Backpropagation without Multiplication
Patrice Simard, Hans Graf
The back propagation algorithm has been modified to work with(cid:173) out any m ultiplications and to tolerate comput.ations with a low resolution, which makes it. more attractive for a hardware imple(cid:173) mentatioll. Numbers are repre sented in float.ing point format with 1 bit mantissa and 3 bits in the exponent for the states, and 1 bit mantissa and 5 bit exponent. for the gradients, whil e the weights are 16 bit fixed-point numbers. In this way, all the computations can be executed with shift and add operations. Large nehvorks with over 100,0 00 weights were t.rained and demonstrat.ed the same per(cid:173) formance as net works comput.ed with full precision. An estimate of a circuit implementatioll s hows that a large network can be placed on a single chip , reaching more t.han 1 billion weight updat.es pel' second. A speedup is also obtained on any machin e where a mul(cid:173) tiplication is slower than a shift operat.ioJl.
****

Hoeffding Races: Accelerating Model Selection Search for Classification and Func tion Approximation
Oded Maron, Andrew Moore
Selecting a good model of a set of input points by cross validation is a computationally intensive process, especially if the number of possible m odels or the number of training points is high. Tech(cid:173) niques su ch as gradient descent are helpful in searching through the space of mo dels, but problems such as local minima, and more importantly, lack of a d istance metric between various models re(cid:173) duce the applicability of these search methods. Hoeffding Races is a technique for finding a go od model for the data by quickly dis(cid:173) carding bad models, and conc entrating the computational effort at differentiating between the better one s. This paper focuses on the special case of leave-one-out cross validat ion applied to memory(cid:173) based learning algorithms, but we also a rgue that it is applicable to any class of model selection problems.
****

Putting It All Together: Methods for Combining Neural Networks
Michael Perrone
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.
****

Cross-Validation Estimates IMSE
Mark Plutowski, Shinichi Sakata, Halbert White
Integrated Mean Squared Error (IMSE) is a version of the usual mean s quared error criterion, averaged over all possible training If it could be observed, it could be used sets of a given size. to determine optimal network complexity or optimal data sub(cid:173) sets for efficien t training. We show that two common methods of cross-validating average squared error deliver unbiased estimates of IMSE, converging to IMSE wi th probability one. These esti(cid:173) mates thus make possible approximat e IMSE-based choice of net(cid:173) work complexity. We also show that two va riants of cross validation measure provide unbiased IMSE-based estimates potent ially useful for selecting optimal data subsets.
****

The Power of Amnesia
Dana Ron, Yoram Singer, Naftali Tishby
We propose a learning algorithm for a variable memory length Markov p rocess. Human communication, whether given as text, handwriting, or spe ech, has multi characteristic time scales. On short scales it is char acterized mostly by the dynamics that gen(cid:173) erate the process, w hereas on large scales, more syntactic and se(cid:173) mantic information is carried. For that reason the conventionally used fixed memory Marko v models cannot capture effectively the complexity of such structures. On

the other hand using long mem(cid:173) ory models uniformly is not practical even for as short memory as four. The algorithm we propose is based on minimizing the sta(cid:173) tistical prediction error by extending the memory, or state length, adaptively, until the total prediction error is sufficiently small. We demonstrate the algorithm by learning the structure of natural En(cid:173) glish text and applying the learned model to the correction of cor(cid:173) rupted text . Using less than 3000 states the model's performance is far superior to that of fixed memory models with similar num(cid:173) ber of states. We also show how the algorithm can be applied to intergenic E. coli DNA base prediction with results comparable to HMM based methods.

****

On the Non-Existence of a Universal Learning Algorithm for Recurrent Neural Networks

Herbert Wiklicky

We prove that the so called "loading problem" for (recurrent) neural net(cid:173) works is unsolvable. This extends several results which already demon(cid:173) strated that training and related design problems for neural networks are (at least) NP-complete. Our result also implies that it is impossible to find or to formulate a universal training algorithm, which for any neu(cid:173) ral network architecture could determine a correct set of weights. For the simple proof of this, we will just show that the loading problem is equivalent to "Hilbert's tenth problem" which is known to be unsolvable.

****

Constructive Learning Using Internal Representation Conflicts

Laurens Leerink, Marwan Jabri

We present an algorithm for the training of feedforward and recur(cid:173) rent neural networks. It detects internal representation conflicts and uses these conflicts in a constructive manner to add new neu(cid:173) rons to the network . The advantages are twofold: (1) starting with a small network neurons are only allocated when required; (2) by detecting and resolving internal conflicts at an early stage learning time is reduced. Empirical results on two real-world problems sub(cid:173) stantiate the faster learning speed; when applied to the training of a recurrent network on a well researched sequence recognition task (the Reber grammar), training times are significantly less than previously reported .

****

A Local Algorithm to Learn Trajectories with Stochastic Neural Networks

Javier Movellan

This paper presents a simple algorithm to learn trajectories with a continuous time, continuous activation version of the Boltzmann machine. The algorithm takes advantage of intrinsic Brownian noise in the network to easily compute gradients using entirely local computations. The algorithm may be ideal for parallel hardware implementations.

****

Use of Bad Training Data for Better Predictions

Tal Grossman, Alan Lapedes

We show how randomly scrambling the output classes of various fractions of the training data may be used to improve predictive accuracy of a classification algorithm. We present a method for calculating the "noise sensitivity signature" of a learning algorithm which is based on scrambling the output classes. This signature can be used to indicate a good match between the complexity of the classifier and the complexity of the data. Use of noise sensitivity signatures is distinctly different from other schemes to avoid over(cid:173) training, such as cross-validation, which uses only part of the train(cid:173) ing data, or various penalty functions, which are not data-adaptive. Noise sensitivity signature methods use all of the training data and are manifestly data-adaptive and non-parametric. They are well suited for situations with limited training data.

****

Non-Intrusive Gaze Tracking Using Artificial Neural Networks

Shumeet Baluja, Dean Pomerleau
We have developed an artificial neural network based gaze tracking system which can be customized to individual users. Unlike other gaze trackers, which normally require the user to wear cumbersome headgear, or to use a chin rest to ensure head immobility, our system is entirely non-intrusive. Currently, the best intrusive gaze tracking systems are accurate to approxi(cid:173) mately 0.75 degrees. In our experiments, we have been able to achieve an accuracy of 1.5 degrees, while allowing head mobility. In this paper we present an empirical analysis of the performance of a large number of artifi(cid:173) cial neural network architectures for this task.
****

Learning Curves: Asymptotic Values and Rate of Convergence
Corinna Cortes, L. D. Jackel, Sara Solla, Vladimir Vapnik, John Denker
Training classifiers on large databases is computationally demand(cid:173) ing. It is desirable to develop efficient procedures for a reliable prediction of a classifier's suitability for implementing a given task, so that resources can be assigned to the most promising candidates or freed for exploring new classifier candidates. We propose such a practical and principled predictive method. Practical because it avoids the costly procedure of training poor classifiers on the whole training set, and principled because of its theoretical foundation. The effectiveness of the proposed procedure is demonstrated for both single- and multi-layer networks.
****

How to Describe Neuronal Activity: Spikes, Rates, or Assemblies?
Wulfram Gerstner, J. van Hemmen
What is the 'correct' theoretical description of neuronal activity? The analysis of the dynamics of a globally connected network of spiking neurons (the Spike Response Model) shows that a descrip(cid:173) tion by mean firing rates is possible only if active neurons fire in(cid:173) coherently. If firing occurs coherently or with spatio-temporal cor(cid:173) relations, the spike structure of the neural code becomes relevant. Alternatively, neurons can be gathered into local or distributed en(cid:173) sembles or 'assemblies'. A description based on the mean ensemble activity is, in principle, possible but the interaction between differ(cid:173) ent assemblies becomes highly nonlinear. A description with spikes should therefore be preferred.
****

Digital Boltzmann VLSI for constraint satisfaction and learning
Michael Murray, Ming-Tak Leung, Kan Boonyanit, Kong Kritayakirana, James Burg, Gregory Wolff, Tokahiro Watanabe, Edward Schwartz, David Stork, Allen M. Peterson
We built a high-speed, digital mean-field Boltzmann chip and SBus board for general problems in constraint satjsfaction and learning. Each chip has 32 neural processors and 4 weight update processors, supporting an arbitrary topology of up to 160 functional neurons. On-chip learning is at a theoretical maximum rate of 3.5 x 108 con(cid:173) nection updates/sec; recall is 12000 patterns/sec for typical condi(cid:173) tions. The chip's high speed is due to parallel computation of inner products, limited (but adequate) precision for weights and activa(cid:173) tions (5 bits), fast clock (125 MHz), and several design insights.
****

Feature Densities are Required for Computing Feature Correspondences
Subutai Ahmad
The feature correspondence problem is a classic hurdle in visual object-recognition concerned with determining the correct mapping between the features measured from the image and the features ex(cid:173) pected by the model. In this paper we show that determining good correspondences requires information about the joint probability density over the image features. We propose "likelihood based correspondence matching" as a general principle for selecting op(cid:173) timal correspondences. The approach is applicable to non-rigid models, allows nonlinear perspective transformations, and can op(cid:173) timally deal with occlusions and missing features. Experiments with rigid and non-rigid 3D hand gesture r

ecognition support the theory. The likelihood based techniques show almost no d ecrease in classification performance when compared to performance with perfec t correspondence knowledge.
****

Emergence of Global Structure from Local Associations
Thea Ghiselli-Crippa, Paul Munro
A variant of the encoder architecture, where units at the input and out(cid:173) put layers represent nodes on a graph. is applied to the task of mapping locat ions to sets of neighboring locations. The degree to which the re(cid:173) su lting internal (i.e. hidden unit) representations reflect global proper(cid: 173) ties of the environment depends upon several parameters of the learning pr ocedure. Architectural bottlenecks. noise. and incremental learning of landmark s are shown to be important factors in maintaining topograph(cid:173) ic relat ionships at a global scale.
****

Robot Learning: Exploration and Continuous Domains
David Cohn
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.
****

Lower Boundaries of Motoneuron Desynchronization via Renshaw Interneurons
Mitchell Maltenfort, Robert Druzinsky, C. Heckman, W. Rymer
Using a quasi-realistic model of the feedback inhibition ofmotoneurons (MNs) by Renshaw cells, we show that weak inhibition is sufficient to maximally desync hronize MNs, with negligible effects on total MN activity. MN synchrony can produce a 20 - 30 Hz peak in the force power spectrum, which may cause ins tability in feedback loops.
****

Catastrophic interference in connectionist networks: Can It Be predicted, can It be prevented?
Robert French
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.
****

Signature Verification using a "Siamese" Time Delay Neural Network
Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, Roopak Shah
This paper describes an algorithm for verification of signatures written on a p en-input tablet. The algorithm is based on a novel, artificial neural network, called a "Siamese" neural network. This network consists of two identical sub-n etworks joined at their out(cid:173) puts. During training the two sub-networks extract features from two signatures, while the joining neuron measures the dis tance be(cid:173) tween the two feature vectors. Verification consists of compar ing an extracted feature vector ~ith a stored feature vector for the signer. S ignatures closer to this stored representation than a chosen thresh(cid:173) old are accepted, all other signatures are rejected as forgeries.
****

Connectionism for Music and Audition
Andreas Weigend
This workshop explored machine learning approaches to 3 topics: (1) fi nding structure in music (analysis, continuation, and comple(cid:173) tion of an unfinished piece), (2) modeling perception of time (ex(cid:173) tract ion of musical meter, explanation of human data on timing), and (3) inter polation in timbre space.
****

Central and Pairwise Data Clustering by Competitive Neural Networks
Joachim Buhmann, Thomas Hofmann

Data clustering amounts to a combinatorial optimization problem to re(cid:173) duce the complexity of a data representation and to increase its precision. Central and pairwise data clustering are studied in the maximum en(cid:173) tropy framework. For central clustering we derive a set of reestimation equations and a minimization procedure which yields an optimal num(cid:173) ber of clusters, their centers and their cluster probabilities. A meanfield approximation for pairwise clustering is used to estimate assignment probabilities. A se1fconsistent solution to multidimensional scaling and pairwise clustering is derived which yields an optimal embedding and clustering of data points in a d-dimensional Euclidian space.

****

Processing of Visual and Auditory Space and Its Modification by Experience

Josef Rauschecker, Terrence J. Sejnowski

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.

****

A Connectionist Model of the Owl's Sound Localization System

Daniel Rosen, David Rumelhart, Eric Knudsen

,,"'e do not have a good understanding of how theoretical principles of learning are realized in neural systems. To address this problem we built a computational model of development in the owl's sound localization system. The structure of the model is drawn from known experimental data while the learning principles come from recent work in the field of brain style computation. The model accounts for numerous properties of the owl's sound localization system, makes specific and testable predictions for future experi(cid:173) ments, and provides a theory of the developmental process.

