SIMPLIFYING NEURAL NETS BY DISCOVERING FLAT MINIMA
Sepp Hochreiter, Jürgen Schmidhuber
We present a new algorithm for finding low complexity networks with high genera lization capability. The algorithm searches for large connected regions of so-c alled ''fiat'' minima of the error func(cid:173) tion. In the weight-space envir onment of a "flat" minimum, the error remains approximately constant. Using an MDL-based ar(cid:173) gument, flat minima can be shown to correspond to low expe cted overfitting. Although our algorithm requires the computation of second or der derivatives, it has backprop's order of complexity. Experiments with feedfo rward and recurrent nets are described. In an application to stock market predi ction, the method outperforms conventional backprop, weight decay, and "optimal brain surgeon" .
****

A Model of the Neural Basis of the Rat's Sense of Direction
William Skaggs, James Knierim, Hemant Kudrimoti, Bruce McNaughton
In the last decade the outlines of the neural structures subserving the se nse of direction have begun to emerge. Several investigations have shed light on the effects of vestibular input and visual input on the head direct ion representation. In this paper, a model is formulated of the neural m echanisms underlying the head direction system. The model is built out of sim ple ingredients, depending on nothing more complicated than connectional s pecificity, attractor dynamics, Hebbian learning, and sigmoidal nonlinear ities, but it behaves in a sophisticated way and is consistent with most of the observed properties ofreal head direction cells. In addition it ma kes a number of predictions that ought to be testable by reasonably s traightforward experiments.
****

A Mixture Model System for Medical and Machine Diagnosis
Magnus Stensmo, Terrence J. Sejnowski
Diagnosis of human disease or machine fault is a missing data problem since man y variables are initially unknown. Additional information needs to be obtained. The j oint probability distribution of the data can be used to solve this prob lem. We model this with mixture models whose parameters are estimated by the EM algorithm. This gives the benefit that missing data in the database itself ca n also be handled correctly. The request for new information to refine the dia gnosis is performed using the maximum utility principle. Since the system is based on learning it is domain independent and less labor intensive th an expert systems or probabilistic networks. An example using a heart disease database is presented.
****

Learning with Preknowledge: Clustering with Point and Graph Matching Distance Me asures
Steven Gold, Anand Rangarajan, Eric Mjolsness
Prior constraints are imposed upon a learning problem in the form of distance m easures. Prototypical 2-D point sets and graphs are learned by clustering with point matching and graph matching dis(cid:173) tance measures. The point matchin g distance measure is approx. invariant under affine transformations - translat ion, rotation, scale and shear - and permutations. It operates between noisy im ages with missing and spurious points. The graph matching distance measure ope rates on weighted graphs and is invariant under per(cid:173) mutations. Learning is formulated as an optimization problem . Large objectives so formulated ('" million variables) are efficiently minimized using a combination of optimizatio n techniques - alge(cid:173) braic transformations, iterative projective scaling , clocked objec(cid:173) tives, and deterministic annealing.
****

Learning Local Error Bars for Nonlinear Regression
David Nix, Andreas Weigend
We present a new method for obtaining local error bars for nonlinear reg ression, i.e., estimates of the confidence in predicted values that de(cid:173) pend on the input. We approach this problem by applying a maximum(cid:173) li

kelihood framework to an assumed distribution of errors. We demon(cid:173) strate our method first on computer-generated data with locally varying, normally distributed target noise. We then apply it to laser data from the Santa Fe Time Series Competition where the underlying system noise is known quantization error and the error bars give local estimates of model misspecification. In both cases, the method also provides a weighted(cid:173) regression effect that improves generalization performance.

