We show that
once we have modified the output layer of a multi(cid:173)layer percept
ron to provide mathematically correct probability dis(cid:173)tributions, and
replaced the usual squared error criterion with a probability-based sc
ore, the result is equivalent to Maximum Mu(cid:173)tual Information tra
ining, which has been used successfully to im(cid:173)prove the perform
ance of hidden Markov models for speech recog(cid:173)nition. If the netwo
rk is specially constructed to perform the recog(cid:173)nition computations of
a given kind of stochastic model based clas(cid:173)sifier then we obtain a
method for discrimination-based training of the parameters of the models.
Examples include an HMM-based word discriminator, which we call an 'Alp
hanet'.
****

Can Simple Cells Learn Curves? A Hebbian Model in a Structured Environment
William Softky, Daniel Kammen
In the mammalian visual cortex, orientation-selective 'simple cells' which det
ect straight lines may be adapted to detect curved lines instead. We tes
t a biologically plausible, Hebbian, single-neuron model, which learns or
iented receptive fields upon exposure to un(cid:173)structured (noise) input
and maintains orientation selectivity upon exposure to edges or bars of al
l orientations and positions. This model can also learn arc-shaped rec
eptive fields upon exposure to an environment of only circular rings. Thu
s, new experiments which try to induce an abnormal (curved) receptive field
may pro(cid:173)vide insight into the plasticity of simple cells. The model su
ggests that exposing cells to only a single spatial frequency may induce
more striking spatial frequency and orientation dependent effects than he
retofore observed.
****

Development and Regeneration of Eye-Brain Maps: A Computational Model
Jack Cowan, A. Friedman
We outline a computational model of the development and regenera(cid:173)tion
of specific eye-brain circuits. The model comprises a self-organiz(cid:173)ing
map-forming network which uses local Hebb rules. constrained by molecular marke
rs. Various simulations of the development of eye(cid:173)brain maps in fish
and frogs are described.
****

Predicting Weather Using a Genetic Memory: A Combination of Kanerva's Sparse Dis
tributed Memory with Holland's Genetic Algorithms
David Rogers
Kanerva's sparse distributed memory (SDM) is an associative-memo(cid:173)
ry model based on the mathematical properties of high-dimensional binar
y address spaces. Holland's genetic algorithms are a search tech(cid:173)
nique for high-dimensional spaces inspired by evolutionary processes of

DNA. "Genetic Memory" is a hybrid of the above two systems, in which the memory uses a genetic algorithm to dynamically recon(cid:173) figure its physical storage locations to reflect correlations between the stored addresses and data. For example, when presented with raw weather station data, the Genetic Memory discovers specific fea(cid:173) tures in the weather data which correlate well with upcoming rain, and reconfigures the memory to utilize this information effectively. This architecture is designed to maximize the ability of the system to scale-up to handle real-world problems.

****

Neural Network Analysis of Distributed Representations of Dynamical Sensory-Motor Transformations in the Leech

Shawn Lockery, Yan Fang, Terrence J. Sejnowski

Interneurons in leech ganglia receive multiple sensory inputs and make synaptic contacts with many motor neurons. These "hidden" units coordinate several different behaviors. We used physiological and anatomical constraints to construct a model of the local bending reflex. Dynamical networks were trained on experimentally derived input-output patterns using recurrent back-propagation. Units in the model were modified to include electrical synapses and multiple synaptic time constants. The properties of the hidden units that emerged in the simulations matched those in the leech. The model and data support distributed rather than localist representations in the local bending reflex. These results also explain counterintuitive aspects of the local bending circuitry.

****

A Computer Modeling Approach to Understanding the Inferior Olive and Its Relationships to the Cerebellar Cortex in Rats

Maurice Lee, James Bower

This paper presents the results of a simulation of the spatial relationship between the inferior olivary nucleus and folium crus IIA of the lateral hemisphere of the rat cerebellum. The principal objective of this modeling effort was to resolve an apparent conflict between a proposed zonal organization of olivary projections to cerebellar cortex suggested by anatomical tract-tracing experiments (Brodal & Kawamura 1980; Campbell & Armstrong 1983) and a more patchy organization apparent with physiological mapping (Robertson 1987). The results suggest that several unique features of the olivocerebellar circuit may contribute to the appearance of zonal organization using anatomical techniques, but that the detailed patterns of patchy tactile projections seen with physiological techniques are a more accurate representation of the afferent organization of this region of cortex.

****

Generalization and Scaling in Reinforcement Learning

David Ackley, Michael Littman

In associative reinforcement learning, an environment generates input vectors, a learning system generates possible output vectors, and a re(cid:173) inforcement function computes feedback signals from the input-output pairs. The task is to discover and remember input-output pairs that generate rewards. Especially difficult cases occur when rewards are rare, since the expected time for any algorithm can grow exponentially with the size of the problem. Nonetheless, if a reinforcement function possesses regularities, and a learning algorithm exploits them, learning time can be reduced below that of non-generalizing algorithms. This paper describes a neural network algorithm called complementary re(cid:173) inforcement back-propagation (CRBP), and reports simulation results on problems designed to offer differing opportunities for generalization.

****

Neural Network Weight Matrix Synthesis Using Optimal Control Techniques

O. Farotimi, Amir Dembo, Thomas Kailath
T. Kailath

****

Collective Oscillations in the Visual Cortex

Daniel Kammen, Christof Koch, Philip Holmes
The firing patterns of populations of cells in the cat visual cor(cid:173) tex can exhibit oscillatory responses in the range of 35 - 85 Hz. Furthermore, groups of neurons many mm's apart can be highly synchroniz ed as long as the cells have similar orientation tuning. We investigat e two basic network architectures that incorporate ei(cid:173) ther nearest-neig hbor or global feedback interactions and conclude that non-local feedback plays a fundamental role in the initial syn(cid:173) chronization and dynamic stabil ity of the oscillations.
****

Neural Networks: The Early Days
Jack Cowan
A short account is given of various investigations of neural network propert ies, beginning with the classic work of McCulloch & Pitts. Early work on neurodynamics and statistical mechanics, analogies with magnetic materials, fau lt tolerance via parallel distributed processing, memory, learning, and patter n recognition, is described.
****

Operational Fault Tolerance of CMAC Networks
Michael Carter, Franklin Rudolph, Adam Nucci
The performance sensitivity of Albus' CMAC network was studied for the scenario in which faults are introduced into the adjustable weights after training has been accomplished. It was found that fault sensitivity was reduced with increa sed generalization when "loss of weight" faults were considered, but sensitivi ty was increased for "saturated weight" faults.
****

Effects of Firing Synchrony on Signal Propagation in Layered Networks
G. Kenyon, Eberhard Fetz, R. Puff
Spiking neurons which integrate to threshold and fire were used to st udy the transmission of frequency modulated (FM) signals through layere d networks. Firing correlations between cells in the input layer were fo und to modulate the transmission of FM sig(cid:173) nals under certain dynamical conditions. A tonic level of activity was maintained by provi ding each cell with a source of Poisson(cid:173) distributed synaptic in put. When the average membrane depo(cid:173) larization produced by the synaptic input was sufficiently below threshold, the firing correlation s between cells in the input layer could greatly amplify the signal prese nt in subsequent layers. When the depolarization was sufficiently close to t hreshold, however, the firing synchrony between cells in the initial laye rs could no longer effect the propagation of FM signals. In this latter ca se, integrate(cid:173) and-fire neurons could be effectively modeled by simpler analog elements governed by a linear input-output relation.
****

Acoustic-Imaging Computations by Echolocating Bats: Unification of Diversely-Rep resented Stimulus Features into Whole Images
James Simmons
The echolocating bat, Eptesicus fuscus, perceives the distance to sonar targets from the delay of echoes and the shape of targets from the s pectrum of echoes. However, shape is perceived in terms of the target's range proftle. The time separation of echo components from parts of the target located at different distances is reconstructed from the echo sp ectrum and added to the estimate of absolute delay already derived fro m the arrival-time of echoes. The bat thus perceives the distance to targets and range depth within dimension, which is computed. The imag e corresponds to the crosscorrelation function of echoes. Fusion of ph ysiologically distinct time- and frequency-domain representations into a fmal, common time-domain image illustrates the binding of within(cid:173) modality features into a unified, whole image. To support the struct ure of images along the dimension of range, bats can perceive echo de lay with a hyperacuity of 10 nanoseconds.