****

Identifying Fault-Prone Software Modules Using Feed-Forward Networks: A Case Study

N. Karunanithi

Functional complexity of a software module can be measured in terms of static complexity metrics of the program text. Classify(cid:173) ing software modules, based on their static complexity measures, into different fault-prone categories is a difficult problem in soft(cid:173) ware engineering. This research investigates the applicability of neural network classifiers for identifying fault-prone software mod(cid:173) ules using a data set from a commercial software system. A pre(cid:173) liminary empirical comparison is performed between a minimum distance based Gaussian classifier, a perceptron classifier and a multilayer layer feed-forward network classifier constructed using a modified Cascade-Correlation algorithm. The modified version of the Cascade-Correlation algorithm constrains the growth of the network size by incorporating a cross-validation check during the output layer training phase. Our preliminary results suggest that a multilayer feed-forward network can be used as a tool for iden(cid:173) tifying fault-prone software modules early during the development cycle. Other issues such as representation of software metrics and selection of a proper training samples are also discussed.

****

Event-Driven Simulation of Networks of Spiking Neurons

Lloyd Watts

A fast event-driven software simulator has been developed for sim(cid:173) ulating large networks of spiking neurons and synapses. The prim(cid:173) itive network elements are designed to exhibit biologically realis(cid:173) tic behaviors, such as spiking, refractoriness, adaptation, axonal delays, summation of post-synaptic current pulses, and tonic cur(cid:173) rent inputs. The efficient event-driven representation allows large networks to be simulated in a fraction of the time that would be required for a full compartmental-model simulation. Correspond(cid:173) ing analog CMOS VLSI circuit primitives have been designed and characterized, so that large-sca

le circuits may be simulated prior to fabrication.
****

Grammatical Inference by Attentional Control of Synchronization in an Oscillating Elman Network
Bill Baird, Todd Troyer, Frank Eeckman
We show how an "Elman" network architecture, constructed from recurrently connected oscillatory associative memory network mod(cid:173)ules, can employ selective "attentional" control of synchronization to direct the flow of communication and computation within the architecture to solve a grammatical inference problem. Previously we have shown how the discrete time "Elman" network algorithm can be implemented in a network completely described by continuous ordinary differential equations. The time steps (ma(cid:173)chine cycles) of the system are implemented by rhythmic variation (clocking) of a bifurcation parameter. In this architecture, oscilla(cid:173)tion amplitude codes the information content or activity of a mod(cid:173)ule (unit), whereas phase and frequency are used to "softwire" the network. Only synchronized modules communicate by exchang(cid:173)ing amplitude information; the activity of non-resonating modules contributes incoherent crosstalk noise. Attentional control is modeled as a special subset of the hidden modules with ouputs which affect the resonant frequencies of other hidden modules. They control synchrony among the other mod(cid:173)ules and direct the flow of computation (attention) to effect transi(cid:173)tions between two subgraphs of a thirteen state automaton which the system emulates to generate a Reber grammar. The internal crosstalk noise is used to drive the required random transitions of the automaton.
****

Bayesian Self-Organization
Alan L. Yuille, Stelios Smirnakis, Lei Xu
Recent work by Becker and Hinton (Becker and Hinton, 1992) shows a promising mechanism, based on maximizing mutual in(cid:173)formation assuming spatial coherence, by which a system can self(cid:173)organize itself to learn visual abilities such as binocular stereo. We introduce a more general criterion, based on Bayesian probability theory, and thereby demonstrate a connection to Bayesian theo(cid:173)ries of visual perception and to other organization principles for early vision (Atick and Redlich, 1990). Methods for implementa(cid:173)tion using variants of stochastic learning are described and, for the special case of linear filtering, we derive an analytic expression for the output.
****

The "Softmax" Nonlinearity: Derivation Using Statistical Mechanics and Useful Properties as a Multiterminal Analog Circuit Element
I. M. Elfadel, J. L. Wyatt, Jr.
We use mean-field theory methods from Statistical Mechanics to derive the "softmax" nonlinearity from the discontinuous winner(cid:173)take-all (WTA) mapping. We give two simple ways of implementing "soft max" as a multiterminal network element. One of these has a number of important network-theoretic properties. It is a recipro(cid:173)cal, passive, incrementally passive, nonlinear, resistive multitermi(cid:173)nal element with a content function having the form of information(cid:173)theoretic entropy. These properties should enable one to use this element in nonlinear RC networks with such other reciprocal el(cid:173)ements as resistive fuses and constraint boxes to implement very high speed analog optimization algorithms using a minimum of hardware.
****

Robust Parameter Estimation and Model Selection for Neural Network Regression
Yong Liu
In this paper, it is shown that the conventional back-propagation (BPP) algorithm for neural network regression is robust to lever(cid:173)ages (data with :n corrupted), but not to outliers (data with y corrupted). A robust model is to model the error as a mixture of normal distribution. The influence function for this mixture model is calculated and the condition for the model to be robust to outliers is given. EM algorithm [5

] is used to estimate the parameter. The usefulness of model selection criteria is also discussed. Illustrative simulations are performed.
****

Computational Elements of the Adaptive Controller of the Human Arm
Reza Shadmehr, Ferdinando Mussa-Ivaldi
We consider the problem of how the CNS learns to control dynam(cid:173) ics of a mechanical system. By using a paradigm where a subject's hand interacts with a virtual mechanical environment, we show that learning control is via composition of a model of the imposed dynamics. Some properties of the computational elements with which the CNS composes this model are inferred through the gen(cid:173) eralization capabilities of the subject outside the training data.
****

Transition Point Dynamic Programming
Kenneth Buckland, Peter Lawrence
Transition point dynamic programming (TPDP) is a memory(cid:173) based, reinforcement learning, direct dynamic programming ap(cid:173) proach to adaptive optimal control that can reduce the learning time and memory usage required for the control of continuous stochastic dynamic systems. TPDP does so by determining an ideal set of transition points (TPs) which specify only the control action changes necessary for optimal control. TPDP converges to an ideal TP set by using a variation of Q-Iearning to assess the mer(cid:173) its of adding, swapping and removing TPs from states throughout the state space. When applied to a race track problem, TPDP learned the optimal control policy much sooner than conventional Q-Iearning, and was able to do so using less memory.
****

Dopaminergic Neuromodulation Brings a Dynamical Plasticity to the Retina
Eric Boussard, Jean-François Vibert
The fovea of a mammal retina was simulated with its detailed bio(cid:173) logical properties to study the local preprocessing of images. The direct visual pathway (photoreceptors, bipolar and ganglion cells) and the horizontal units, as well as the D-amacrine cells were sim(cid:173) ulated. The computer program simulated the analog non-spiking transmission between photoreceptor and bipolar cells, and between bipolar and ganglion cells, as well as the gap-junctions between hor(cid:173) izontal cells, and the release of dopamine by D-amacrine cells and its diffusion in the extra-cellular space. A 64 x 64 photoreceptors retina, containing 16,448 units, was carried out. This retina dis(cid:173) played contour extraction with a Mach effect, and adaptation to brightness. The simulation showed that the dopaminergic amacrine cells were necessary to ensure adaptation to local brightness.
****

Monte Carlo Matrix Inversion and Reinforcement Learning
Andrew Barto, Michael Duff
We describe the relationship between certain reinforcement learn(cid:173) ing (RL) methods based on dynamic programming (DP) and a class of unorthodox Monte Carlo methods for solving systems of linear equations proposed in the 1950's. These methods recast the solu(cid:173) tion of the linear system as the expected value of a statistic suitably defined over sample paths of a Markov chain. The significance of our observations lies in arguments (Curtiss, 1954) that these Monte Carlo methods scale better with respect to state-space size than do standard, iterative techniques for solving systems of linear equa(cid:173) tions. This analysis also establishes convergence rate estimates. Because methods used in RL systems for approximating the evalu(cid:173) ation function of a fixed control policy also approximate solutions to systems of linear equations, the connection to these Monte Carlo methods establishes that algorithms very similar to TD algorithms (Sutton, 1988) are asymptotically more efficient in a precise sense than other methods for evaluating policies. Further, all DP-based RL methods have some of the properties of these Monte Carlo al(cid:173) gorithms, which suggests that although RL is often perceived to be slow, for sufficiently large pro

blems, it may in fact be more ef(cid:173) ficient than other known classes of me thods capable of producing the same results.

****

Globally Trained Handwritten Word Recognizer using Spatial Representation, Convo lutional Neural Networks, and Hidden Markov Models

Yoshua Bengio, Yann LeCun, Donnie Henderson

We introduce a new approach for on-line recognition of handwrit(cid:173) ten wor ds written in unconstrained mixed style. The preprocessor performs a word-level normalization by fitting a model of the word structure using the EM algorithm. Words are then coded into low resolution "annotated images" where each pixel c ontains informa(cid:173) tion about trajectory direction and curvature. The reco gnizer is a convolution network which can be spatially replicated. From the ne twork output, a hidden Markov model produces word scores. The entire system is globally trained to minimize word-level errors.

****

Robust Reinforcement Learning in Motion Planning

Satinder Singh, Andrew Barto, Roderic Grupen, Christopher Connolly

While exploring to find better solutions, an agent performing on(cid:173) line r einforcement learning (RL) can perform worse than is accept(cid:173) able. In so me cases, exploration might have unsafe, or even catas(cid:173) trophic, results , often modeled in terms of reaching 'failure' states of the agent's environmen t. This paper presents a method that uses domain knowledge to reduce the number of failures during explo(cid:173) ration. This method formulates the set of act ions from which the RL agent composes a control policy to ensure that explorati on is conducted in a policy space that excludes most of the unacceptable polic ies. The resulting action set has a more abstract relationship to the task bein g solved than is common in many applications of RL. Although the cost of this a dded safety is that learning may result in a suboptimal solution, we argue that this is an appropri(cid:173) ate tradeoff in many problems. We illustrate this method in the domain of motion planning.

****

Optimal Stopping and Effective Machine Complexity in Learning

Changfeng Wang, Santosh Venkatesh, J. Judd

We study tltt' problem of when to stop If'arning a class of feedforw ard networks - networks with linear outputs I1PUrOIl and fixed input w eights - when they are trained with a gradient descent algorithm on a finite number of examples. Under general regularity conditions, it is shown that there a.re in general three distinct phases in the generali zation performance in the learning process, and in particular, the network has hetter gt'neralization pPTformance when learning is stopped at a certain time before til(' global miniIl111lu of the empirical error is re achert. A notion of effective size of a machine is rtefil1e

****

Exploiting Chaos to Control the Future

Gary Flake, Guo-Zhen Sun, Yee-Chun Lee

Recently, Ott, Grebogi and Yorke (OGY) [6] found an effective method to control chaotic systems to unstable fixed points by us(cid:173) ing only small control forces; however, OGY's method is based on and limited to a linear theory and re quires considerable knowledge of the dynamics of the system to be controlled. I n this paper we use two radial basis function networks: one as a model of an un known plant and the other as the controller. The controller is trained with a recurrent learning algorithm to minimize a novel objective function such that t he controller can locate an unstable fixed point and drive the system into the fixed point with no a priori knowl(cid:173) edge of the system dynamics. Our res ults indicate that the neural controller offers many advantages over OGY's tech nique.

****

Backpropagation Convergence Via Deterministic Nonmonotone Perturbed Minimization

O. L. Mangasarian, M. V. Solodov

The fundamental backpropagation (BP) algorithm for training ar(cid:173) tificial neural networks is cast as a deterministic nonmonotone per(cid:173) turbed gradient method. Under certain natural assumptions, such as the series of learning rates diverging while the series of their squares converging, it is established that every accumulation point of the online BP iterates is a stationary point of the BP error func(cid:173) tion. The results presented cover serial and parallel online BP, modified BP with a momentum term, and BP with weight decay.

****

Fool's Gold: Extracting Finite State Machines from Recurrent Network Dynamics

John Kolen

Several recurrent networks have been proposed as representations for the task of formal language learning. After training a recurrent network rec(cid:173) ognize a formal language or predict the next symbol of a sequence, the next logical step is to understand the information processing carried out by the network. Some researchers have begun to extracting finite state machines from the internal state trajectories of their recurrent networks. This paper describes how sensitivity to initial conditions and discrete measurements can trick these extraction methods to return illusory finite state descriptions.

****

Two Iterative Algorithms for Computing the Singular Value Decomposition from Input/Output Samples

Terence Sanger

The Singular Value Decomposition (SVD) is an important tool for linear algebra and can be used to invert or approximate matrices. Although many authors use "SVD" synonymously with "Eigen(cid:173) vector Decomposition" or "Principal Components Transform", it is important to realize that these other methods apply only to symmetric matrices, while the SVD can be applied to arbitrary nonsquare matrices. This property is important for applications to signal transmission and control. I propose two new algorithms for iterative computation of the SVD given only sample inputs and outputs from a matrix. Although there currently exist many algorithms for Eigenvector Decomposi(cid:173) tion (Sanger 1989, for example), these are the first true sample(cid:173) based SVD algorithms.