****

Reinforcement Learning Methods for Continuous-Time Markov Decision Problems

Steven Bradtke, Michael Duff

Semi-Markov Decision Problems are continuous time generaliza(cid:173) tions of discrete time Markov Decision Problems. A number of reinforcement learning algorithms have been developed recently for the solution of Markov Decision Problems, based on the ideas of asynchronous dynamic programming and stochastic approxima(cid:173) tion. Among these are TD(,x), Q-Iearning, and Real-time Dynamic Programming. After reviewing semi-Markov Decision Problems and Bellman's optimality equation in that context, we propose al(cid:173) gorithms similar to those named above, adapted to the solution of semi-Markov Decision Problems. We demonstrate these algorithms by applying them to the problem of determining the optimal con(cid:173) trol for a simple queueing system. We conclude with a discussion of circumstances under which these algorithms may be usefully ap(cid:173) plied.

****

Connectionist Speaker Normalization with Generalized Resource Allocating Networks

Cesare Furlanello, Diego Giuliani, Edmondo Trentin

The paper presents a rapid speaker-normalization technique based on neural network spectral mapping. The neural network is used as a front-end of a continuous speech recognition system (speaker(cid:173) dependent, HMM-based) to normalize the input acoustic data from a new speaker. The spectral difference between speakers can be reduced using a limited amount of new acoustic data (40 phonet(cid:173) ically rich sentences). Recognition error of phone units from the acoustic-phonetic continuous speech corpus APASCI is decreased with an adaptability ratio of 25%. We used local basis networks of elliptical Gaussian kernels, with recursive allocation of units and on-line optimization of parameters (GRAN model). For this ap(cid:173) plication, the model included a linear term. The results compare favorably with multivariate linear mapping based on constrained orthonormal transformations.

****

A Novel Reinforcement Model of Birdsong Vocalization Learning

Kenji Doya, Terrence J. Sejnowski

Songbirds learn to imitate a tutor song through auditory and motor learn(cid:173) ing. We have developed a theoretical framework for song learning that accounts for response properties of neurons that have been observed in many of the nuclei that are involved in song learning. Specifically, we suggest that the anteriorforebrain pathway, which is not needed for song production in the adult but is essential for song acquisition, provides synaptic perturbations and adaptive evaluations for syllable vocalization learning. A computer model based on reinforcement learning was con(cid:173) structed that could replicate a real zebra finch song with 90% accuracy based on a spectrographic measure. The second generation of the bird(cid:173) song model replicated the tutor song with 96% accuracy.

****

Bias, Variance and the Combination of Least Squares Estimators

Ronny Meir

We consider the effect of combining several least squares estimators on the expected performance of a regression problem. Computing the exact bias and variance curves as a function of the sample size we are able to quantitatively compare the effect of the combination on the bias and variance separately, an

d thus on the expected error which is the sum of the two. Our exact calcu lations, demonstrate that the combination of estimators is particularly useful in the case where the data set is small and noisy and the function to be lear ned is unrealizable. For large data sets the single estimator produces superior results. Finally, we show that by splitting the data set into several independent parts and training each estimator on a different subset, the performance can in some cases be significantly improved.
****

Hierarchical Mixtures of Experts Methodology Applied to Continuous Speech Recogn ition

Ying Zhao, Richard Schwartz, Jason Sroka, John Makhoul

In this paper, we incorporate the Hierarchical Mixtures of Experts (HME) method of probability estimation, developed by Jordan [1], into an H MM(cid:173) based continuous speech recognition system. The resulting sys tem can be thought of as a continuous-density HMM system, but instead of usin g gaussian mixtures, the HME system employs a large set of hierarchically orga nized but relatively small neural networks to perform the probability dens ity estimation. The hierarchical structure is reminiscent of a decision tree except for two important differences: each "expert" or neural net performs a "soft" decision rather than a hard decision, and, unlike ordinary decision trees, the parameters of all the neural nets in the HME are a utomatically trainable using the EM algorithm. We report results on the ARPA 5,OOO-word and 4O,OOO-word Wall Street Journal corpus using HME model s.
****

A Comparison of Discrete-Time Operator Models for Nonlinear System Identificatio n

Andrew Back, Ah Tsoi

We present a unifying view of discrete-time operator models used in the context of finite word length linear signal processing. Comparisons are made between the recently presented gamma operator model, and the delta and rho operator mod els for performing nonlinear system identification and prediction using neural networks. A new model based on an adaptive bilinear transformation which ge neralizes all of the above models is presented.
****