****

Time Dependent Adaptive Neural Networks
Fernando Pineda

A comparison of algorithms that minimize error functions to train the trajectories of recurrent networks, reveals how complexity is traded off for causality. These algorithms are also related to time-independent fonnalisms. It is suggested that causal and scalable algorithms are possible when the activation dynamics of adaptive neurons is fast compared to the behavior to be learned. Standard continuous-time recurrent backpropagation is used in an example.

****

Neural Network Visualization
Jakub Wejchert, Gerald Tesauro

We have developed graphics to visualize static and dynamic infor(cid:173)mation in layered neural network learning systems. Emphasis was placed on creating new visuals that make use of spatial arrange(cid:173)ments, size information, animation and color. We applied these tools to the study of back-propagation learning of simple Boolean predicates, and have obtained new insights into the dynamics of the learning process.

****

Discovering the Structure of a Reactive Environment by Exploration
Michael C. Mozer, Jonathan Bachrach

Consider a robot wandering around an unfamiliar environment. performing ac(cid:173)tions and sensing the resulting environmental states. The robot's task is to con(cid:173)struct an internal model of its environment. a model that will allow it to predict the consequences of its actions and to determine what sequences of actions to take to reach particular goal states. Rivest and Schapire (1987&, 1987b; Schapire. 1988) have studied this problem and have designed a symbolic algo(cid:173)rithm to strategically explore and infer the structure of "finite state" environ(cid:173)ments. The heart of this algorithm is a clever representation of the environment called an update graph. We have developed a connectionist implementation of the update graph using a highly-specialized network architecture. With back propagation learning and a trivial exploration strategy - tions - gorithm on simple problems. The network has the additional strength that it can accommodate stochastic environments. Perhaps the greatest virtue of the connectionist approach is that it suggests generalizations of the update graph representation that do not arise from a traditional, symbolic perspective.

****

A Self-organizing Associative Memory System for Control Applications
Michael Hormel

The CHAC storage scheme has been used as a basis for a software implementation of an associative .emory system AHS, which itself is a major part of the learning control loop LERNAS. A major this CHAC-concept is that the disadvantage of degree of local generalization (area of interpo(cid:173)lation) is fixed. This paper deals with an algo(cid:173)rithm for self-organizing variable generaliza(cid:173)tion for the AKS, based on ideas of T. Kohonen.

****

Complexity of Finite Precision Neural Network Classifier
Amir Dembo, Kai-Yeung Siu, Thomas Kailath

A rigorous analysis on the finite precision computational <)Spects of neural network as a pattern classifier via a probabilistic approach is presented. Even though there exist negative results on the capa(cid:173)bility of perceptron, we show the following positive results: Given n pattern vectors each represented by en bits where e > 1, that are uniformly distributed, with high probability the perceptron can perform all possible binary classifications of the patterns. More(cid:173)over, the resulting neural network requires a vanishingly small pro(cid:173)portion O(log n/n) of the memory that would be required for com(cid:173)plete storage of the patterns. Furth

er, the perceptron algorithm takes O(n2) arithmetic operations with high probability, whereas other methods such as linear programming takes O (n3 .5) in the worst case. We also indicate some mathematical connections with VLSI circuit testing and the theory of random matrices.
****

Neuronal Group Selection Theory: A Grounding in Robotics
Jim Donnett, Tim Smithers
In this paper, we discuss a current attempt at applying the organi(cid:173) zational principle Edelman calls Neuronal Group Selection to the control of a real, two-link robotic manipulator. We begin by moti(cid:173) vating the need for an alternative to the position-control paradigm of classical robotics, and suggest that a possible avenue is to look at the primitive animal limb 'neurologically ballistic' control mode. We have been considering a selectionist approach to coordinating a simple perception-action task.
****

Dimensionality Reduction and Prior Knowledge in E-Set Recognition
Kevin Lang, Geoffrey E. Hinton
It is well known that when an automatic learning algorithm is applied to a fixed corpus of data, the size of the corpus places an upper bound on the number of degrees of freedom that the model can contain if it is to generalize well. Because the amount of hardware in a neural network typically increases with the dimensionality of its inputs, it can be challenging to build a high-performance network for classifying large input patterns. In this paper, several techniques for addressing this problem are discussed in the context of an isolated word recognition task.
****

A self-organizing multiple-view representation of 3D objects
Daphna Weinshall, Shimon Edelman, Heinrich Bülthoff
We demonstrate the ability of a two-layer network of thresholded summation units to support representation of 3D objects in which several distinct 2D views are stored for ea.ch object. Using unsu(cid:173) pervised Hebbian relaxation, the network learned to recognize ten objects from different viewpoints. The training process led to the emergence of compact representations of the specific input views. When tested on novel views of the same objects, the network ex(cid:173) hibited a substantial generalization capability. In simulated psy(cid:173) chophysical experiments, the network's behavior was qualitatively similar to that of human subjects.
****

Associative Memory in a Simple Model of Oscillating Cortex
Bill Baird
A generic model of oscillating cortex, which assumes "minimal" coupling justified by known anatomy, is shown to function as an as(cid:173) sociative memory, using previously developed theory. The network has explicit excitatory neurons with local inhibitory interneuron feedback that forms a set of nonlinear oscillators coupled only by long range excitatofy connections. Using a local Hebb-like learning rule for primary and higher order synapses at the ends of the long range connections, the system learns to store the kinds of oscil(cid:173) lation amplitude patterns observed in olfactory and visual cortex. This rule is derived from a more general "projection algorithm" for recurrent analog networks, that analytically guarantees content addressable memory storage of continuous periodic sequences - capacity: N /2 Fourier components for an N node network - "spurious" attractors.
****

Adjoint Operator Algorithms for Faster Learning in Dynamical Neural Networks
Jacob Barhen, Nikzad Toomarian, Sandeep Gulati
A methodology for faster supervised learning in dynamical nonlin(cid:173) ear neural networks is presented. It exploits the concept of adjoint operntors to enable computation of changes in the network's re(cid:173) sponse due to perturbations in all system parameters, using the so(cid:173) lution of a single set of appropriately constructed linear equations. The lower bound on

speedup per learning iteration over conven(cid:173) tional methods for c
alculating the neuromorphic energy gradient is O(N2), where N is the number
of neurons in the network.
****

Computer Simulation of Oscillatory Behavior in Cerebral Cortical Networks
Matthew Wilson, James Bower
It has been known for many years that specific regions of the work(cid:173) ing
cerebral cortex display periodic variations in correlated cellular activity.
While the olfactory system has been the focus of much of this work, similar
behavior has recently been observed in primary visual cortex. We have
developed models of both the olfactory and visual cortex which replic
ate the observed oscillatory proper(cid:173) ties of these networks. Using
these models we have examined the dependence of oscillatory behavior on s
ingle cell properties and net(cid:173) work architectures. We discuss the ide
a that the oscillatory events recorded from cerebral cortex may be intrinsi
c to the architecture of cerebral cortex as a whole, and that these rh
ythmic patterns may be important in coordinating neuronal activity during sens
ory processmg.
****