****

Efficient Computation of Complex Distance Metrics Using Hierarchical Filtering

Patrice Simard

By their very nature, memory based algorithms such as KNN or Parzen windows require a computationally expensive search of a large database of prototypes. In this paper we optimize the search(cid:173) ing process for tangent distance (Simard, LeCun and Denker, 1993) to improve speed performance. The closest prototypes are found by recursively searching included subset.s of the database using dis(cid:173) tances of increasing complexit.y. This is done by using a hierarchy of tangent distances (increasing the Humber of tangent. vectors from o to its maximum) and multiresolution (using wavelets). At each stage, a confidence level of the classification is computed. If the confidence is high enough, the c.omputation of more complex dis(cid:173) tances is avoided. The resulting algorithm applied to character recognition is close to t.hree orders of magnitude faster than com(cid:173) puting the full tangent dist.ance on every prot.ot.ypes .

****

Statistics of Natural Images: Scaling in the Woods

Daniel Ruderman, William Bialek

In order to best understand a visual system one should attempt to characterize the natural images it processes. We gather images from the woods and find that these scenes possess an ensemble scale invariance. Further, they are highly non-Gaussian, and this non(cid:173) Gaussian character cannot be removed through local linear filter(cid:173) ing. We find that including a simple "gain control" nonlinearity in the filtering process makes the filter output quite Gaussian, mean(cid:173) ing information is maximized at fixed channel variance. Finally, we use the measured power spectrum to place an upper bound on the information conveyed about natural scenes by an array of

receptors.
****

GDS: Gradient Descent Generation of Symbolic Classification Rules
Reinhard Blasig
Imagine you have designed a neural network that successfully learns a complex classification task. What are the relevant input features the classifier relies on and how are these features combined to pro(cid:173)duce the classification decisions? There are applications where a deeper insight into the structure of an adaptive system and thus into the underlying classification problem may well be as important as the system's performance characteristics, e.g. in economics or medicine. GDSi is a backpropagation-based training scheme that produces networks transformable into an equivalent and concise set of IF-THEN rules. This is achieved by imposing penalty terms on the network parameters that adapt the network to the expressive power of this class of rules. Thus during training we simultaneously minimize classification and transformation error. Some real-world tasks demonstrate the viability of our approach.
****

Packet Routing in Dynamically Changing Networks: A Reinforcement Learning Approach
Justin Boyan, Michael Littman
This paper describes the Q-routing algorithm for packet routing, in which a reinforcement learning module is embedded into each node of a switching network. Only local communication is used by each node to keep accurate statistics on which routing decisions lead to minimal delivery times. In simple experiments involving a 36-node, irregularly connected network, Q-routing proves supe(cid:173)rior to a nonadaptive algorithm based on precomputed shortest paths and is able to route efficiently even when critical aspects of the simulation, such as the network load, are allowed to vary dy(cid:173)namically. The paper concludes with a discussion of the tradeoff between discovering shortcuts and maintaining stable policies.
****

A Massively-Parallel SIMD Processor for Neural Network and Machine Vision Applications
Michael Glover, W. Miller
This paper describes the MM32k, a massively-parallel SIMD com(cid:173)puter which is easy to program, high in performance, low in cost and effective for implementing highly parallel neural network ar(cid:173)chitectures. The MM32k has 32768 bit serial processing elements, each of which has 512 bits of memory, and all of which are inter(cid:173)connected by a switching network. The entire system resides on a single PC-AT compatible card. It is programmed from the host computer using a C++ language class library which abstracts the parallel processor in terms of fast arithmetic operators for vectors of variable precision integers.
****

How to Choose an Activation Function
H. N. Mhaskar, C. A.. Micchelli
We study the complexity problem in artificial feedforward neural networks desig(cid:173)ned to approximate real valued functions of several real variables; i.e., we es(cid:173)timate the number of neurons in a network required to ensure a given degree of approximation to every function in a given function class. We indicate how to construct networks with the indicated number of neurons evaluating standard activation functions. Our general theorem shows that the smoother the activation function, the better the rate of approximation.
****

Efficient Simulation of Biological Neural Networks on Massively Parallel Supercomputers with Hypercube Architecture
Ernst Niebur, Dean Brettle
We present a neural network simulation which we implemented on the ma

ssively parallel Connection Machine 2. In contrast to previous work, t
his simulator is based on biologically realistic neu(cid:173) rons with nontri
vial single-cell dynamics, high connectivity with a structure modelled in
agreement with biological data, and preser(cid:173) vation of the temporal dy
namics of spike interactions. We simulate neural networks of 16,384 neurons c
oupled by about 1000 synapses per neuron, and estimate the performance for muc
h larger systems. Communication between neurons is identified as the computat
ion(cid:173) ally most demanding task and we present a novel method to over
(cid:173) come this bottleneck. The simulator has already been used to study t
he primary visual system of the cat.

****

Convergence of Stochastic Iterative Dynamic Programming Algorithms

Tommi Jaakkola, Michael Jordan, Satinder Singh

Increasing attention has recently been paid to algorithms based on dynamic pr
ogramming (DP) due to the suitability of DP for learn(cid:173) ing problems
involving control. In stochastic environments where the system being con
trolled is only incompletely known, however, a unifying theoretical acco
unt of these methods has been missing. In this paper we relate DP-base
d learning algorithms to the pow(cid:173) erful techniques of stochastic ap
proximation via a new convergence theorem, enabling us to establish a clas
s of convergent algorithms to which both TD(") and Q-Iearning belong.

****

High Performance Neural Net Simulation on a Multiprocessor System with "Intellig
ent" Communication

Urs A. Müller, Michael Kocheisen, Anton Gunzinger

The performance requirements in experimental research on arti(cid:173) ficial ne
ural nets often exceed the capability of workstations and PCs by a great amount
. But speed is not the only requirement. Flexibility and implementation time fo
r new algorithms are usually of equal importance. This paper describes the simu
lation of neural nets on the MUSIC parallel supercomputer, a system that shows
a good balance between the three issues and therefore made many research proje
cts possible that were unthinkable before. (MUSIC stands for Multiprocessor Sys
tem with Intelligent Communication)

****

An Analog VLSI Model of Central Pattern Generation in the Leech

Micah Siegel

I detail the design and construction of an analog VLSI model of the ne
ural system responsible for swimming behaviors of the leech. Why the leech?
The biological network is small and relatively well understood, and th
e silicon model can therefore span three levels of organization in the
leech nervous system (neuron, ganglion, system); it represents one of the fir
st comprehensive models of leech swimming operating in real-time. The circui
t employs biophysically motivated analog neurons networked to form multiple
biologically inspired silicon ganglia. These ganglia are coupled using kn
own interganglionic connections. Thus the model retains the flavor of
its biological counterpart, and though simplified, the output of the silic
on circuit is similar to the output of the leech swim central pattern generato
r. The model operates on the same time- and spatial-scale as the leech nervous
system and will provide an excellent platform with which to explore r
eal-time adaptive locomotion in the leech and other "simple" invertebra
te nervous systems.

****

Non-Linear Statistical Analysis and Self-Organizing Hebbian Networks

Jonathan Shapiro, Adam Prügel-Bennett

Neurons learning under an unsupervised Hebbian learning rule can perform a non
linear generalization of principal component analysis. This relationship betwee
n nonlinear PCA and nonlinear neurons is reviewed. The stable fixed points
of the neuron learning dynamics correspond to the maxima of the statist,ic
optimized under non(cid:173) linear PCA. However, in order to predict.
what the neuron learns, knowledge of the basins of attractions of the n

euron dynamics is required. Here the correspondence between nonlinear PCA and neural networks breaks down. This is shown for a simple model. Methods of statistical mechanics can be used to find the optima of the objective function of non-linear PCA. This determines what the neurons can learn. In order to find how the solutions are parti(cid:173) tioned amoung the neurons, however, one must solve the dynamics.

****

Locally Adaptive Nearest Neighbor Algorithms

Dietrich Wettschereck, Thomas Dietterich

Four versions of a k-nearest neighbor algorithm with locally adap(cid:173) tive k are introduced and compared to the basic k-nearest neigh(cid:173) bor algorithm (kNN). Locally adaptive kNN algorithms choose the value of k that should be used to classify a query by consulting the results of cross-validation computations in the local neighborhood of the query. Local kNN methods are shown to perform similar to kNN in experiments with twelve commonly used data sets. Encour(cid:173) aging results in three constructed tasks show that local methods can significantly outperform kNN in specific applications. Local methods can be recommended for on-line learning and for appli(cid:173) cations where different regions of the input space are covered by patterns solving different sub-tasks.

****

Bayesian Backprop in Action: Pruning, Committees, Error Bars and an Application to Spectroscopy

Hans Thodberg

MacKay's Bayesian framework for backpropagation is conceptually appealing as well as practical. It automatically adjusts the weight decay parameters during training, and computes the evidence for each trained network. The evidence is proportional to our belief in the model. The networks with highest evidence turn out to generalise well. In this paper, the framework is extended to pruned nets, leading to an Ockham Factor for "tuning the architecture to the data". A committee of networks, selected by their high evidence, is a natural Bayesian construction. The evidence of a committee is computed. The framework is illustrated on real-world data from a near infrared spectrometer used to determine the fat content in minced meat. Error bars are computed, including the contribution from the dissent of the committee members.

****

Solvable Models of Artificial Neural Networks

Sumio Watanabe

Solvable models of nonlinear learning machines are proposed, and learning in artificial neural networks is studied based on the theory of ordinary differential equations. A learning algorithm is con(cid:173) structed, by which the optimal parameter can be found without any recursive procedure. The solvable models enable us to analyze the reason why experimental results by the error backpropagation often contradict the statistical learning theory.

****

A Comparative Study of a Modified Bumptree Neural Network with Radial Basis Function Networks and the Standard Multi Layer Perceptron

Richard Bostock, Alan Harget

Bumptrees are geometric data structures introduced by Omohundro (1991) to provide efficient access to a collection of functions on a Euclidean space of interest. We describe a modified bumptree structure that has been employed as a neural network classifier, and compare its performance on several classification tasks against that of radial basis function networks and the standard mutIi-Iayer perceptron.

****

Generalization Error and the Expected Network Complexity

Chuanyi Ji

O(E~) O((EK)d l N) ,

****

Combined Neural Networks for Time Series Analysis

Iris Ginzburg, David Horn

We propose a method for improving the performance of any net(cid:173) work designed to predict the next value of a time series. Vve advo(cid:173) cate analyzing the deviations of the network's predictions from the data in th e training set. This can be carried out by a secondary net(cid:173) work train ed on the time series of these residuals. The combined system of the two networks is viewed as the new predictor. We demonstrate the simpl icity and success of this method, by apply(cid:173) ing it to the sunspots data. The small corrections of the secondary network can be regarded as resulting from a Taylor expansion of a complex network which includes the combined system. \Te find that the complex network is more difficu lt to train and performs worse than the two-step procedure of the combin ed system.

****

Illumination-Invariant Face Recognition with a Contrast Sensitive Silicon Retina

Joachim Buhmann, Martin Lades, Frank Eeckman

Changes in lighting conditions strongly effect the performance and reli(cid:173) ability of computer vision systems. We report face recognition results under d rastically changing lighting conditions for a computer vision sys(cid:173) tem w hich concurrently uses a contrast sensitive silicon retina and a con(cid:173) ve ntional, gain controlled CCO camera. For both input devices the face recognitio n system employs an elastic matching algorithm with wavelet based features to c lassify unknown faces. To assess the effect of analog on -chip preprocessing by the silicon retina the CCO images have been "digitally preprocessed" with a ba ndpass filter to adjust the power spec(cid:173) trum. The silicon retina with it s ability to adjust sensitivity increases the recognition rate up to 50 percent . These comparative experiments demonstrate that preprocessing with an analog V LSI silicon retina gen(cid:173) erates image data enriched with object-constant features.

****

Address Block Location with a Neural Net System

Hans Graf, Eric Cosatto

We developed a system for finding address blocks on mail pieces that can proces s four images per second. Besides locating the address block, our system al so determines the writing style, handwritten or machine printed, and moreover, it measures the skew angle of the text lines and cleans noisy images. A layout analysis of all the elements present in the image is perfor med in order to distinguish drawings and dirt from text and to separate text of advertisement from that of the destination address. A speed of more than f our images per second is obtained on a modular hardware platform, containing a board with two of the NET32K neural net chips, a SP ARC2 processor board , and a board with 2 digital signal processors. The system has been test ed with more than 100,000 images. Its performance depends on the quality of the images, and lies between 85% correct location in very noisy images to over 98% in cleaner images.