Learning Many Related Tasks at the Same Time with Backpropagation

Rich Caruana

Hinton [6] proposed that generalization in artificial neural nets shoul d improve if nets learn to represent the domain's underlying regularities. Abu-Mustafa's hints work [1] shows that the outputs of a backprop net can be used as inputs through which domain(cid:173) specific information can be given to the net. We extend these ideas by showing that a backprop ne t learning many related tasks at the same time can use these tasks as inducti ve bias for each other and thus learn better. We identify five mechanisms by which multitask backprop improves generalization and give empirical evidenc e that multi task backprop generalizes better in real domains.
****

Multidimensional Scaling and Data Clustering

Thomas Hofmann, Joachim Buhmann

Visualizing and structuring pairwise dissimilarity data are difficult combinator ial op(cid:173) timization problems known as multidimensional scaling or pairwis e data clustering. Algorithms for embedding dissimilarity data set in a Euclidi an space, for clustering these data and for actively selecting data to support the clustering process are discussed in the maximum entropy framework. Active d ata selection provides a strategy to discover structure in a data set efficient ly with partially unknown data.
****

Predicting the Risk of Complications in Coronary Artery Bypass Operations using Neural Networks

Richard P. Lippmann, Linda Kukolich, David Shahian

Experiments demonstrated that sigmoid multilayer perceptron (MLP) networks provide slightly better risk prediction than conventional logistic regression when used to predict the risk of death, stroke, and renal failure on 1257 patients who underwent coronary artery bypass operations at the Lahey Clinic. MLP networks with no hidden layer and networks with one hidden layer were trained using stochastic gradient descent with early stopping. MLP networks and logistic regression used the same input features and were evaluated using bootstrap sampling with 50 replications. ROC areas for predicting mortality using preoperative input features were 70.5% for logistic regression and 76.0% for MLP networks. Regularization provided by early stopping was an important component of improved perfonnance. A simplified approach to generating confidence intervals for MLP risk predictions using an auxiliary "confidence MLP" was developed. The confidence MLP is trained to reproduce confidence intervals that were generated during training using the outputs of 50 MLP networks trained with different bootstrap samples.

****

Spatial Representations in the Parietal Cortex May Use Basis Functions

Alexandre Pouget, Terrence J. Sejnowski

The parietal cortex is thought to represent the egocentric posi(cid:173)tions of objects in particular coordinate systems. We propose an alternative approach to spatial perception of objects in the pari(cid:173)etal cortex from the perspective of sensorimotor transformations. The responses of single parietal neurons can be modeled as a gaus(cid:173)sian function of retinal position multiplied by a sigmoid function of eye position, which form a set of basis functions. We show here how these basis functions can be used to generate receptive fields in either retinotopic or head-centered coordinates by simple linear transformations. This raises the possibility that the parietal cortex does not attempt to compute the positions of objects in a partic(cid:173)ular frame of reference but instead computes a general purpose representation of the retinal location and eye position from which any transformation can be synthesized by direct projection. This representation predicts that hemineglect, a neurological syndrome produced by parietal lesions, should not be confined to egocentric coordinates, but should be observed in multiple frames of reference in single patients, a prediction supported by several experiments.

****

Reinforcement Learning Algorithm for Partially Observable Markov Decision Problems

Tommi Jaakkola, Satinder Singh, Michael Jordan

Increasing attention has been paid to reinforcement learning algo(cid:173)rithms in recent years, partly due to successes in the theoretical analysis of their behavior in Markov environments. If the Markov assumption is removed, however, neither generally the algorithms nor the analyses continue to be usable. We propose and analyze a new learning algorithm to solve a certain class of non-Markov decision problems. Our algorithm applies to problems in which the environment is Markov, but the learner has restricted access to state information. The algorithm involves a Monte-Carlo pol(cid:173)icy evaluation combined with a policy improvement method that is similar to that of Markov decision problems and is guaranteed to converge to a local maximum. The algorithm operates in the space of stochastic policies, a space which can yield a policy that per(cid:173)forms considerably better than any deterministic policy. Although the space of stochastic policies is continuous-even for a discrete action space-our algorithm is computationally tractable.