Coupled Markov Random Fields and Mean Field Theory
Davi Geiger, Federico Girosi
Federico Girosi Artificial Intelligence Laboratory, MIT 545 Tech. Sq. #
788 Cambridge, MA 02139
****

Pulse-Firing Neural Chips for Hundreds of Neurons
Michael Brownlow, Lionel Tarassenko, Alan Murray, Alister Hamilton, Il Han, H. R
eekie
We announce new CMOS synapse circuits using only three and four MOSFET
sisynapse. Neural states are asynchronous pulse streams, upon which arithm
etic is performed directly. Chips implementing over 100 fully programma
ble synapses are described and projections to networks of hundreds of
neurons are made.
****

Rule Representations in a Connectionist Chunker
David Touretzky, Gillette Elvgreen
We present two connectionist architectures for chunking of symbolic rewr
ite rules. One uses backpropagation learning, the other competitive learning.
Although they were developed for chunking the same sorts of rules, th
e two differ in their representational abilities and learning behaviors
.
****

A Neural Network for Real-Time Signal Processing
Donald Malkoff
This paper describes a neural network algorithm that (1) performs temporal pa
ttern matching in real-time, (2) is trained on-line, with a single pass, (3)
requires only a single template for training of each representative class, (
4) is continuously adaptable to changes in background noise, (5) deals wi
th transient signals having low signal(cid:173) to-noise ratios, (6) works in t
he presence of non-Gaussian noise, (7) makes use of context dependencies and (
8) outputs Bayesian proba(cid:173) bility estimates. The algorithm has been ada
pted to the problem of passive sonar signal detection and classification. It r
uns on a Con(cid:173) nection Machine and correctly classifies, within 50
0 ms of onset, signals embedded in noise and subject to considerable uncertai
nty.
****

Speaker Independent Speech Recognition with Neural Networks and Speech Knowledge
Yoshua Bengio, Renato de Mori, Régis Cardin
We attempt to combine neural networks with knowledge from speech science
to build a speaker independent speech recogni(cid:173) tion system. This kno
wledge is utilized in designing the preprocessing, input coding, output c

oding, output supervision and architectural constraints. To handle the t emporal aspect of speech we combine delays, copies of activations of hidd en and output units at the input level, and Back-Propagation for Sequ ences (BPS), a learning algorithm for networks with local self-loops. This strategy is demonstrated in several experi(cid:173) ments, in particula r a nasal discrimination task for which the application of a speech theory hypothesis dramatically im(cid:173) proved generalization.
****

Analytic Solutions to the Formation of Feature-Analysing Cells of a Three-Layer Feedforward Visual Information Processing Neural Net
Dun-Sung Tang
Analytic solutions to the information-theoretic evolution equa(cid:173) ti on of the connection strength of a three-layer feedforward neural net for vis ual information processing are presented. The results are (1) the recepti ve fields of the feature-analysing cells corre(cid:173) spond to the eigenv ector of the maximum eigenvalue of the Fred(cid:173) holm integral equation of t he first kind derived from the evolution equation of the connection streng th; (2) a symmetry-breaking mechanism (parity-violation) has been identi fied to be respon(cid:173) sible for the changes of the morphology of the receptive field; (3) the conditions for the formation of different mo rphologies are explicitly identified.
****

An Analog VLSI Model of Adaptation in the Vestibulo-Ocular Reflex
Stephen DeWeerth, Carver Mead
The vestibulo-ocular reflex (VOR) is the primary mechanism that controls the compensatory eye movements that stabilize retinal im(cid:173) ages during rapid head motion. The primary pathways of this sys(cid:173) tem are feed-forward, w ith inputs from the semicircular canals and outputs to the oculomotor syste m. Since visual feedback is not used directly in the VOR computation, the system must exploit motor learning to perform correctly. Lisberger(19 88) has proposed a model for adapting the VOR gain using image-slip infor mation from the retina. We have designed and tested analog very large (cid:173) scale integrated (VLSI) circuitry that implements a simplified ver(ci d:173) sion of Lisberger's adaptive VOR model.
****

A Reconfigurable Analog VLSI Neural Network Chip
Srinagesh Satyanarayana, Yannis Tsividis, Hans Graf
1024 distributed-neuron synapses have been integrated in an active area of 6.1m m x 3.3mm using a 0.9p.m, double-metal, single-poly, n-well CMOS technolo gy. The distributed-neuron synapses are ar(cid:173) ranged in blocks of 16, whi ch we call '4 x 4 tiles'. Switch matrices are interleaved between each of t hese tiles to provide programma(cid:173) bility of interconnections. With a s mall area overhead (15 %), the 1024 units of the network can be rearrange d in various configura(cid:173) tions. Some of the possible configurations ar e, a 12-32-12 network, a 16-12-12-16 network, two 12-32 networks etc. (th e numbers sep(cid:173) arated by dashes indicate the number of units per layer, including the input layer). Weights are stored in analog form on MaS ca(cid:173) pacitors. The synaptic weights are usable to a resolution of 1 % of their full scale value. The limitation arises due to charge injecti on from the access switch and charge leakage. Other parameters like gain and shape of nonlinearity are also programmable.
****

Handwritten Digit Recognition with a Back-Propagation Network
Yann LeCun, Bernhard Boser, John Denker, Donnie Henderson, R. Howard, Wayne Hubb ard, Lawrence Jackel
We present an application of back-propagation networks to hand(cid:173) written digit recognition. Minimal preprocessing of the data was required, but architec ture of the network was highly constrained and specifically designed for the ta sk. The input of the network consists of normalized images of isolated digits. The method has 1 % error rate and about a 9% reject rate on zipcode digits prov

ided by the U.S. Postal Service.
****

Digital-Analog Hybrid Synapse Chips for Electronic Neural Networks
Alexander Moopenn, T. Duong, A. Thakoor
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.
****

Computational Efficiency: A Common Organizing Principle for Parallel Computer Maps and Brain Maps?
Mark Nelson, James Bower
It is well-known that neural responses in particular brain regions are spatially organized, but no general principles have been de(cid:173)veloped that relate the structure of a brain map to the nature of the associated computation. On parallel computers, maps of a sort quite similar to brain maps arise when a computation is distributed across multiple processors. In this paper we will discuss the rela(cid:173)tionship between maps and computations on these computers and suggest how similar considerations might also apply to maps in the brain.
****

A Cost Function for Internal Representations
Anders Krogh, C. Thorbergsson, John Hertz
We introduce a cost function for learning in feed-forward neural networks which is an explicit function of the internal representa(cid:173)tion in addition to the weights. The learning problem can then be formulated as two simple perceptrons and a search for internal representations. Back-propagation is recovered as a limit. The frequency of successful solutions is better for this algorithm than for back-propagation when weights and hidden units are updated on the same timescale i.e. once every learning step.
****

HMM Speech Recognition with Neural Net Discrimination
William Huang, Richard P. Lippmann
Two approaches were explored which integrate neural net classifiers with Hidden Markov Model (HMM) speech recognizers. Both at(cid:173)tempt to improve speech pattern discrimination while retaining the temporal processing advantages of HMMs. One approach used neu(cid:173)ral nets to provide second-stage discrimination following an HMM recognizer. On a small vocabulary task, Radial Basis Function (RBF) and back-propagation neural nets reduced the error rate substantially (from 7.9% to 4.2% for the RBF classifier). In a larger vocabulary task, neural net classifiers did not reduce the error rate. They, however, outperformed Gaussian, Gaussian mixture, and k(cid:173)nearest neighbor (KNN) classifiers. In another approach, neural nets functioned as low-level acoustic-phonetic feature extractors. When classifying phonemes based on single 10 msec. frames, dis(cid:173)criminant RBF neural net classifiers outperformed Gaussian mix(cid:173)ture classifiers. Performance, however, differed little when classi(cid:173)fying phones by accumulating scores across all frames in phonetic segments using a single node HMM recognizer.
****

The Effects of Circuit Integration on a Feature Map Vector Quantizer
Jim Mann
The effects of parameter modifications imposed by hardware con(cid:173)straints on a self-organizing feature map algorithm were examined. Performance was measured by the error rate of a speech recogni(cid:173)tion system which included this algorithm as part of the front-end processing. System parameters which were varied included weight (connection strength) quantization, adap tation quantization, dis(cid:173)tance measures and circuit approximations which include device characteristics and process vari

ability. Experiments using the TI isolated word database for 16 speakers d emonstrated degradation in performance when weight quantization fell below 8 bits. The com(cid:173) petitive nature of the algorithm rela..xes constrain ts on uniformity and linearity which makes it an excellent candidate for a fully ana(cid:173) log circuit implementation. Prototype circuits have been f abricated and characterized following the constraints established through the simulation efforts.