****

Assessing the Quality of Learned Local Models

Stefan Schaal, Christopher Atkeson

An approach is presented to learning high dimensional functions in the case where the learning algorithm can affect the generation of new da ta. A local modeling algorithm, locally weighted regression, is used to repre sent the learned function. Architectural parameters of the approach, such as d istance metrics, are also localized and become a function of the query point in stead of being global. Statistical tests are given for when a local model is g ood enough and sampling should be moved to a new area. Our methods explicitly d eal with the case where prediction accuracy requirements exist during explorati on: By gradually shifting a "center of exploration" and controlling the speed o f the shift with local pre(cid:173) diction accuracy, a goal-directed explorati

on of state space takes place along the fringes of the current data support un til the task goal is achieved. We illustrate this approach with simulation results and results from a real robot learning a complex juggling task.
****

Synchronization, oscillations, and 1/f noise in networks of spiking neurons
Martin Stemmler, Marius Usher, Christof Koch, Zeev Olami
We investigate a model for neural activity that generates long range temporal c orrelations, 1/ f noise, and oscillations in global activity. The model consi sts of a two-dimensional sheet of leaky integrate(cid:173) and-fire neurons with feedback connectivity consisting of local ex(cid:173) citation and sur round inhibition. Each neuron is independently driven by homogeneous ex ternal noise. Spontaneous symmetry breaking occurs, resulting in the forma tion of "hotspots" of activ(cid:173) ity in the network. These localized p atterns of excitation appear as clusters that coalesce, disintegrate, or f luctuate in size while si(cid:173) multaneously moving in a random walk constr ained by the interac(cid:173) tion with other clusters. The emergent cross-c orrelation functions have a dual structure, with a sharp peak around z ero on top of a much broader hill . The power spectrum associated wit h single units shows a 1/ f decay for small frequencies and is flat at h igher frequencies, while the power spectrum of the spiking activity aver(cid:1 73) aged over many cells-equivalent to the local field potential-shows no 1/ f decay but a prominent peak around 40 Hz.
****

What Does the Hippocampus Compute?: A Precis of the 1993 NIPS Workshop
Mark Gluck
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.
****

Unsupervised Learning of Mixtures of Multiple Causes in Binary Data
Eric Saund
This paper presents a formulation for unsupervised learning of clus(cid:173) ter s reflecting multiple causal structure in binary data. Unlike the standard mixture model, a multiple cause model accounts for ob(cid:173) served d ata by combining assertions from many hidden causes, each of which can pertain to varying degree to any subset of the observ(cid:173) able dimensions. A cruc ial issue is the mixing-function for combin(cid:173) ing beliefs from dif ferent cluster-centers in order to generate data reconstructions whose err ors are minimized both during recognition and learning. We demonstrate a we akness inherent to the popular weighted sum followed by sigmoid squashing, and offer an alterna(cid:173) tive form of the nonlinearity. Results are presented demonstrating the algorithm's ability successfully to discover coherent multiple causal representat.ions of noisy test data and in im ages of printed characters.
****

A Hodgkin-Huxley Type Neuron Model That Learns Slow Non-Spike Oscillation
Kenji Doya, Allen Selverston, Peter Rowat
A gradient descent algorithm for parameter estimation which is similar to those used for continuous-time recurrent neural networks was derived for Hodgkin-Huxley type neuron models. Using mem(cid:173) brane potential trajec tories as targets, the parameters (maximal conductances, thresholds and s lopes of activation curves, time con(cid:173) stants) were successfully estima ted. The algorithm was applied to modeling slow non-spike oscillation o f an identified neuron in the lobster stomatogastric ganglion. A model wit h three ionic currents was trained with experimental data. It revealed a novel role of A-current for slow oscillation below -50 mY.
****

Memory-Based Methods for Regression and Classification
Thomas Dietterich, Dietrich Wettschereck, Chris G. Atkeson, Andrew Moore

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 Network Methods for Optimization Problems
Arun Jagota
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.
****

A Comparison of Dynamic Reposing and Tangent Distance for Drug Activity Prediction
Thomas Dietterich, Ajay Jain, Richard Lathrop, Tomás Lozano-Pérez
In drug activity prediction (as in handwritten character recogni(cid:173) tion), the features extracted to describe a training example depend on the pose (location, orientation, etc.) of the example. In hand(cid:173) written character recognition, one of the best techniques for ad(cid:173) dressing this problem is the tangent distance method of Simard, LeCun and Denker (1993). Jain, et al. (1993a; 1993b) introduce a new technique-dynamic reposing-that also addresses this prob(cid:173) lem. Dynamic reposing iteratively learns a neural network and then reposes the examples in an effort to maximize the predicted out(cid:173) put values. New models are trained and new poses computed until models and poses converge. This paper compares dynamic reposing to the tangent distance method on the task of predicting the bio(cid:173) logical activity of musk compounds. In a 20-fold cross-validation,
****

Hoo Optimality Criteria for LMS and Backpropagation
Babak Hassibi, Ali H. Sayed, Thomas Kailath
We have recently shown that the widely known LMS algorithm is an H OO optimal estimator. The H OO criterion has been introduced, initially in the control theory literature, as a means to ensure ro(cid:173) bust performance in the face of model uncertainties and lack of statistical information on the exogenous signals. We extend here our analysis to the nonlinear setting often encountered in neural networks, and show that the backpropagation algorithm is locally H OO optimal. This fact provides a theoretical justification of the widely observed excellent robustness properties of the LMS and backpropagation algorithms. We further discuss some implications of these results.
****

Classifying Hand Gestures with a View-Based Distributed Representation
Trevor J. Darrell, Alex P. Pentland
We present a method for learning, tracking, and recognizing human hand gestures recorded by a conventional CCD camera without any special gloves or other sensors. A view-based representation is used to model aspects of the hand relevant to the trained gestures, and is found using an unsupervised clustering techniqu e. We use normalized correlation net(cid:173) works, with dynamic time warping in the temporal domain, as a distance function for unsupervised clustering. Views are computed separably for space and time dimensions; the distributed response of the combination of these units characterizes the input data with a low dimensional repre(cid:173) sentation. A supervised classification stage uses labeled outputs of the spatio-temporal units as training data. Our system can correctly classify gestures in real time with a low-cost image processing accelerator.
****

Optimal Unsupervised Motor Learning Predicts the Internal Representation of Barn Owl Head Movements
Terence Sanger
(Masino and Knudsen 1990) showed some remarkable results which suggest that he

ad motion in the barn owl is controlled by distinct circuits coding for the horizontal and vertical components of move(cid:173) ment. This implies the exis tence of a set of orthogonal internal co(cid:173) ordinates that are related to meaningful coordinates of the external world. No coherent computational the ory has yet been proposed to explain this finding. I have proposed a simp le model which pro(cid:173) vides a framework for a theory of low-level motor learning. I show that the theory predicts the observed microstimulation results in the barn owl. The model rests on the concept of "Optimal U n(cid:173) supervised Motor Learning", which provides a set of criteri a that predict optimal internal representations. I describe two iterativ e Neural Network algorithms which find the optimal solution and demons trate possible mechanisms for the development of internal representations in animals.

****

Amplifying and Linearizing Apical Synaptic Inputs to Cortical Pyramidal Cells

Öjvind Bernander, Christof Koch, Rodney Douglas

Intradendritic electrophysiological recordings reveal a bewildering reperto ire of complex electrical spikes and plateaus that are dif(cid:173) ficu lt to reconcile with conventional notions of neuronal function. In this p aper we argue that such dendritic events are just an exu(cid:173) berant exp ression of a more important mechanism - a proportional current amplifier whose primary task is to offset electrotonic losses. Using the example of functionall y important synaptic inputs to the superficial layers of an anatomically and electrophysiologically re(cid:173) constructed layer 5 pyramidal neuron, we derive and simulate the properties of conductances that linearize and ampli fy distal synap(cid:173) tic input current in a graded manner. The ampli fication depends on a potassium conductance in the apical tuft and calcium con duc(cid:173) tances in the apical trunk.

****

Decoding Cursive Scripts

Yoram Singer, Naftali Tishby

Online cursive handwriting recognition is currently one of the most intriguin g challenges in pattern recognition. This study presents a novel approach to this problem which is composed of two comple(cid:173) mentary phases. The first is dynamic encoding of the writing tra(cid:173) jectory into a compact sequence of discrete motor control symbols. In this compact represent ation we largely remove the redundancy of the script, while preserving most of its intelligible components. In the second phase these control sequences are used to train adaptive probabilistic acyclic automata (PAA) for the impor tant ingredients of the writing trajectories, e.g. letters. We present a new and effi(cid:173) cient learning algorithm for such stochastic automata, and demon(cid:173) strate its utility for spotting and segmentation of cur sive scripts. Our experiments show that over 90% of the letters are correctly spotted and identified, prior to any higher level language mo del. Moreover, both the training and recognition algorithms are very e fficient compared to other modeling methods, and the models are 'on-line' ada ptable to other writers and styles.

****

Foraging in an Uncertain Environment Using Predictive Hebbian Learning

P. Montague, Peter Dayan, Terrence J. Sejnowski

Survival is enhanced by an ability to predict the availability of food , the likelihood of predators, and the presence of mates. We present a concrete model that uses diffuse neurotransmitter systems to implement a pred ictive version of a Hebb learning rule embedded in a neural ar(cid:173) chitecture based on anatomical and physiological studies on bees. The mo del captured the strategies seen in the behavior of bees and a number of other animals when foraging in an uncertain environment. The predictive model sugge sts a unified way in which neuromodulatory influences can be used to bias ac tions and control synaptic plasticity.

****

Analysis of Short Term Memories for Neural Networks
Jose C. Principe, Hui-H. Hsu, Jyh-Ming Kuo
Short term memory is indispensable for the processing of time varying informati on with artificial neural networks. In this paper a model for linear memories i s presented, and ways to include memories in connectionist topologies are discu ssed. A comparison is drawn among different memory types, with indication of wh at is the salient characteristic of each memory model.
****
Credit Assignment through Time: Alternatives to Backpropagation
Yoshua Bengio, Paolo Frasconi
Learning to recognize or predict sequences using long-term con(cid:173) text has many applications. However, practical and theoretical problems are found in training recurrent neural networks to per(cid:173) form ta sks in which input/output dependencies span long intervals. Starting from a mat hematical analysis of the problem, we consider and compare alternative algori thms and architectures on tasks for which the span of the input/output depen dencies can be controlled. Results on the new algorithms show performance qu alitatively su(cid:173) perior to that obtained with backpropagation.
****
Analyzing Cross-Connected Networks
Thomas Shultz, Jeffrey Elman
Jeffrey L. Elman
****
Recognition-based Segmentation of On-Line Cursive Handwriting
Nicholas Flann
This paper introduces a new recognition-based segmentation ap(cid:173) proach to recognizing on-line cursive handwriting from a database of 10,000 English word s. The original input stream of z, y pen coor(cid:173) dinates is encoded as a s equence of uniform stroke descriptions that are processed by six feed-forward n eural-networks, each designed to recognize letters of different sizes. Words ar e then recognized by performing best-first search over the space of all possibl e segmen(cid:173) tations. Results demonstrate that the method is effective at b oth writer dependent recognition (1.7% to 15.5% error rate) and writer indepen dent recognition (5.2% to 31.1% error rate).
****
Implementing Intelligence on Silicon Using Neuron-Like Functional MOS Transistor s
Tadashi Shibata, Koji Kotani, Takeo Yamashita, Hiroshi Ishii, Hideo Kosaka, Tada hiro Ohmi
We will present the implementation of intelligent electronic circuits r ealized for the first time using a new functional device called Neuron MOS Transistor (neuMOS or vMOS in short) simulating the behavior of b iological neurons at a single transistor level. Search for the most r esembling data in the memory cell array, for instance, can be automat ically software manipulation. Soft Hardware, which we named, can arbitrar ily change its logic function in real time by external control signal s without any hardware modification. Implementation of a neural network equipped with an on-chip self-learning capability is also described. Thr ough the studies of vMOS intelligent circuit implementation, we noticed an interesting similarity in the architectures of vMOS logic circuitry and biological systems.
****
Probabilistic Anomaly Detection in Dynamic Systems
Padhraic Smyth
This paper describes probabilistic methods for novelty detection when u sing pattern recognition methods for fault monitoring of dynamic system s. The problem of novelty detection is particular(cid:173) ly acute when prior knowledge and training data only allow one to construct an incom plete classification model. Allowance must be made in model design so that the classifier will be robust to data generated by classes not

included in the training phase. For diagnosis applications one practica l approach is to construct both an input density model and a discrimi native class model. Using Bayes' rule and prior estimates of the relat ive likelihood of data of known and unknown origin the resulting classificat ion equations are straightforward. The paper describes the application o f this method in the context of hidden Markov models for online fault monitoring of large ground antennas for spacecraft tracking, with particu lar application to the detection of transient behaviour of unknown ori gin.

****

Learning Complex Boolean Functions: Algorithms and Applications

Arlindo Oliveira, Alberto Sangiovanni-Vincentelli

The most commonly used neural network models are not well suited to direct digi tal implementations because each node needs to per(cid:173) form a large number of operations between floating point values. Fortunately, the ability to learn from examples and to generalize is not restricted to networks ofthis type. Inde ed, networks where each node implements a simple Boolean function (Boolean netw orks) can be designed in such a way as to exhibit similar properties. Two algo rithms that generate Boolean networks from examples are pre(cid:173) sented. The results show that these algorithms generalize very well in a class of problems that accept compact Boolean network descriptions. The techniques described are general and can be ap(cid:173) plied to tasks that are not known to have that c haracteristic. Two examples of applications are presented: image reconstruction and hand-written character recognition.