****

Generalization and Parameter Estimation in Feedforward Nets: Some Experiments
N. Morgan, H. Bourlard

We have done an empirical study of the relation of the number of parameters (we ights) in a feedforward net to generalization perfor(cid:173) mance. Two experim ents are reported. In one, we use simulated data sets with well-controlled para meters, such as the signal-to-noise ratio of continuous-valued data. In the sec ond, we train the network on vector-quantized mel cepstra from real speech samp les. In each case, we use back-propagation to train the feedforward net to disc riminate in a multiple class pattern classification problem. We report the resu lts of these studies, and show the application of cross-validation techniques to prevent overfitting.

****

VLSI Implementation of a High-Capacity Neural Network Associative Memory
Tzi-Dar Chiueh, Rodney Goodman

In this paper we describe the VLSI design and testing of a high capaci ty associative memory which we call the exponential cor(cid:173) relatio n associative memory (ECAM). The prototype 3J.'-CMOS programmable chip is capable of storing 32 memory patterns of 24 bits each. The high capac ity of the ECAM is partly due to the use of special exponentiation neurons, w hich are implemented via sub-threshold MOS transistors in this design. The p rototype chip is capable of performing one associative recall in 3 J.'S.

****

The Cascade-Correlation Learning Architecture
Scott Fahlman, Christian Lebiere

Cascade-Correlation is a new architecture and supervised learning algo(cid:173) rithm for artificial neural networks. Instead of just adjusting the weights in a network of fixed topology. Cascade-Correlation begins with a min(cid:173) ima l network, then automatically trains and adds new hidden units one by one, creating a multi-layer structure. Once a new hidden unit has been adde d to the network, its input-side weights are frozen. This unit then becomes a permanent feature-detector in the network, available for producing outputs or for creating other, more complex feature detec(cid:173) tors. The Casca de-Correlation architecture has several advantages over existing algorithms: it learns very quickly, the network . determines its own size and topology , it retains the structures it has built even if the training set changes, and it requires no back-propagation of error signals through the connections of the network.

****

Sigma-Pi Learning: On Radial Basis Functions and Cortical Associative Learning
Bartlett Mel, Christof Koch

The goal in this work has been to identify the neuronal elements of the cortical column that are most likely to support the learning of nonlinear associative maps. We show that a particular style of network learning algo rithm based on locally-tuned receptive fields maps naturally onto cortical hardware, and gives coherence to a variety of features of cortical anatom y, physiology, and biophysics whose relations to learning remain poorly und erstood.

****

Optimal Brain Damage
Yann LeCun, John Denker, Sara Solla

We have used information-theoretic ideas to derive a class of prac(cid:173) tical and nearly optimal schemes for adapting the size of a neural netwo

rk. By removing unimportant weights from a network, sev(cid:173) eral i mprovements can be expected: better generalization, fewer training exampl es required, and improved speed of learning and/or classification. The basi c idea is to use second-derivative informa(cid:173) tion to make a tradeo ff between network complexity and training set error. Experiments confirm the usefulness of the methods on a real-world application.

****

Learning Aspect Graph Representations from View Sequences

Michael Seibert, Allen Waxman

In our effort to develop a modular neural system for invariant learn(cid:173) in g and recognition of 3D objects, we introduce here a new module architectur e called an aspect network constructed around adaptive axo-axo-dendritic synapses. This builds upon our existing system (Seibert & Waxman, 1989) whi ch processes 20 shapes and classifies t.hem into view categories (i.e., aspects) invariant to illumination, position, orientat.ion, scale, and p rojective deformations. From a sequence 'of views, the aspect network l earns the transitions be(cid:173) tween these aspects, crystallizing a g raph-like structure from an initially amorphous network . Object recogni tion emerges by ac(cid:173) cumulating evidence over multiple views which activate competing object hypotheses.

****

The Effect of Catecholamines on Performance: From Unit to System Behavior

David Servan-Schreiber, Harry Printz, Jonathan D. Cohen

At the level of individual neurons. catecholamine release increases the respon sivity of cells to excitatory and inhibitory inputs. We present a model of catecholamine effects in a network of neural-like elements. We argue that changes in the responsivity of individual elements do not affect their ability to detect a signal and ignore noise. However. the same changes in cell responsivity in a network of such elements do improve the signa l detection performance of the network as a whole. We show how this result ca n be used in a computer simulation of behavior to account for the effect o f eNS stimulants on the signal detection performance of human subjects.

****

Meiosis Networks

Stephen Hanson

A central problem in connectionist modelling is the control of network and architectural resources during learning. In the present approach, weights reflect a coarse prediction history as coded by a distributio n of values and parameterized in the mean and standard deviation of these weight distributions. Weight updates are a function of both the mean and standard deviation of each connection in the network and vary as a function of the error signal ("stochastic delta rule"; Hanson, 1990). Consequently, the weights their maintain in "uncertainty" establi shing a policy concerning the size of the nodal complexity of the ne twork and growth of new nodes. For example, during problem solving th e present network can undergo "meiosis", producing two nodes where the re was one "overtaxed" node as measured by its coefficient of variation . It is shown in a number of benchmark problems that meiosis networks can find minimal architectures, reduce computational complexity, and ove rall increase the efficiency of the representation learning interaction.

****

Synergy of Clustering Multiple Back Propagation Networks

William Lincoln, Josef Skrzypek

The properties of a cluster of multiple back-propagation (BP) networks are exa mined and compared to the performance of a single BP net(cid:173) work. The underlying idea is that a synergistic effect within the cluster improves the perfonnance and fault tolerance. Five networks were ini(cid:173) tially tr ained to perfonn the same input-output mapping. Following training, a cl uster was created by computing an average of the outputs generated by the indiv idual networks. The output of the cluster can be used as the desired output du

ring training by feeding it back to the indi(cid:173) vidual networks. In comparison to a single BP network, a cluster of multiple BP's generalization and significant fault tolerance. It appear that cluster advantage follows from simple maxim "you can fool some of the single BP's in a cluster all of the time but you cannot fool all of them all of the time" {Lincoln}

****

Incremental Parsing by Modular Recurrent Connectionist Networks

Ajay Jain, Alex Waibel

We present a novel, modular, recurrent connectionist network architec(cid:173) ture which learns to robustly perform incremental parsing of complex sentences. From sequential input, one word at a time, our networks learn to do semantic role assignment, noun phrase attachment, and clause structure recognition for sentences with passive constructions and center embedded clauses. The networks make syntactic and semantic predictions at every point in time, and previous predictions are revised as expectations are affirmed or violated with the arrival of new informa(cid:173) tion. Our networks induce their own "grammar rules" for dynamically transforming an input sequence of words into a syntactic/semantic in(cid:173) terpretation. These networks generalize and display tolerance to input which has been corrupted in ways common in spoken language.