****

A Unified Gradient-Descent/Clustering Architecture for Finite State Machine Indu ction

Sreerupa Das, Michael C. Mozer

Although recurrent neural nets have been moderately successful in learn ing to emulate finite-state machines (FSMs), the continu(cid:173) ous in ternal state dynamics of a neural net are not well matched to the di screte behavior of an FSM. We describe an architecture, called DOLCE, t hat allows discrete states to evolve in a net as learn(cid:173) ing progresses . DOLCE consists of a standard recurrent neural net trained by gradien t descent and an adaptive clustering technique that quantizes the state space. DOLCE is based on the assumption that a finite set of discrete i nternal states is required for the task, and that the actual network state belongs to this set but has been corrupted by noise due to inaccuracy in the weights. DOLCE learns to recover the discrete state with maximum a p osteriori probabil(cid:173) ity from the noisy state. Simulations show t hat DOLCE leads to a significant improvement in generalization performan ce over earlier neural net approaches to FSM induction.

****

Convergence of Indirect Adaptive Asynchronous Value Iteration Algorithms

Vijaykumar Gullapalli, Andrew Barto

Reinforcement Learning methods based on approximating dynamic programming (DP) are receiving increased attention due to their utility in forming re active control policies for systems embedded in dynamic environments. E nvironments are usually modeled as controlled Markov processes, but whe n the environment model is not known a priori, adaptive methods are necess ary. Adaptive con(cid:173) trol methods are often classified as being direct or indirect. Direct methods directly adapt the control policy from experience, whereas indirect methods adapt a model of the controlled process and com(ci d:173) pute control policies based on the latest model. Our focus is on indirect adaptive DP-based methods in this paper. We present a con vergence result for indirect adaptive asynchronous value itera(cid:173) t ion algorithms for the case in which a look-up table is used to store the valu e function. Our result implies convergence of several ex(cid:173) isting reinforcement learning algorithms such as adaptive real-time dynamic programmi ng (ARTDP) (Barto, Bradtke, & Singh, 1993) and prioritized sweeping (Moor

e & Atkeson, 1993). Although the emphasis of researchers studying DP-bas ed reinforcement learning has been on direct adaptive methods such as Q-Learn ing (Watkins, 1989) and methods using TD algorithms (Sutton, 1988), it is not clear that these direct methods are preferable in practice to indirect methods such as those analyzed in this paper.

****

Fast Pruning Using Principal Components

Asriel Levin, Todd Leen, John Moody

We present a new algorithm for eliminating excess parameters and improving network generalization after supervised training. The method, "Principal Components Pruning (PCP)", is based on prin(cid:173) cipal component analysis of the node activations of successive layers of the network. It is simple, che ap to implement, and effective. It requires no network retraining, and d oes not involve calculating the full Hessian of the cost function. Only th e weight and the node activity correlation matrices for each layer of nodes are required. We demonstrate the efficacy of the method on a regression prob lem using polynomial basis functions, and on an economic time series pred iction problem using a two-layer, feedforward network.

****

Hidden Markov Models for Human Genes

Pierre Baldi, Søren Brunak, Yves Chauvin, Jacob Engelbrecht, Anders Krogh

Human genes are not continuous but rather consist of short cod(cid:173) ing regions (exons) interspersed with highly variable non-coding region s (introns). We apply HMMs to the problem of modeling ex(cid:173) ons, intro ns and detecting splice sites in the human genome. Our most interesting result so far is the detection of particular oscilla(cid:173) tory pattern s, with a minimal period ofroughly 10 nucleotides, that seem to be characterist ic of exon regions and may have significant biological implications.

****

Postal Address Block Location Using a Convolutional Locator Network

Ralph Wolf, John Platt

This paper describes the use of a convolutional neural network to perf orm address block location on machine-printed mail pieces. Locating the address block is a difficult object recognition problem because there is often a la rge amount of extraneous printing on a mail piece and because address blocks vary dramatically in size and shape. We used a convolutional locator netwo rk with four outputs, each trained to find a different corner of the a ddress block. A simple set of rules was used to generate ABL candidates from the network output. The system performs very well: when allowed five guesses , the network will tightly bound the address delivery information in 98.2% of the cases.

****

An Analog VLSI Saccadic Eye Movement System

Timothy Horiuchi, Brooks Bishofberger, Christof Koch

In an effort to understand saccadic eye movements and their rela(cid:173) tio n to visual attention and other forms of eye movements, we - in collabora tion with a number of other laboratories - are carry(cid:173) ing out a larg e-scale effort to design and build a complete primate oculomotor system us ing analog CMOS VLSI technology. Using this technology, a low power, compa ct, multi-chip system has been built which works in real-time using real-w orld visual inputs. We describe in this paper the performance of an ear ly version of such a system including a 1-D array of photoreceptors mim icking the retina, a circuit computing the mean location of activity represent (cid:173) ing the superior colliculus, a saccadic burst generator, and a one degree-of-freedom rotational platform which models the dynamic pr operties of the primate oculomotor plant.

****

An Optimization Method of Layered Neural Networks based on the Modified Informat ion Criterion

Sumio Watanabe

This paper proposes a practical optimization method for layered neural networks , by which the optimal model and parameter can be found simultaneously. 'i\Te m odify the conventional information criterion into a differentiable function of parameters, and then, min(cid:173) imize it, while controlling it back to the or dinary form. Effective(cid:173) ness of this method is discussed theoretically a nd experimentally.

****

A Computational Model for Cursive Handwriting Based on the Minimization Principl e

Yasuhiro Wada, Yasuharu Koike, Eric Vatikiotis-Bateson, Mitsuo Kawato

We propose a trajectory planning and control theory for continuous move ments such as connected cursive handwriting and continuous natural spee ch. Its hardware is based on our previously proposed forward-inverse-re laxation neural network (Wada & Kawato, 1993). Computationally, its opti mization principle is the minimum torque(cid:173) change criterion. Regard ing the representation level, hard constraints satisfied by a trajectory are represented as a set of via-points extracted from a handwritten character. Accordingly, we propose a via-point estimation algorithm that estimates via-points by repeating the trajectory formation of a character and the via- point extraction from the character. In experiments, good quantitative ag reement is found between human handwriting data and the trajectories genera ted by the theory. Finally, we propose a recognition schema based on the movement generation. We show a result in which the recognition sc hema is applied to the handwritten character recognition and can be extended to the phoneme timing estimation of natural speech.

****

Discontinuous Generalization in Large Committee Machines

H. Schwarze, J. Hertz

The problem of learning from examples in multilayer networks is studied within the framework of statistical mechanics. Using the replica formalism we calculat e the average generalization error of a fully connected committee machine in th e limit of a large number of hidden units. If the number of training examples i s proportional to the number of inputs in the network, the generalization error as a function of the training set size approaches a finite value. If the numb er of training examples is proportional to the number of weights in the network we find first-order phase transitions with a discontinuous drop in the general ization error for both binary and continuous weights.

****

Inverse Dynamics of Speech Motor Control

Makoto Hirayama, Eric Vatikiotis-Bateson, Mitsuo Kawato

Progress ha.s been made in comput.ational implementation of speech production based on physiological dat.a. An inverse dynamics model of the speech articulator's l1111sculo-skeletal system. which is the mapping from art. iculator t.rajectories to e\ectromyogl'aphic (EMG) signals, was modeled u sing the acquired forward dynamics model and temporal (smoot.hness of EM G activation) and range constraints. This inverse dynamics model allows th e use of a faster speech mot.or control scheme, which can be applied to pho neme-to(cid:173) speech synthesis via musclo-skeletal system dynamics, or to future use in speech recognition. The forward acoustic model, which is the mapping from articulator trajectories t.o the acoustic parameters, was i mproved by adding velocity and voicing information inputs to distinguish acollst.ic paramet.er differences caused by changes in source characte rist.ics.

****

Learning Temporal Dependencies in Connectionist Speech Recognition

Steve Renals, Mike Hochberg, Tony Robinson

Hybrid connectionistfHMM systems model time both using a Markov chain and throu gh properties of a connectionist network. In this paper, we discuss the nature of the time dependence currently employed in our systems using recurrent networ ks (RNs) and feed-forward multi-layer perceptrons (MLPs). In particular, we int

roduce local recurrences into a MLP to produce an enhanced input representation . This is in the form of an adaptive gamma filter and incorporates an automatic approach for learning temporal dependencies. We have experimented on a speaker (cid:173) independent phone recognition task using the TIMIT database. Results using the gamma filtered input representation have shown improvement over the b aseline MLP system. Improvements have also been obtained through merging the ba seline and gamma filter models.

****

Odor Processing in the Bee: A Preliminary Study of the Role of Central Input to the Antennal Lobe

Christiane Linster, David Marsan, Claudine Masson, Michel Kerszberg

Based on precise anatomical data of the bee's olfactory system, we pro pose an investigation of the possible mechanisms of modulation and control betw een the two levels of olfactory information processing: the antennallobe glom eruli and the mushroom bodies. We use simplified neurons, but realistic architecture. As a first conclusion, we postulate that the feature extra ction performed by the antennallobe (glomeruli and interneurons) necessitates central input from the mushroom bodies for fine tuning. The central input thus facilitates the evolution from fuzzy olfactory images in the glomerular layer towards more focussed images upon odor presentation.

****

Surface Learning with Applications to Lipreading

Christoph Bregler, Stephen Omohundro

Most connectionist research has focused on learning mappings from one space to another (eg. classification and regression). This paper introduces the more general task of learning constraint surfaces. It describes a simple bu t powerful architecture for learning and manipulating nonlinear surfaces from data. We demonstrate the technique on low dimensional synthetic sur faces and compare it to nearest neighbor approaches. We then show its u tility in learning the space of lip images in a system for improving speech rec ognition by lip reading. This learned surface is used to improve the visu al tracking performance during recognition.

****

Figure of Merit Training for Detection and Spotting

Eric Chang, Richard P. Lippmann

Spotting tasks require detection of target patterns from a background of richly varied non-target inputs. The performance measure of interest for these tasks , called the figure of merit (FOM), is the detection rate for target p atterns when the false alarm rate is in an acceptable range. A new a pproach to training spotters is presented which computes the FOM gradient for each input pattern and then directly maximizes the FOM using b ackprop agati on. This eliminates the need for thresholds during training. It al so uses network resources to model Bayesian a posteriori probability funct ions accurately only for patterns which have a significant effect on th e detection accuracy over the false alarm rate of interest. FOM training increased detection accuracy by 5 percentage points for a hybrid radial ba sis function (RBF) - hidden Markov model (HMM) wordspotter on the credit-card speech corpus.

****

Generation of Internal Representation by α-Transformation

Ryotaro Kamimura

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.

****

Autoencoders, Minimum Description Length and Helmholtz Free Energy

Geoffrey E. Hinton, Richard Zemel

An autoencoder network uses a set of recognition weights to convert an input ve ctor into a code vector. It then uses a set of generative weights to convert t

he code vector into an approximate reconstruction of the input vector. We derive an objective function for training autoencoders based on the Minimum Description Length (MDL) principle. The aim is to minimize the information required to describe both the code vector and the reconstruction error. We show that this information is minimized by choosing code vectors stochastically according to a Boltzmann distri(cid:173) bution, where the generative weights define the energy of each possible code vector given the input vector. Unfortunately, if the code vectors use distributed representations, it is exponentially expensive to compute this Boltzmann distribution because it involves all possible code vectors. We show that the recognition weights of an autoencoder can be used to compute an approximation to the Boltzmann distribution and that this ap(cid:173) proximation gives an upper bound on the description length. Even when this bound is poor, it can be used as a Lyapunov function for learning both the generative and the recognition weights. We demonstrate that this approach can be used to learn factorial codes.

****

Dual Mechanisms for Neural Binding and Segmentation

Paul Sajda, Leif Finkel

We propose that the binding and segmentation of visual features is mediated by two complementary mechanisms; a low resolu(cid:173) tion, spatial-based, resource-free process and a high resolution, temporal-based, resource-limited process. In the visual cortex, the former depends upon the orderly topographic organization in stri(cid:173) ate and extrastriate areas while the latter may be related to ob(cid:173) served temporal relationships between neuronal activities . Com(cid:173) puter simulations illustrate the role the two mechanisms play in figure/ ground discrimination, depth-from-occlusion, and the vivid(cid:173) ness of perceptual completion.

****

The Statistical Mechanics of k-Satisfaction

Scott Kirkpatrick, Géza Györgyi, Naftali Tishby, Lidror Troyansky

The satisfiability of random CNF formulae with precisely k vari(cid:173) ables per clause ("k-SAT") is a popular testbed for the performance of search algorithms. Formulae have M clauses from N variables, randomly negated, keeping the ratio a = M / N fixed . For k = 2, this model has been proven to have a sharp threshold at a = 1 between formulae which are almost aways satisfiable and formulae which are almost never satisfiable as N --jo 00 . Computer experi(cid:173) ments for k = 2, 3, 4, 5 and 6, (carried out in collaboration with B. Selman of ATT Bell Labs). show similar threshold behavior for each value of k. Finite-size scaling, a theory of the critical point phenomena used in statistical physics, is shown to characterize the size dependence near the threshold. Annealed and replica-based mean field theories give a good account of the results.

****

Lipreading by neural networks: Visual preprocessing, learning, and sensory integration

Gregory Wolff, K. Prasad, David Stork, Marcus Hennecke

We have developed visual preprocessing algorithms for extracting phonologically relevant features from the grayscale video image of a speaker, to provide speaker-independent inputs for an automat(cid:173) ic lipreading ("speechreading") system. Visual features such as mouth open/closed, tongue visible/not-visible, teeth visible/not(cid:173) visible, and several shape descriptors of the mouth and its motion are all rapidly computable in a manner quite insensitive to lighting conditions. We formed a hybrid speechreading system consisting of two time delay neural networks (video and acoustic) and inte(cid:173) grated their responses by means of independent opinion pooling - the Bayesian optimal method given conditional independence, which seems to hold for our data. This hybrid system had an er(cid:173) ror rate 25% lower than that of the acoustic subsystem alone on a five-utterance speaker-independent task, indicating that video can be used to improve speech recognition.

****

Dynamic Modulation of Neurons and Networks
Eve Marder
Biological neurons have a variety of intrinsic properties because of the large number of voltage dependent currents that control their activity. Neuromodulatory substances modify both the balance of conductances that determine intrinsic properties and the strength of synapses. These mechanisms alter circuit dynamics, and suggest that functional circuits exist only in the modulatory environment in which they operate.
****

Training Neural Networks with Deficient Data
Volker Tresp, Subutai Ahmad, Ralph Neuneier
We analyze how data with uncertain or missing input features can be incorporated into the training of a neural network. The gen(cid:173) eral solution requires a weighted integration over the unknown or uncertain input although computationally cheaper closed-form so(cid:173) lutions can be found for certain Gaussian Basis Function (GBF) networks. We also discuss cases in which heuristical solutions such as substituting the mean of an unknown input can be harmful.
****

When will a Genetic Algorithm Outperform Hill Climbing
Melanie Mitchell, John Holland, Stephanie Forrest
We analyze a simple hill-climbing algorithm (RMHC) that was pre(cid:173) viously shown to outperform a genetic algorithm (GA) on a simple "Royal Road" function. We then analyze an "idealized" genetic algorithm (IGA) that is significantly faster than RMHC and that gives a lower bound for GA speed. We identify the features of the IGA that give rise to this speedup, and discuss how these features can be incorporated into a real GA.
****

Unsupervised Parallel Feature Extraction from First Principles
Mats Österberg, Reiner Lenz
We describe a number of learning rules that can be used to train un(cid:173) supervised parallel feature extraction systems. The learning rules are derived using gradient ascent of a quality function. We con(cid:173) sider a number of quality functions that are rational functions of higher order moments of the extracted feature values. We show that one system learns the principle components of the correla(cid:173) tion matrix. Principal component analysis systems are usually not optimal feature extractors for classification. Therefore we design quality functions which produce feature vectors that support unsu(cid:173) pervised classification. The properties of the different systems are compared with the help of different artificially designed datasets and a database consisting of all Munsell color spectra.
****

Classification of Electroencephalogram using Artificial Neural Networks
A C Tsoi, D S C So, A Sergejew
In this paper, we will consider the problem of classifying electroencephalo(cid:173) gram (EEG) signals of normal subjects, and subjects suffering from psychi(cid:173) atric disorder, e.g., obsessive compulsive disorder, schizophrenia, using a class of artificial neural networks, viz., multi-layer perceptron. It is shown that the multilayer perceptron is capable of classifying unseen test EEG signals to a high degree of accuracy.
****

Comparison Training for a Rescheduling Problem in Neural Networks
Didier Keymeulen, Martine de Gerlache
Airline companies usually schedule their flights and crews well in advance to optimize their crew pools activities. Many events such as flight delays or the absence of a member require the crew pool rescheduling team to change the initial schedule (rescheduling). In this paper, we show that the neural network comparison paradigm applied to the backgammon game by Tesauro (Tesauro and Se(cid:173) jnowski, 1989) can also be applied

to the rescheduling problem of an aircrew pool. Indeed both problems cor respond to choosing the best solut.ion from a set of possible ones without r anking them (called here best choice problem). The paper explains from a math(cid:173) ematical point of view the architecture and the learning strategy of the backpropagation neural network used for the best choice prob(cid:173) lem . We also show how the learning phase of the network can be accelerat ed. Finally we apply the neural network model to some real rescheduling problems for the Belgian Airline (Sabena).

****

Bounds on the complexity of recurrent neural network implementations of finite s tate machines

Bill Horne, Don Hush

In this paper the efficiency of recurrent neural network implementa(cid:173) tio ns of m-state finite state machines will be explored. Specifically, it will be shown that the node complexity for the unrestricted case can be bounded above b y 0 (fo) . It will also be shown that the node complexity is 0 (y'm log m) whe n the weights and thresholds are restricted to the set {-I, I}, and 0 (m) when the fan-in is re(cid:173) stricted to two. Matching lower bounds will be provide d for each of these upper bounds assuming that the state of the FSM can be enc oded in a subset of the nodes of size rlog m 1.

****

Classification of Multi-Spectral Pixels by the Binary Diamond Neural Network

Yehuda Salu

A new neural network, the Binary Diamond, is presented and its use as a classi fier is demonstrated and evaluated. The network is of the feed-forward type . It learns from examples in the 'one shot' mode, and recruits new neuron s as needed. It was tested on the problem of pixel classification, and performe d well. Possible applications of the network in associative memories are outli ned.

****

Optimal Brain Surgeon: Extensions and performance comparisons

Babak Hassibi, David Stork, Gregory Wolff

a second-order

****

Agnostic PAC-Learning of Functions on Analog Neural Nets

Wolfgang Maass

There exist a number of negative results ([J), [BR), [KV]) about learning on neural nets in Valiant's model [V) for probably approx(cid:173) imately correct learning ("PAC-learning"). These negative results are based on an asymptotic analysis where one lets the number of nodes in the neural net go to infinit.y. Hence this analysis is less ad(cid:173) equate for the investigation of learning on a small fixed neural net. with relatively few analog inputs (e.g. the principal components of some sensory data). The latter type of learning problem gives rise to a different kind of asymptotic question: Can the true error of the neural net be brought arbitrarily close to that of a neural net with "optimal" weights through sufficiently long training? In this paper we employ some new arguments ill order to give a positive answer to this question in Haussler's rather realistic refinement of Valiant's model for PAC-learning ([H), [KSS)). In this more realistic model no a-priori assumptions are required about the "learning target" , noise is permitted in the training data, and the inputs and outputs are not restricted to boolean values. As a special case our result implies one of the first positive results about learning on multi-layer neural net.s in Valiant's original PAC-learning model. At the end of this paper we will describe an efficient parallel implementation of this new learning algorit.hm.

****

Mixtures of Controllers for Jump Linear and Non-Linear Plants

Timothy Cacciatore, Steven Nowlan
We describe an extension to the Mixture of Experts architecture for modelling a nd controlling dynamical systems which exhibit multi(cid:173) ple modes of be havior. This extension is based on a Markov process model, and suggests a r ecurrent network for gating a set of linear or non-linear controllers. Th e new architecture is demonstrated to be capable of learning effective control strategies for jump linear and non-linear plants with multiple mod es of behavior.
****

A Hybrid Radial Basis Function Neurocomputer and Its Applications
Steven Watkins, Paul Chau, Raoul Tawel, Bjorn Lambrigtsen, Mark Plutowski
A neurocomputer was implemented using radial basis functions and a combinatio n of analog and digital VLSI circuits. The hybrid system uses custom analog circuits for the input layer and a digital signal processing b oard for the hidden and output layers. The system combines the advantages of b oth analog and digital circuits. featuring low power consumption while minimizi ng overall system error. The analog circuits have been fabricated and tested, t he system has been built, and several applications have been executed on the system. One application provides significantly better results for a re mote sensing problem than have been previously obtained using conventional meth ods.
****

Resolving motion ambiguities
K. I. Diamantaras, D. Geiger
We address the problem of optical flow reconstruction and in par(cid:173) ticula r the problem of resolving ambiguities near edges. They oc(cid:173) cur due to (i) the aperture problem and (ii) the occlusion problem, where pixels on both si des of an intensity edge are assigned the same velocity estimates (and confiden ce). However, these measurements are correct for just one side of the edge (the non occluded one). Our approach is to introduce an uncertamty field with respe ct to the estimates and confidence measures. We note that the confi(cid:173) de nce measures are large at intensity edges and larger at the con(cid:173) vex sid es of the edges, i.e. inside corners, than at the concave side. We resolve the ambiguities through local interactions via coupled Markov random fields (MRF) . The result is the detection of motion for regions of images with large global convexity.
****

Developing Population Codes by Minimizing Description Length
Richard Zemel, Geoffrey E. Hinton
The Minimum Description Length principle (MDL) can be used to train the hidden units of a neural network to extract a representa(cid:173) tion that is cheap to describe but nonetheless allows the input to be reconstructed accurately. We s how how MDL can be used to develop highly redundant population codes. Each hidd en unit has a location in a low-dimensional implicit space. If the hidden unit activities form a bump of a standard shape in this space, they can be cheaply encoded by the center ofthis bump. So the weights from the input units to the h idden units in an autoencoder are trained to make the activities form a standar d bump. The coordinates of the hidden units in the implicit space are also lear ned, thus allow(cid:173) ing flexibility, as the network develops a discontinuou s topography when presented with different input classes. Population-coding in a space other than the input enables a network to extract nonlinear higher-ord er properties of the inputs.
****

Learning in Computer Vision and Image Understanding
Hayit Greenspan
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.
****

Neural Network Definitions of Highly Predictable Protein Secondary Structure Cla
sses
Alan Lapedes, Evan Steeg, Robert Farber
We use two co-evolving neural networks to determine new classes of protein seco
ndary structure which are significantly more pre(cid:173) dictable from local am
ino sequence than the conventional secondary structure classification. Accurate
prediction of the conventional secondary structure classes: alpha helix, beta
strand, and coil, from primary sequence has long been an important problem in c
ompu(cid:173) tational molecular biology. Neural networks have been a popular m
ethod to attempt to predict these conventional secondary struc(cid:173) ture cla
sses. Accuracy has been disappointingly low. The algo(cid:173) rithm presented h
ere uses neural networks to similtaneously exam(cid:173) ine both sequence and s
tructure data, and to evolve new classes of secondary structure that can be pre
dicted from sequence with significantly higher accuracy than the conventional c
lasses. These new classes have both similarities to, and differences with the c
on(cid:173) ventional alpha helix, beta strand and coil.
****

Coupled Dynamics of Fast Neurons and Slow Interactions
A.C.C. Coolen, R. Penney, D. Sherrington
A simple model of coupled dynamics of fast neurons and slow inter(cid:173) actio
ns, modelling self-organization in recurrent neural networks, leads naturally t
o an effective statistical mechanics characterized by a partition function whic
h is an average over a replicated system. This is reminiscent of the replica tr
ick used to study spin-glasses, but with the difference that the number of repl
icas has a physi(cid:173) cal meaning as the ratio of two temperatures and can b
e varied throughout the whole range of real values. The model has inter(cid:173
) esting phase consequences as a function of varying this ratio and external st
imuli, and can be extended to a range of other models.
****

Using Local Trajectory Optimizers to Speed Up Global Optimization in Dynamic Pro
gramming
Christopher Atkeson
Dynamic programming provides a methodology to develop planners and controllers
for nonlinear systems. However, general dynamic programming is computati
onally intractable. We have developed procedures that allow more complex pla
nning and control problems to be solved. We use second order local traj
ectory optimization to generate locally optimal plans and local models
of the value function and its derivatives. We maintain global consistency
of the local models of the value function, guaranteeing that our local
ly optimal plans are actually globally optimal, up to the resolution of our s
earch procedures.
****

Connectionist Modeling and Parallel Architectures
Joachim Diederich, Ah Tsoi
The introduction of specialized hardware platforms for connectionist mod
eling
("connectionist supercomputer") has created a number of research topics. S
ome of
these issues are controversial, e.g. the efficient implementation of incrementa
l learn(cid:173)
ing techniques, the need for the dynamic reconfiguration of networks and possibl
e
programming environments for these machines.
****

Structural and Behavioral Evolution of Recurrent Networks
Gregory Saunders, Peter Angeline, Jordan Pollack
This paper introduces GNARL, an evolutionary program which induces recurrent ne
ural networks that are structurally unconstrained. In contrast to constructive
and destructive algorithms, GNARL employs a popula(cid:173) tion of networks and
uses a fitness function's unsupervised feedback to guide search through n

etwork space. Annealing is used in generating both gaussian weight changes and structural modifications. Applying GNARL to a complex search and collection task demonstrates that the system is capable of inducing networks with complex internal dynamics.
****

Two-Dimensional Object Localization by Coarse-to-Fine Correlation Matching
Chien-Ping Lu, Eric Mjolsness
We present a Mean Field Theory method for locating two(cid:173) dimensional objects that have undergone rigid transformations. The resulting algorithm is a form of coarse-to-fine correlation matching. We first consider problems of matching synthetic point data, and derive a point matching objective function. A tractable line segment matching objective function is derived by considering each line segment as a dense collection of points, and approximat(cid:173) ing it by a sum of Gaussians. The algorithm is tested on real images from which line segments are extracted and matched.
****