****

Reading a Neural Code

William Bialek, Fred Rieke, Robert van Steveninck, David Warland

Traditional methods of studying neural coding characterize the en(cid:173) coding of known stimuli in average neural responses. Organisms face nearly the opposite task - decoding short segments of a spike train to extract information about an unknown, time-varying stim(cid:173) ulus. Here we present strategies for characterizing the neural code from the point of view of the organism, culminating in algorithms for real-time stimulus reconstruction based on a single sample of the spike train. These methods are applied to the design and anal(cid:173) ysis of experiments on an identified movement-sensitive neuron in the fly visual system. As far as we know this is the first instance in which a direct "reading" of the neural code has been accomplished.

****

A Method for the Associative Storage of Analog Vectors

Amir Atiya, Yaser Abu-Mostafa

A method for storing analog vectors in Hopfield's continuous feed(cid:173) back model is proposed. By analog vectors we mean vectors whose components are real-valued. The vectors to be stored are set as equilibria of the network. The network model consists of one layer of visible neurons and one layer of hidden neurons. We propose a learning algorithm, which results in adjusting the positions of the equilibria, as well as guaranteeing their stability. Simulation results confirm the effectiveness of the method.

****

Using Local Models to Control Movement

Christopher Atkeson

This paper explores the use of a model neural network for motor learning. Steinbuch and Taylor presented neural network designs to do nearest neighbor lookup in the early 1960s. In this paper their nearest neighbor network is augmented with a local model network, which fits a local model to a set of nearest neighbors. The network design is equivalent to local regression. This network architecture can represent smooth nonlinear functions, yet has simple training rules with a single global optimum. The network has been used for motor learning of a simulated arm and a simulated running machine.

****

Learning to Control an Unstable System with Forward Modeling

Michael Jordan, Robert Jacobs

The forward modeling approach is a methodology for learning con(cid:173) trol when data is available in distal coordinate systems. We extend previous work by considering how this methodology can be applied to the optimization of quan tities that are distal not only in space but also in time.

****

Learning in Higher-Order "Artificial Dendritic Trees

Tony Bell

If neurons sum up their inputs in a non-linear way, as some simula(ci d:173) tions suggest, how is this distributed fine-grained non-linearity ex(cid:173) ploited during learning? How are all the small sigmoids in synapse, spine and dendritic tree lined up in the right areas of the ir respective input spaces? In this report, I show how an abstract atemp oral highly nested tree structure with a quadratic transfer function a ssociated with each branchpoint, can self organise using only a single global reinforcement scalar, to perform binary classification tasks. Th e pro(cid:173) cedure works well, solving the 6-multiplexer and a difficul t phoneme classification task as well as back-propagation does, and fas ter. Furthermore, it does not calculate an error gradient, but uses a statis t(cid:173) ical scheme to build moving models of the reinforcement signal.

****

Dynamic Behavior of Constained Back-Propagation Networks

Yves Chauvin

The learning dynamics of the back-propagation algorithm are in(cid:173) vestigat ed when complexity constraints are added to the standard Least Mean Square (LMS) cost function. It is shown that loss of generalization performance due to ove rtraining can be avoided when using such complexity constraints. Furthermore, " energy," hidden representations and weight distributions are observed and comp ared during learning. An attempt is made at explaining the results in terms of linear and non-linear effects in relation to the gradient descent learning algo rithm.

****

Designing Application-Specific Neural Networks Using the Genetic Algorithm

Steven Harp, Tariq Samad, Aloke Guha

We present a general and systematic method for neural network design based on the genetic algorithm. The technique works in conjunction wit h network learning rules, addressing aspects of the network's gross ar chitecture, connectivity, and learning rule parameters. Networks can be optimiled for various application(cid:173) specific criteria, such as lea rning speed, generalilation, robustness and connectivity. The approach is model-independent. We describe a prototype system, NeuroGENESYS, that em ploys the backpropagation learning rule. Experiments on several small p roblems have been conducted. In each case, NeuroGENESYS has produced n etworks that perform significantly better than the randomly generated netw orks of its initial population. The com(cid:173) putational feasibility of o ur approach is discussed.

****

A Computational Basis for Phonology

David Touretzky, Deirdre Wheeler

The phonological structure of human languages is intricate, yet highly cons trained. Through a combination of connectionist modeling and linguistic analysis, we are attempting to develop a computational basis for the nature of phonology. We present a connectionist architecture that performs mul tiple simultaneous insertion, deletion, and mutation operations on sequenc es of phonemes, and introduce a novel additional primitive, clustering. Clus tering provides an interesting alternative to both iterative and relaxatio n accounts of assimilation processes such as vowel harmony. Our resulting m odel is efficient because it processes utterances entirely in parallel using only feed-forward circuitry.

****

Neural Network Simulation of Somatosensory Representational Plasticity

Kamil Grajski, Michael Merzenich
The brain represents the skin surface as a topographic map in the so matosensory cortex. This map has been shown experimentally to be modif iable in a use-dependent fashion throughout life. We present a neural network simulation of the competitive dynamics underlying this cortical plasticity by detailed analysis of receptive field properties of mode l neurons during simulations of skin co(cid:173) activation, cortical les ion, digit amputation and nerve section.
****

A Neural Network for Feature Extraction
Nathan Intrator
The paper suggests a statistical framework for the parameter esti(cid:173) m ation problem associated with unsupervised learning in a neural network, leading to an exploratory projection pursuit network that performs feature ex traction, or dimensionality reduction.
****

Generalized Hopfield Networks and Nonlinear Optimization
Gintaras Reklaitis, Athanasios Tsirukis, Manoel Tenorio
A nonlinear neural framework, called the Generalized Hopfield network, is proposed, which is able to solve in a parallel distributed manner systems of nonlinear equations. The method is applied to the general non linear optimization problem. We demonstrate GHNs implementing the three most important optimization algorithms, namely the Augmented Lagrangian, Ge neralized Reduced Gradient and Successive Quadratic Programming methods. Th e study results in a dynamic view of the optimization problem and offers a straightforward model for the parallelization of the optimization compu tations, thus significantly extending the practical limits of problems t hat can be formulated as an optimization problem and which can gain from the introduction of nonlinearities in their structure (eg. pattern recogn ition, supervised learning, design of content-addressable memories).
****

Connectionist Architectures for Multi-Speaker Phoneme Recognition
John Hampshire, Alex Waibel
We present a number of Time-Delay Neural Network (TDNN) based architect ures for multi-speaker phoneme recognition (/b,d,g/ task). We use speech of two females and four males to compare the performance of the various archite ctures against a baseline recognition rate of 95.9% for a single IDNN on the si x-speaker /b,d,g/ task. This series of modu(cid:173) lar designs leads to a hig hly modular multi-network architecture capable of performing the six-speaker re cognition task at the speaker dependent rate of 98.4%. In addition to its high recognition rate, the so-called "Meta-Pi" architecture learns - wit hout direct supervision - ognize the speech of one particular male speaker usin g internal models of other male speakers exclusively.
****

Bayesian Inference of Regular Grammar and Markov Source Models
Kurt Smith, Michael Miller
In this paper we develop a Bayes criterion which includes the Rissanen complexi ty, for inferring regular grammar models. We develop two methods for regular grammar Bayesian inference. The fIrst method is based on treating the regu lar grammar as a I-dimensional Markov source, and the second is based on th e combinatoric characteristics of the regular grammar itself. We apply the res ulting Bayes criteria to a particular example in order to show the efficiency o f each method.
****

Dataflow Architectures: Flexible Platforms for Neural Network Simulation
Ira Smotroff
Dataflow architectures are general computation engines optimized for the execut ion of fme-grain parallel algorithms. Neural networks can be simulated on thes e systems with certain advantages. In this paper, we review dataflow archit ectures, examine neural network simulation performance on a new generat

ion dataflow machine, compare that performance to other simulation alternat
ives, and discuss the benefits and drawbacks of the dataflow approach.
****

The Perceptron Algorithm Is Fast for Non-Malicious Distributions
Eric Baum
Within the context of Valiant's protocol for learning, the Perceptron
algorithm is shown to learn an arbitrary half-space in time O(r;;) if D, the p
roba(cid:173)
bility distribution of examples, is taken uniform over the unit sphere sn. He
re f is
the accuracy parameter. This is surprisingly fast, as "standard" approaches
involve
solution of a linear programming problem involving O(7') constraints in n
dimen(cid:173)
sions. A modification of Valiant's distribution independent protocol for
learning
is proposed in which the distribution and the function to be learned may
be cho(cid:173)
sen by adversaries, however these adversaries may not communicate. It is
argued
that this definition is more reasonable and applicable to real world l
earning than
Valiant's. Under this definition, the Perceptron algorithm is shown to be
a distri(cid:173)
bution independent learning algorithm. In an appendix we show that, for un
iform
distributions, some classes of infinite V-C dimension including convex s
ets and a
class of nested differences of convex sets are learnable.
****

Discovering High Order Features with Mean Field Modules
Conrad Galland, Geoffrey E. Hinton
A new form of the deterministic Boltzmann machine (DBM) learn(cid:173) ing proc
edure is presented which can efficiently train network mod(cid:173) ules to di
scriminate between input vectors according to some cri(cid:173) terion. T
he new technique directly utilizes the free energy of these "mean field mod
ules" to represent the probability that the criterion is met, the free ene
rgy being readily manipulated by the learning procedure. Although convent
ional deterministic Boltzmann learn(cid:173) ing fails to extract the high
er order feature of shift at a network bottleneck, combining the new
mean field modules with the mu(cid:173) tual information objective function
rapidly produces modules that perfectly extract this important higher order
feature without direct external supervision.
****

Sequential Decision Problems and Neural Networks
A. G. Barto, R. S. Sutton, C. J. C. H. Watkins
Decision making tasks that involve delayed consequences are very common yet dif
ficult to address with supervised learning methods. If there is an accurate mod
el of the underlying dynamical system, then these tasks can be formulated as se
quential decision problems and solved by Dynamic Programming. This paper discus
ses rein(cid:173) forcement learning in terms of the sequential decision framewo
rk and shows how a learning algorithm similar to the one implemented by the Ad
aptive Critic Element used in the pole-balancer of Barto, Sutton, and Anderson
(1983), and further developed by Sutton (1984), fits into this framework. Adapt
ive neural networks can play significant roles as modules for approximating the
functions required for solving sequential decision problems.
****

Asymptotic Convergence of Backpropagation: Numerical Experiments
Subutai Ahmad, Gerald Tesauro, Yu He
Yu He Dept. of Physics Ohio State Univ. Columbus, OH 43212

****

Training Connectionist Networks with Queries and Selective Sampling
Les Atlas, David Cohn, Richard Ladner
"Selective sampling" is a form of directed search that can greatly increase the ability of a connectionist network to generalize accu(cid:173) rately. Based on information from previous batches of samples, a network may be trained on data selectively sampled from regions in the domain that are unknown. This is realizable in cases when the distribution is known, or when the cost of drawing points from the target distribution is negligible compared to the cost of label(cid:173) ing them with the proper classification. The approach is justified by its applicability to the problem of training a network for power system security analysis. The benefits of selective sampling are studied analytically, and the results are confirmed experimentally.
****

Unsupervised Learning in Neurodynamics Using the Phase Velocity Field Approach
Michail Zak, Nikzad Toomarian
A new concept for unsupervised learning based upon examples in(cid:173) troduced to the neural network is proposed. Each example is con(cid:173) sidered as an interpolation node of the velocity field in the phase space. The velocities at these nodes are selected such that all the streamlines converge to an attracting set imbedded in the subspace occupied by the cluster of examples. The synaptic interconnections are found from learning procedure providing selected field. The theory is illustrated by examples.
****

A Continuous Speech Recognition System Embedding MLP into HMM
Hervé Bourlard, Nelson Morgan
We are developing a phoneme based. speaker-dependent continuous speech recognition system embedding a Multilayer Perceptron (MLP) (Le .• a feed forward Artificial Neural Network). into a Hidden Markov Model (HMM) approach. In [Bourlard & Wellekens]. it was shown that MLPs were approximating Maximum a Posteriori (MAP) probabilities and could thus be embedded as an emission probability estimator in HMMs. By using contextual information from a sliding window on the input frames. we have been able to improve frame or phoneme clas(cid:173) sification performance over the corresponding performance for Simple Maximum Likelihood (ML) or even MAP probabilities that are esti(cid:173) mated without the benefit of context. However. recognition of words in continuous speech was not so simply improved by the use of an MLP. and several modifications of the original scheme were necessary for getting acceptable performance. It is shown here that word recogniti on performance for a simple discrete density HMM system appears to be somewhat better when MLP methods are used to estimate the emission probabilities.
****

Analysis of Linsker's Simulations of Hebbian Rules
David MacKay, Kenneth Miller
Linsker has reported the development of centre---surround receptive fields and oriented receptive fields in simulations of a Hebb-type equation in a linear network. The dynamics of the learning rule are analysed in terms of the eigenvectors of the covariance matrix of cell activities. Analytic and computational results for Linsker's covariance matrices, and some general theorems, lead to an expla(cid:173) nation of the emergence of centre---surround and certain oriented structures.
****

Analog Neural Networks of Limited Precision I: Computing with Multilinear Threshold Functions
Zoran Obradovic, Ian Parberry
Experimental evidence has shown analog neural networks to be ex(cid:173) ~mely fault-tolerant; in particular. their performance does not ap(cid:173) pear to be significantly impaired when precision is limited. Analog neurons with limited precision essentially compute k-ary weighted mul

tilinear threshold functions. which divide R" into k regions with k-l hyperplanes. The behaviour of k-ary neural networks is investi(cid:173) gated. There is no canonical set of threshold values for k>3. although they exist for binary and ternary neural networks. The weights can be made integers of only 0 «z +k) log (z +k » bits. where z is the number of processors. without increasing hardware or run(cid:173) ning time. The weights can be made ±1 while increasing running time by a constant multiple and hardware by a small polynomial in z and k. Binary neurons can be used if the running time is allowed to increase by a larger constant multiple and the hardware is allowed to increase by a slightly larger polynomial in z and k. Any symmetric k-ary function can be computed in constant depth and size o (n k- 1/(k-2)!). and any k-ary function can be computed in constant depth and size 0 (nk"). The alternating neural networks of Olafsson and Abu-Mostafa. and the quantized neural networks of Fleisher are closely related to this model.

****

A Systematic Study of the Input/Output Properties of a 2 Compartment Model Neuron With Active Membranes

Paul Rhodes

The input/output properties of a 2 compartment model neuron are systematically explored. Taken from the work of MacGregor (MacGregor, 1987), the model neuron compartments contain several active conductances, including a potassium conductance in the dendritic compartment driven by the accumulation of intradendritic calcium. Dynamics of the conductances and potentials are governed by a set of coupled first order differential equations which are integrated numerically. There are a set of 17 internal parameters to this model, specificying conductance rate constants, time constants, thresholds, etc.

****

Subgrouping Reduces Complexity and Speeds Up Learning in Recurrent Networks

David Zipser

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.

****

Neural Implementation of Motivated Behavior: Feeding in an Artificial Insect

Randall Beer, Hillel Chiel

Most complex behaviors appear to be governed by internal moti(cid:173) vational states or drives that modify an animal's responses to its environment. It is therefore of considerable interest to understand the neural basis of these motivational states. Drawing upon work on the neural basis of feeding in the marine mollusc Aplysia, we have developed a heterogeneous artificial neural network for con(cid:173) trolling the feeding behavior of a simulated insect. We demonstrate that feeding in this artificial insect shares many characteristics with the motivated behavior of natural animals.