WATTLE: A Trainable Gain Analogue VLSI Neural Network
Richard Coggins, Marwan Jabri
This paper describes a low power analogue VLSI neural network called Wattle. Wattle is a 10:6:4 three layer perceptron with multi(cid:173) plying DAC synapses and on chip switched capacitor neurons fabri(cid:173) cated in 1.2um CMOS. The on chip neurons facillitate variable gain per neuron and lower energy/connection than for previous designs. The intended application of this chip is Intra Cardiac Electrogram classification as part of an implantable pacemaker / defibrillator sys(cid:173) tem. Measurements of t.he chip indicate that 10pJ per connection is achievable as part of an integrated system. Wattle has been suc(cid:173) cessfully trained in loop on parity 4 and ICEG morphology classi(cid:173) fication problems.
****

Neurobiology, Psychophysics, and Computational Models of Visual Attention
Ernst Niebur, Bruno Olshausen
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.
****

Bayesian Backpropagation Over I-O Functions Rather Than Weights
David H. Wolpert
The conventional Bayesian justification of backprop is that it finds the MAP weight vector. As this paper shows, to find the MAP i-o function instead one must add a correction tenn to backprop. That tenn biases one towards i-o functions with small description lengths, and in particular fa(cid:173) vors (some kinds of) feature-selection, pruning, and weight-sharing.
****

Fast Non-Linear Dimension Reduction
Nanda Kambhatla, Todd Leen
We present a fast algorithm for non-linear dimension reduction. The algorithm builds a local linear model of the data by merging PCA with clustering based on a new distortion measure. Exper(cid:173) iments with speech and image data indicate that the local linear algorithm produces encodings with lower distortion than those built by five layer auto-associative networks. The local linear algorithm is also more than an order of magnitude faster to train.
****

Observability of Neural Network Behavior
Max Garzon, Fernanda Botelho
We prove that except possibly for small exceptional sets, discrete(cid:173) time analog neural nets are globally observable, i.e. all their cor(cid:173) rupted pseudo-orbits on computer simulations actually reflect the true dynamical behavior of the network. Locally finite discrete (boolean) neural networks are obser

vable without exception.
****
Optimal Signalling in Attractor Neural Networks
Isaac Meilijson, Eytan Ruppin
In [Meilijson and Ruppin, 1993] we presented a methodological framework describing the two-iteration performance of Hopfield(cid:173) like attractor neural networks with history-dependent, Bayesian dynamics. We now extend this analysis in a number of directions: input patterns applied to small subsets of neurons, general con(cid:173) nectivity architectures and more efficient use of history. We show that the optimal signal (activation) function has a slanted sigmQidal shape, and provide an intuitive account of activation functions with a non-monotone shape. This function endows the model with some properties characteristic of cortical neurons' firing.
****
Neural Network Exploration Using Optimal Experiment Design
David Cohn
Consider the problem of learning input/output mappings through exploration, e.g. learning the kinematics or dynamics of a robotic manipulator. If actions are expensive and computation is cheap, then we should explore by selecting a trajectory through the in(cid:173) put space which gives us the most amount of information in the fewest number of steps. I discuss how results from the field of opti(cid:173) mal experiment design may be used to guide such exploration, and demonstrate its use on a simple kinematics problem.
****
Directional Hearing by the Mauthner System
Audrey Guzik, Robert Eaton
We provide a computational description of the function of the Mau(cid:173) thner system. This is the brainstem circuit which initiates fast(cid:173) start escapes in teleost fish in response to sounds. Our simula(cid:173) tions, using back propagation in a realistically constrained feedfor(cid:173) ward network, have generated hypotheses which are directly inter(cid:173) pretable in terms of the activity of the auditory nerve fibers, the principle cells of the system and their associated inhibitory neu(cid:173) rons.
****
Asynchronous Dynamics of Continuous Time Neural Networks
Xin Wang, Qingnan Li, Edward Blum
Motivated by mathematical modeling, analog implementation and distributed simulation of neural networks, we present a definition of asynchronous dynamics of general CT dynamical systems defined by ordinary differential equations, based on notions of local times and communication times. We provide some preliminary results on globally asymptotical convergence of asynchronous dynamics for contractive and monotone CT dynamical systems. When ap(cid:173) plying the results to neural networks, we obtain some conditions that ensure additive-type neural networks to be asynchronizable.
****
Development of Orientation and Ocular Dominance Columns in Infant Macaques
Klaus Obermayer, Lynne Kiorpes, Gary Blasdel
Maps of orientation preference and ocular dominance were recorded optically from the cortices of 5 infant macaque monkeys, ranging in age from 3.5 to 14 weeks. In agreement with previous observations, we found that basic features of orientation and ocular dominance maps, as well as correlations between them, are present and robust by 3.5 weeks of age. We did observe changes in the strength of ocular dominance signals, as well as in the spacing of ocular dom(cid:173) inance bands, both of which increased steadily between 3.5 and 14 weeks of age. The latter finding suggests that the adult spacing of ocular dominance bands depends on cortical growth in neonatal animals. Since we found no corresponding increase in the spacing of orientation preferences, however, there is a possibility that the orientation prefer

ences of some cells change as the cortical surface expands. Since correlations between the patterns of orientation selectivity and ocular dominance are present at an age, when the visual system is still immature, it seems more likely that their de(cid:173) velopment may be an innate process and may not require extensive visual experience.

****

Counting function theorem for multi-layer networks

Adam Kowalczyk

We show that a randomly selected N-tuple x of points ofRn with probability> 0 is such that any multi-layer percept ron with the first hidden layer composed of hi threshold logic units can imple- ment exactly 2 2:~~~ (Nil) different dichotomies of x. If N > hin then such a perceptron must have all units of the first hidden layer fully connected to inputs. This implies the maximal capacities (in the sense of Cover) of 2n input patterns per hidden unit and 2 input patterns per synaptic weight of such networks (both capacities are achieved by networks with single hidden layer and are the same as for a single neuron). Comparing these results with recent estimates of VC-dimension we find that in contrast to the single neuron case, for sufficiently large nand hl, the VC-dimension exceeds Cover's capacity.

****

Supervised Learning with Growing Cell Structures

Bernd Fritzke

We present a new incremental radial basis function network suit(cid:173) able for classification and regression problems. Center positions are continuously updated through soft competitive learning. The width of the radial basis functions is derived from the distance to topological neighbors. During the training the observed error is accumulated locally and used to determine where to insert the next unit. This leads (in case of classification problems) to the placement of units near class borders rather than near frequency peaks as is done by most existing methods. The resulting networks need few training epochs and seem to generalize very well. This is demonstrated by examples.

****

A Learning Analog Neural Network Chip with Continuous-Time Recurrent Dynamics

Gert Cauwenberghs

We present experimental results on supervised learning of dynam(cid:173) ical features in an analog VLSI neural network chip. The recur(cid:173) rent network, containing six continuous-time analog neurons and 42 free parameters (connection strengths and thresholds), is trained to generate time-varying outputs approximating given periodic signals presented to the network. The chip implements a stochastic pertur(cid:173) bative algorithm, which observes the error gradient along random directions in the parameter space for error-descent learning. In ad(cid:173) dition to the integrated learning functions and the generation of pseudo-random perturbations, the chip provides for teacher forc(cid:173) ing and long-term storage of the volatile parameters. The network learns a 1 kHz circular trajectory in 100 sec. The chip occupies 2mm x 2mm in a 2JLm CMOS process, and dissipates 1.2 m W.

****

Functional Models of Selective Attention and Context Dependency

Thomas Hildebrandt

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.

****

Learning Classification with Unlabeled Data

Virginia de

One of the advantages of supervised learning is that the final error met(cid:173) ric is available during training. For classifiers, the algorithm can directly reduce the number of misclassifications on the training set. Unfortu(

cid:173) nately, when modeling human learning or constructing classifiers for au (cid:173) tonomous robots, supervisory labels are often not available or too ex(cid:173) pensive. In this paper we show that we can substitute for t he labels by making use of structure between the pattern distributions to diffe rent sen(cid:173) sory modalities. We show that minimizing the disagreement bet ween the outputs of networks processing patterns from these different modalitie s is a sensible approximation to minimizing the number of misclassifications i n each modality, and leads to similar results. Using the Peterson-Barney vowel dataset we show that the algorithm performs well in finding ap(cid:173) pr opriate placement for the codebook vectors particularly when the con(cid:173) fu seable classes are different for the two modalities.

****

Temporal Difference Learning of Position Evaluation in the Game of Go

Nicol Schraudolph, Peter Dayan, Terrence J. Sejnowski

The game of Go has a high branching factor that defeats the tree search appro ach used in computer chess, and long-range spa(cid:173) tiotemporal inter actions that make position evaluation extremely difficult. Development of conv entional Go programs is hampered by their knowledge-intensive nature. We de monstrate a viable alternative by training networks to evaluate Go positions v ia tem(cid:173) poral difference (TD) learning. Our approach is based on ne twork architectures that reflect the spatial organization of both input and reinforcement signals on the Go board, and training protocols that pro vide exposure to competent (though unlabelled) play. These techniques yiel d far better performance than undifferentiated networks trained by self(cid:173) play alone. A network with less than 500 weights learned within 3,000 games of 9x9 Go a position evaluation function that enables a primitive one-ply sear ch to defeat a commercial Go program at a low playing level.

****

Recovering a Feed-Forward Net From Its Output

Charles Fefferman, Scott Markel

We study feed-forward nets with arbitrarily many layers, using the stan(cid:173) dard sigmoid, tanh x. Aside from technicalities, our theorems are: 1. Complete knowledge of the output of a neural net for arbitrary inputs uniq uely specifies the architecture, weights and thresholds; and 2. There are only finitely many critical points on the error surface for a generic training problem.

****

Optimal Stochastic Search and Adaptive Momentum

Todd Leen, Genevieve Orr

Stochastic optimization algorithms typically use learning rate schedules that behave asymptotically as J.t(t) = J.to/t. The ensem(cid:173) ble dyna mics (Leen and Moody, 1993) for such algorithms provides an easy path to resu lts on mean squared weight error and asymp(cid:173) totic normality. We ap ply this approach to stochastic gradient algorithms with momentum. We sh ow that at late times, learning is governed by an effective learning rate J.tejJ = J.to/(l - f3) where f3 is the momentum parameter. We descri be the behavior of the asymptotic weight error and give conditions on J.tejJ that insure optimal convergence speed. Finally, we use the resu lts to develop an adaptive form of momentum that achieves optimal convergence speed independent of J.to.

****

The Parti-Game Algorithm for Variable Resolution Reinforcement Learning in Multi dimensional State-Spaces

Andrew Moore

Parti-game is a new algorithm for learning from delayed rewards in high dimensional real-valued state-spaces. In high dimensions it is essenti al that learning does not explore or plan over state space uniformly. Part i-game maintains a decision-tree partitioning of state-space and app lies game-theory and computational geom(cid:173) etry techniques to efficie ntly and reactively concentrate high reso(cid:173) lution only on critical

areas. Many simulated problems have been tested, ranging from 2-dimensio
nal to 9-dimensional state-spaces, including mazes, path planning, non-line
ar dynamics, and uncurl(cid:173) ing snake robots in restricted spaces. I
n all cases, a good solution is found in less than twenty trials and
a few minutes.
****

Complexity Issues in Neural Computation and Learning
V. P. Roychowdhury, K.-Y. Siu
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.
****

Stability and Observability
Max Garzon, Fernanda Botelho
he theme was the effect of perturbations of the defining parameters of a neural
net(cid:173)
work due to: 1) mea"urement" (particularly with analog networks); 2) di"creti
zation
due to a) digital implementation of analog nets; b) bounded-precision implemen
ta(cid:173)
tion of digital networks; or c) inaccurate evaluation of the transfer functi
on(s}; 3)
noise in or incomplete input and/or output of the net or individual cells (part
icu(cid:173)
larly with analog networks).
****

The Role of MT Neuron Receptive Field Surrounds in Computing Object Shape from V
elocity Fields
G. Buracas, T. Albright
The goal of this work was to investigate the role of primate MT neurons in s
olving the structure from motion (SFM) problem. Three types of recepti
ve field (RF) surrounds found in area MT neurons (K.Tanaka et al.,1986; Allm
an et al.,1985) correspond, as our analysis suggests, to the oth, pt and 2nd
order fuzzy space-differential operators. The large surround/center both d
ifferentiation of smooth velocity fields and discontinuity detection at
boundaries of objects. The model is in agreement with recent psychoph
ysical data on surface interpolation involvement in SFM. We suggest tha
t area MT partially segregates information about object shape from inf
ormation about spatial relations necessary for navigation and manipulation
.
****

Estimating analogical similarity by dot-products of Holographic Reduced Represen
tations
Tony A. Plate
Models of analog retrieval require a computationally cheap method of estimating
similarity between a probe and the candidates in a large pool of memory items.
The vector dot-product operation would be ideal for this purpose if it were po
ssible to encode complex structures as vector representations in such a way tha
t the superficial similarity of vector representations reflected underlying str
uctural similarity. This paper de(cid:173) scribes how such an encoding is provi
ded by Holographic Reduced Rep(cid:173) resentations (HRRs), which are a method
for encoding nested relational structures as fixed-width distributed representa
tions. The conditions un(cid:173) der which structural similarity is reflected i
n the dot-product rankings of HRRs are discussed.
****

Segmental Neural Net Optimization for Continuous Speech Recognition
Ying Zhao, Richard Schwartz, John Makhoul, George Zavaliagkos
Previously, we had developed the concept of a Segmental Neural Net (SNN) for ph
onetic modeling in continuous speech recognition (CSR). This kind of neu(cid:173

) ral network technology advanced the state-of-the-art of large-vocabulary CSR, which employs Hidden Marlcov Models (HMM), for the ARPA 1oo0-word Re(cid:173) source Management corpus. More Recently, we started porting the neural net syste m to a larger, more challenging corpus - the ARPA 20,Ooo-word Wall Street Journ al (WSJ) corpus. During the porting, we explored the following research directi ons to refine the system: i) training context-dependent models with a reg(cid:17 3) ularization method; ii) training SNN with projection pursuit; and ii) combini ng different models into a hybrid system. When tested on both a development set and an independent test set, the resulting neural net system alone yielded a p er(cid:173) fonnance at the level of the HMM system, and the hybrid SNN/HMM syst em achieved a consistent 10-15% word error reduction over the HMM system. This paper describes our hybrid system, with emphasis on the optimization methods e mployed.