****

A Neural Network to Detect Homologies in Proteins

Yoshua Bengio, Samy Bengio, Yannick Pouliot, Patrick Agin

In order to detect the presence and location of immunoglobu(cid:173) lin (Ig) domains from amino acid sequences we built a system based on a neural network with one hidden layer trained with back propagation. The program was designed to efficiently identify proteins exhibiting such domains, characterized by a few localized conserved regions and a low overall homology. When the National Biomedical Research Foundation (NBRF) NEW protein sequence database was scanned to evaluate the program's performance, we obtained very low rates of false negatives coupled with a moderate rate of false positives.

****

Higher Order Recurrent Networks and Grammatical Inference

C. Giles, Guo-Zheng Sun, Hsing-Hen Chen, Yee-Chun Lee, Dong Chen

A higher order single layer recursive network easily learns to simulat

e a deterministic finite state machine and recognize regular grammars. When an enhanced version of this neural net state machine is connected through a common error term to an external analog stack memory, the combination can be interpreted as a neural net pushdown automata. The neural net finite state machine is given the primitives, push and POP. and is able to read the top of the stack. Through a gradient descent learning rule derived from the common error function, the hybrid network learns to effectively use the stack actions to manipUlate the stack memory and to learn simple context(cid:173) free grammars. INTRODUCTION
****

Practical Characteristics of Neural Network and Conventional Pattern Classifiers on Artificial and Speech Problems
Yuchun Lee, Richard P. Lippmann
Eight neural net and conventional pattern classifiers (Bayesian(cid:173) unimodal Gaussian, k-nearest neighbor, standard back-propagation, adaptive-step size back-propagation, hypersphere, feature-map, learn(cid:173) ing vector quantizer, and binary decision tree) were implemented on a serial computer and compared using two speech recognition and two artificial tasks. Error rates were statistically equivalent on almost all tasks, but classifiers differed by orders of magnitude in memory requirements, training time, classification time, and ease of adaptivity. Nearest-neighbor classifiers trained rapidly but re(cid:173) quired the most memory. Tree classifiers provided rapid classifica(cid:173) tion but were complex to adapt. Back-propagation classifiers typ(cid:173) ically required long training times and had intermediate memory requirements. These results suggest that classifier selection should often depend more heavily on practical considerations concerning memory and computation resources, and restrictions on training and classification times than on error rate.
****

Contour-Map Encoding of Shape for Early Vision
Pentti Kanerva
Contour maps provide a general method for recognizing two-dimensional shapes. All but blank images give rise to such maps, and people are good at recognizing objects and shapes from them. The maps are encoded easily in long feature vectors that are suitable for recognition by an associative memory. These properties of contour maps suggest a role for them in early visual perception. The prevalence of direction-sensitive neurons in the visual cortex of mammals supports this view.
****

Maximum Likelihood Competitive Learning
Steven Nowlan
One popular class of unsupervised algorithms are competitive algo(cid:173) rithms. In the traditional view of competition, only one competitor, the winner, adapts for any given case. I propose to view compet(cid:173) itive adaptation as attempting to fit a blend of simple probability generators (such as gaussians) to a set of data-points. The maxi(cid:173) mum likelihood fit of a model of this type suggests a "softer" form of competition, in which all competitors adapt in proportion to the relative probability that the input came from each competitor. I investigate one application of the soft competitive model, place(cid:173) ment of radial basis function centers for function interpolation, and show that the soft model can give better performance with little additional computational cost.
****

Note on Development of Modularity in Simple Cortical Models
Alex Chernajvsky, John Moody
The existence of modularity in the organization of nervous systems (e.g. cortical columns and olfactory glomeruli) is well known. We show that localized activity patterns in a layer of cells, collective excitations, can induce the formation of modular structures in the anatomical connections via a Hebbian learning mechanism. The networks are spatially homogeneous before learning, but the spon

(cid:173) taneous emergence of localized collective excitations and subse(cid:17
3) quently modularity in the connection patterns breaks translational symmetry.
This spontaneous symmetry breaking phenomenon is similar to those which drive
pattern formation in reaction-diffusion systems. We have identified requirement
s on the patterns of lateral connections and on the gains of internal units whi
ch are essential for the development of modularity. These essential requirement
s will most likely remain operative when more complicated (and bi(cid:173) olog
ically realistic) models are considered.
****

Model Based Image Compression and Adaptive Data Representation by Interacting Fi
lter Banks
Toshiaki Okamoto, Mitsuo Kawato, Toshio Inui, Sei Miyake
introduced. Based on
****

On the Distribution of the Number of Local Minima of a Random Function on a Grap
h
Pierre Baldi, Yosef Rinott, Charles Stein
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-auth
ors prior to requesting a name change in the electronic proceedings.
****

Using a Translation-Invariant Neural Network to Diagnose Heart Arrhythmia
Susan Lee
Distinctive electrocardiogram (EeG) patterns are created when the heart is beat
ing normally and when a dangerous arrhythmia is present. Some devices which mon
itor the EeG and react to arrhythmias parameterize the ECG signal and make a di
agnosis based on the parameters. The author discusses the use of a neural netwo
rk to classify the EeG signals directly. without parameterization. The input to
such a network must be translation-invariant. since the distinctive features o
f the EeG may appear anywhere in an arbritrarily-chosen EeG segment. The input
must also be insensitive to the episode-to-episode and patient-to-patient vari
ability in the rhythm pattern.
****

Non-Boltzmann Dynamics in Networks of Spiking Neurons
Michael Crair, William Bialek
We study networks of spiking neurons in which spikes are fired as a Poisso
n process. The state of a cell is determined by the instan(cid:173) taneous
firing rate, and in the limit of high firing rates our model reduces to
that studied by Hopfield. We find that the inclusion of spiking resu
lts in several new features, such as a noise-induced asymmetry between "o
n" and "off" states of the cells and probabil(cid:173) ity currents which destro
y the usual description of network dynam(cid:173) ics in terms of energy sur
faces. Taking account of spikes also al(cid:173) lows us to calibrate networ
k parameters such as "synaptic weights" against experiments on real synapse
s. Realistic forms of the post synaptic response alters the network dyn
amics, which suggests a novel dynamical learning mechanism.
****

The "Moving Targets" Training Algorithm
Richard Rohwer
A simple method for training the dynamical behavior of a neu(cid:173)
ral network is derived. It is applicable to any training problem in
discrete-time networks with arbitrary feedback. The algorithm resembles back-
propagation in that an error function is minimized using a gradient-based m
ethod, but the optimization is carried out in the hidden part of state space e
ither instead of, or in addition to weight space. Computational results are p
resented for some simple dynamical training problems, one of which require
s response to a signal 100 time steps in the past.
****

Performance Comparisons Between Backpropagation Networks and Classification Tree

s on Three Real-World Applications
Les Atlas, Ronald Cole, Jerome Connor, Mohamed El-Sharkawi, Robert Marks, Yeshwant Muthusamy, Etienne Barnard
Multi-layer perceptrons and trained classification trees are two very different techniques which have recently become popular. Given enough data and time, both methods are capable of performing arbi(cid:173)trary non-linear classification. We first consider the important differences between multi-layer perceptrons and classification trees and conclude that there is not enough theoretical basis for the clear(cid:173)cut superiority of one technique over the other. For this reason, we performed a number of empirical tests on three real-world problems in power system load forecasting, power system security prediction, and speaker-independent vowel identification. In all cases, even for piecewise-linear trees, the multi-layer perceptron performed as well as or better than the trained classification trees.
****

An Efficient Implementation of the Back-propagation Algorithm on the Connection Machine CM-2
Xiru Zhang, Michael McKenna, Jill Mesirov, David Waltz
In this paper, we present a novel implementation of the widely used Back-propagation neural net learning algorithm on the Connection Machine CM-2 - a general purpose, massively parallel computer with a hypercube topology. This implementation runs at about 180 million interconnections per second (IPS) on a 64K processor CM- 2. The main interprocessor communication operation used is 2D nearest neighbor communication. The techniques developed here can be easily extended to implement other algorithms for layered neural nets on the CM-2, or on other massively parallel computers which have 2D or higher degree connections among their processors.
****

Algorithms for Better Representation and Faster Learning in Radial Basis Function Networks
Avijit Saha, James Keeler
in
****

The Cocktail Party Problem: Speech/Data Signal Separation Comparison between Backpropagation and SONN
John Kassebaum, Manoel Tenorio, Christoph Schaefers
This work introduces a new method called Self Organizing Neural Network (SONN) algorithm and compares its performance with Back Propagation in a signal separation application. The problem is to separate two signals; a modem data signal and a male speech signal, added and transmitted through a 4 khz channel. The signals are sam(cid:173)pled at 8 khz, and using supervised learning, an attempt is made to reconstruct them. The SONN is an algorithm that constructs its own network topology during training, which is shown to be much smaller than the BP network, faster to trained, and free from the trial-and(cid:173)error network design that characterize BP.
****

Real-Time Computer Vision and Robotics Using Analog VLSI Circuits
Christof Koch, Wyeth Bair, John Harris, Timothy Horiuchi, Andrew Hsu, Jin Luo
The long-term goal of our laboratory is the development of analog resistive network-based VLSI implementations of early and inter(cid:173)mediate vision algorithms. We demonstrate an experimental cir(cid:173)cuit for smoothing and segmenting noisy and sparse depth data using the resistive fuse and a 1-D edge-detection circuit for com(cid:173)puting zero-crossings using two resistive grids with different space(cid:173)constants. To demonstrate the robustness of our algorithms and of the fabricated analog CMOS VLSI chips, we are mounting these circuits onto small mobile vehicles operating in a real-time, labo(cid:173)ratory environment.
****

The CHIR Algorithm for Feed Forward Networks with Binary Weights
Tal Grossman
A new learning algorithm, Learning by Choice of Internal Rep(cid:173) resetation s (CHIR), was recently introduced. Whereas many algo(cid:173) rithms reduce the learning process to minimizing a cost function over the weights, ou r method treats the internal representations as the fundamental entities to b e determined. The algorithm applies a search procedure in the space of in ternal representations, and a cooperative adaptation of the weights (e.g. by using the perceptron learning rule). Since the introduction of its basic, sin gle output ver(cid:173) sion, the CHIR algorithm was generalized to train any fe ed forward network of binary neurons. Here we present the generalised version of the CHIR algorithm, and further demonstrate its versatility by descri bing how it can be modified in order to train networks with binary (±1) weights. Preliminary tests of this binary version on the random teacher problem are also reported.
****

A Large-Scale Neural Network Which Recognizes Handwritten Kanji Characters
Yoshihiro Mori, Kazuki Joe
We propose a new way to construct a large-scale neural network for 3.000 handwritten Kanji characters recognition. This neural network consists o f 3 parts: a collection of small-scale networks which are trained indivi dually on a small number of Kanji characters; a network which integrates the output from the small-scale networks, and a process to facilitate the i ntegration of these neworks. The recognition rate of the total system is co mparable with those of the small-scale networks. Our results indicate that the proposed method is effective for constructing a large-scale network wit hout loss of recognition performance.
****

Mechanisms for Neuromodulation of Biological Neural Networks
Ronald Harris-Warrick
The pyloric Central Pattern Generator of the crustacean stomatogastric ganglion is a well-defined biological neural network. This 14-neuron network is modulated by many inputs. These inputs reconfigure the network to produc e multiple output patterns by three simple mechanisms: 1) detennining wh ich cells are active; 2) modulating the synaptic efficacy; 3) changing the intrinsic response properties of individual neurons. The importance of m odifiable intrinsic response properties of neurons for network function and m odulation is discussed.
****

Recognizing Hand-Printed Letters and Digits
Gale Martin, James Pittman
We are developing a hand-printed character recognition system using a multi(cid: 173) layered neural net trained through backpropagation. We report on results o f training nets with samples of hand-printed digits scanned off of bank checks and hand-printed letters interactively entered into a computer through a sty(ci d:173) lus digitizer. Given a large training set, and a net with sufficient ca pacity to achieve high performance on the training set, nets typically achie ved error rates of 4-5% at a 0% reject rate and 1-2% at a 10% reject rate. The topology and capacity of the system, as measured by the number of connections in the net, have surprisingly little effect on generalization. For those de veloping practical pattern recognition systems, these results suggest that a l arge and representative training sample may be the single, most important factor in achieving high recognition accuracy. From a scientific standpoint, t hese re(cid:173) sults raise doubts about the relevance to backpropagation of le arning models that estimate the likelihood of high generalization from estimate s of capacity. Reducing capacity does have other benefits however, especially w hen the re(cid:173) duction is accomplished by using local receptive fields wi th shared weights. In this latter case, we find the net evolves feature detecto rs resembling those in visual cortex and Linsker's orientation-selective nodes.

****

TRAFFIC: Recognizing Objects Using Hierarchical Reference Frame Transformations
Richard Zemel, Michael C. Mozer, Geoffrey E. Hinton
We describe a model that can recognize two-dimensional shapes in an unsegmented image, independent of their orientation, position, and scale. The model, called TRAFFIC, efficiently represents the structural relation between an object and each of its component features by encoding the fixed viewpoint-invariant transformation from the feature's reference frame to the object's in the weights of a connectionist network. Using a hierarchy of such transformations, with increasing complexity of features at each successive layer, the network can recognize multiple objects in parallel. An implemen(cid:173)tation of TRAFFIC is described, along with experimental results demonstrating the network's ability to recognize constellations of stars in a viewpoint-invariant manner.

****

Performance of Connectionist Learning Algorithms on 2-D SIMD Processor Arrays
Fernando Nuñez, José Fortes
The mapping of the back-propagation and mean field theory learning algorithms onto a generic 2-D SIMD computer is described. This architecture proves to be very adequate for these applications since efficiencies close to the optimum can be attained. Expressions to find the learning rates are given and then particularized to the DAP array procesor.

****

Neurally Inspired Plasticity in Oculomotor Processes
Paul Viola
We have constructed a two axis camera positioning system which is roughly analogous to a single human eye. This Artificial-Eye (A(cid:173)eye) combines the signals generated by two rate gyroscopes with motion information extracted from visual analysis to stabilize its camera. This stabilization process is similar to the vestibulo-ocular response (VOR); like the VOR, A-eye learns a system model that can be incrementally modified to adapt to changes in its structure, performance and environment. A-eye is an example of a robust sen(cid:173)sory system that performs computations that can be of significant use to the designers of mobile robots.

****

Combining Visual and Acoustic Speech Signals with a Neural Network Improves Intelligibility
Terrence J. Sejnowski, Ben Yuhas, Moise Goldstein, Robert Jenkins
R.E. Jenkins The Applied Physics Laboratory The Johns Hopkins University Laurel, MD 20707

****

The Computation of Sound Source Elevation in the Barn Owl
Clay D. Spence, John Pearson
The midbrain of the barn owl contains a map-like representation of sound source direction which is used to precisely orient the head to(cid:173)ward targets of interest. Elevation is computed from the interaural difference in sound level. We present models and computer simula(cid:173)tions of two stages of level difference processing which qualitatively agree with known anatomy and physiology, and make several strik(cid:173)ing predictions.

****