****

Adaptive knot Placement for Nonparametric Regression

Hossein L. Najafi, Vladimir Cherkassky

We show how an "Elman" network architecture, constructed from recurrently conne cted oscillatory associative memory network mod(cid:173) ules, can employ select ive "attentional" control of synchronization to direct the flow of communicatio n and computation within the architecture to solve a grammatical inference prob lem. Previously we have shown how the discrete time "Elman" network algorithm can be implemented in a network completely described by continuous ordinary dif ferential equations. The time steps (ma(cid:173) chine cycles) of the system are implemented by rhythmic variation (clocking) of a bifurcation parameter. In th is architecture, oscilla(cid:173) tion amplitude codes the information content o r activity of a mod(cid:173) ule (unit), whereas phase and frequency are used to "softwire" the network. Only synchronized modules communicate by exchang(cid:1 73) ing amplitude information; the activity of non-resonating modules contribut es incoherent crosstalk noise. Attentional control is modeled as a special subs et of the hidden modules with ouputs which affect the resonant frequencies of o ther hidden modules. They control synchrony among the other mod(cid:173) ules a nd direct the flow of computation (attention) to effect transi(cid:173) tions be tween two subgraphs of a thirteen state automaton which the system emulates to generate a Reber grammar. The internal crosstalk noise is used to drive the req uired random transitions of the automaton.

****

Bayesian Modeling and Classification of Neural Signals

Michael Lewicki

Signal processing and classification algorithms often have limited applica bility resulting from an inaccurate model of the signal's un(cid:173) derlying structure. We present here an efficient, Bayesian algo(cid:173) rithm for modeling a signal composed of the superposition of brief, Poisson-distributed functions. This methodology is applied to the specific problem of modelin g and classifying extracellular neural waveforms which are composed of a superposition of an unknown number of action potentials CAPs). Previous ap proaches have had limited success due largely to the problems of determining th e spike shapes, deciding how many are shapes distinct, and decomposing overlapping APs. A Bayesian solution to each of these problems is obtained by inferring a probabilistic model of the waveform. This approach quantifie s the uncertainty of the form and number of the inferred AP shapes and is used to obtain an efficient method for decomposing complex overlaps. Th is algorithm can extract many times more information than previous methods and facilitates the extracellular investigation of neuronal classes and o f interactions within neuronal circuits.

****

Supervised learning from incomplete data via an EM approach

Zoubin Ghahramani, Michael Jordan

Real-world learning tasks may involve high-dimensional data sets with arb itrary patterns of missing data. In this paper we present a framework based on maximum likelihood density estimation for learning from such data

set.s. VVe use mixture models for the den(cid:173) sity estimates and ma ke two distinct appeals to the Expectation(cid:173) Maximization (EM) pri nciple (Dempster et al., 1977) in deriving a learning algorithm-EM is used both for the estimation of mix(cid:173) ture components and for cop ing wit.h missing dat.a. The result(cid:173) ing algorithm is applicable t.o a wide range of supervised as well as unsupervised learning prob lems. Result.s from a classification benchmark-t.he iris data set-are pres ented.

****

Correlation Functions in a Large Stochastic Neural Network

Iris Ginzburg, Haim Sompolinsky

Most theoretical investigations of large recurrent networks focus on the proper ties of the macroscopic order parameters such as popu(cid:173) lation averaged a ctivities or average overlaps with memories. How(cid:173) ever, the statistics o f the fluctuations in the local activities may be an important testing ground f or comparison between models and observed cortical dynamics. We evaluated the n euronal cor(cid:173) relation functions in a stochastic network comprising of ex citatory and inhibitory populations. We show that when the network is in a sta tionary state, the cross-correlations are relatively weak, i.e., their amplitud e relative to that of the auto-correlations are of or(cid:173) der of 1/ N, N be ing the size of the interacting population. This holds except in the neighborho ods of bifurcations to nonstationary states. As a bifurcation point is approach ed the amplitude of the cross-correlations grows and becomes of order 1 and the decay time(cid:173) constant diverges. This behavior is analogous to the phenom enon of critical slowing down in systems at thermal equilibrium near a critica l point. Near a Hopf bifurcation the cross-correlations ex(cid:173) hibit damped oscillations.

****

Clustering with a Domain-Specific Distance Measure

Steven Gold, Eric Mjolsness, Anand Rangarajan

With a point matching distance measure which is invariant under translatio n, rotation and permutation, we learn 2-D point-set ob(cid:173) jects, by clustering noisy point-set images. Unlike traditional clus(cid:173) tering methods which use distance measures that operate on feature vectors - a represe ntation common to most problem domains - this object-based clustering techniqu e employs a distance measure spe(cid:173) cific to a type of object within a problem domain. Formulating the clustering problem as two nested objec tive functions, we derive optimization dynamics similar to the Expectatio n-Maximization algorithm used in mixture models.

****

Speaker Recognition Using Neural Tree Networks

Kevin Farrell, Richard Mammone

A new classifier is presented for text-independent speaker recognition. The new classifier is called the modified neural tree network (MNTN). The NTN is a hie rarchical classifier that combines the properties of decision trees and feed-fo rward neural networks. The MNTN differs from the stan(cid:173) dard NTN in that a new learning rule based on discriminant learning is used, which minimizes the classification error as opposed to a norm of the approximation error. The MNTN also uses leaf probability mea(cid:173) sures in addition to the class labels. The MNTN is evaluated for several speaker identification experiments and is com pared to multilayer percep(cid:173) trons (MLPs) , decision trees, and vector qu antization (VQ) classifiers. The VQ classifier and MNTN demonstrate comparable performance and per(cid:173) form significantly better than the other classifier s for this task. Addition(cid:173) ally, the MNTN provides a logarithmic saving in retrieval time over that of the VQ classifier. The MNTN and VQ classifiers a re also compared for several speaker verification experiments where the MNTN is found to outperform the VQ classifier.

****

Tonal Music as a Componential Code: Learning Temporal Relationships Between and Within Pitch and Timing Components

Catherine Stevens, Janet Wiles
This study explores the extent to which a network that learns the temporal relationships within and between the component features of Western tonal music can account for music theoretic and psychological phenomena such as the tonal hierarchy and rhythmic expectancies. Predicted and generated sequences were recorded as the representation of a 153-note waltz melody was learnt by a predictive, recurrent network. The network learned transitions and relations between and within pitch and timing components: accent and duration values interacted in the development of rhythmic and metric structures and, with training, the network developed chordal expectancies in response to the activation of individual tones. Analysis of the hidden unit representation revealed that musical sequences are represented as transitions between states in hidden unit space.
****

Connectionist Models for Auditory Scene Analysis
Richard Duda
Although the visual and auditory systems share the same basic tasks of informing an organism about its environment, most con(cid:173)nectionist work on hearing to date has been devoted to the very different problem of speech recognition. VVe believe that the most fundamental task of the auditory system is the analysis of acoustic signals into components corresponding to individual sound sources, which Bregman has called auditory scene analysis. Computational and connectionist work on auditory scene analysis is reviewed, and the outline of a general model that includes these approaches is described.
****

Learning in Compositional Hierarchies: Inducing the Structure of Objects from Data
Joachim Utans
I propose a learning algorithm for learning hierarchical models for ob(cid:173)ject recognition. The model architecture is a compositional hierarchy that represents part-whole relationships: parts are described in the lo(cid:173)cal context of substructures of the object. The focus of this report is inducing the structure of learning hierarchical models from data, i.e. model prototypes from observed exemplars of an object. At each node in the hierarchy, a probability distribution governing its parameters must be learned. The connections between nodes reflects the structure of the object. The formulation of substructures is encouraged such that their parts become conditionally independent. The resulting model can be interpreted as a Bayesian Belief Network and also is in many respects similar to the stochastic visual grammar described by Mjolsness.
****

VLSI Phase Locking Architectures for Feature Linking in Multiple Target Tracking Systems
Andreas Andreou, Thomas Edwards
Recent physiological research has shown that synchronization of oscillatory responses in striate cortex may code for relationships between visual features of objects. A VLSI circuit has been de(cid:173)signed to provide rapid phase-locking synchronization of multiple oscillators to allow for further exploration of this neural mechanism. By exploiting the intrinsic random transistor mismatch of devices operated in subthreshold, large groups of phase-locked oscillators can be readily partitioned into smaller phase-locked groups. A mUltiple target tracker for binary images is described utilizing this phase-locking architecture. A VLSI chip has been fabricated and tested to verify the architecture. The chip employs Pulse Ampli(cid:173)tude Modulation (PAM) to encode the output at the periphery of the system.
****

A Network Mechanism for the Determination of Shape-From-Texture
Kô Sakai, Leif Finkel
We propose a computational model for how the cortex discriminates shape and dep

th from texture. The model consists of four stages: (1) extraction of local spa tial frequency, (2) frequency characterization, (3) detection of texture compre ssion by normalization, and (4) integration of the normalized frequency over sp ace. The model accounts for a number of psychophysical observations including e xperiments based on novel random textures. These textures are generated from wh ite noise and manipulated in Fourier domain in order to produce specific frequ ency spectra. Simulations with a range of stimuli, including real images, show qualitative and quantitative agreement with human perception.
****

Encoding Labeled Graphs by Labeling RAAM
Alessandro Sperduti
In this paper we propose an extension to the RAAM by Pollack. This extension, the Labeling RAAM (LRAAM), can encode la(cid:173) beled graph s with cycles by representing pointers explicitly. Data encoded in an L RAAM can be accessed by pointer as well as by content. Direct access by content can be achieved by transform(cid:173) ing the encoder netw ork of the LRAAM into an analog Hopfield network with hidden units. Different access procedures can be defined depending on the access key . Sufficient conditions on the asymptotical stability of the associated Hop field network are briefly introduced.
****

Learning Mackey-Glass from 25 examples, Plus or Minus 2
Mark Plutowski, Garrison Cottrell, Halbert White
We apply active exemplar selection (Plutowski &. White, 1991; 1993) to predicti ng a chaotic time series. Given a fixed set of ex(cid:173) amples, the method ch ooses a concise subset for training. Fitting these exemplars results in the ent ire set being fit as well as de(cid:173) sired. The algorithm incorporates a met hod for regulating network complexity, automatically adding exempla.rs and hidd en units as needed. Fitting examples generated from the Mackey-Glass equa(cid:1 73) tion with fractal dimension 2.1 to an rmse of 0.01 required about 25 exampl ars and 3 to 6 hidden units. The method requires an order of magnitude fewer fl oating point operations than training on the entire set of examples, is signifi cantly cheaper than two contend(cid:173) ing exemplar selection techniques, and suggests a simpler active selection technique that performs comparably.
****

Structured Machine Learning for 'Soft' Classification with Smoothing Spline ANOV A and Stacked Tuning, Testing and Evaluation
Grace Wahba, Yuedong Wang, Chong Gu, Ronald Klein, MD, Barbara Klein, MD
We describe the use of smoothing spline analysis of variance (SS(cid:173) ANOVA) in the penalized log likelihood context, for learning (estimating) the probabi lity p of a '1' outcome, given a train(cid:173) ing set with attribute vectors a nd outcomes. p is of the form pet) = eJ(t) /(1 + eJ(t)), where, if t is a vecto r of attributes, f is learned as a sum of smooth functions of one attribute plu s a sum of smooth functions of two attributes, etc. The smoothing parameters g overning f are obtained by an iterative unbiased risk or iterative GCV method. Confidence intervals for these estimates are available.
****