****

FINANCIAL APPLICATIONS OF LEARNING FROM HINTS

Yaser Abu-Mostafa

The basic paradigm for learning in neural networks is 'learning from examples' where a training set of input-output examples is used to teach the network the

target function. Learning from hints is a gen(cid:173) eralization of learning f rom examples where additional information about the target function can be inco rporated in the same learning process. Such information can come from common se nse rules or special expertise. In financial market applications where the trai n(cid:173) ing data is very noisy, the use of such hints can have a decisive ad vantage. We demonstrate the use of hints in foreign-exchange trading of the U.S . Dollar versus the British Pound, the German Mark, the Japanese Yen, and the S wiss Franc, over a period of 32 months. We explain the general method of learni ng from hints and how it can be applied to other markets. The learning model fo r this method is not restricted to neural networks.

****

An Auditory Localization and Coordinate Transform Chip

Timothy Horiuchi

The localization and orientation to various novel or interesting events in the environment is a critical sensorimotor ability in all animals, predator or prey. In mammals, the superior colliculus (SC) plays a major role in this behavior, the deeper layers ex(cid:173) hibiting topo graphically mapped responses to visual, auditory, and somatosensory stimuli. Sensory information arriving from differ(cid:173) ent modalities should then be represented in the same coordinate frame. Auditory cues, in particula r, are thought to be computed in head-based coordinates which must then be t ransformed to reti(cid:173) nal coordinates. In this paper, an analog VLSI impl ementation for auditory localization in the azimuthal plane is described which ex(cid:173) tends the architecture proposed for the barn owl to a primate eye movement system where further transformation is required. This tran sformation is intended to model the projection in primates from auditory corti cal areas to the deeper layers of the primate superior colliculus. This syste m is interfaced with an analog VLSI-based saccadic eye movement system al so being constructed in our labo(cid:173) ratory.

****

Limits on Learning Machine Accuracy Imposed by Data Quality

Corinna Cortes, L. D. Jackel, Wan-Ping Chiang

Random errors and insufficiencies in databases limit the perfor(cid:173) mance of any classifier trained from and applied to the database. In thi s paper we propose a method to estimate the limiting perfor(cid:173) mance of c lassifiers imposed by the database. We demonstrate this technique on the ta sk of predicting failure in telecommunication paths.

****

Interference in Learning Internal Models of Inverse Dynamics in Humans

Reza Shadmehr, Tom Brashers-Krug, Ferdinando Mussa-Ivaldi

Experiments were performed to reveal some of the computational properties of the human motor memory system. We show that as humans practice re aching movements while interacting with a novel mechanical environment, th ey learn an internal model of the inverse dynamics of that environment. Subje cts show recall of this model at testing sessions 24 hours after the ini tial practice. The representation of the internal model in memory is such that there is interference when there is an attempt to learn a new i nverse dynamics map immediately after an anticorrelated mapping was lear ned. We suggest that this interference is an indication that the same computational elements used to encode the first inverse dynamics map are being used to learn the second mapping. We predict that this leads to a forgetting of the initially learned skill.

****

Optimal Movement Primitives

Terence Sanger

The theory of Optimal Unsupervised Motor Learning shows how a network can discover a reduced-order controller for an unknown nonlinear system b y representing only the most significant modes. Here, I extend the theory to apply to command sequences, so that the most significant components discov ered by the network corre(cid:173) spond to motion "primitives". Combina

tions of these primitives can be used to produce a wide variety of di
fferent movements. I demonstrate applications to human handwriting decom
position and synthesis, as well as to the analysis of electrophysiolog
ical experiments on movements resulting from stimulation of the frog spin
al cord.
****

Using a Saliency Map for Active Spatial Selective Attention: Implementation & In
itial Results

Shumeet Baluja, Dean A. Pomerleau

In many vision based tasks, the ability to focus attention on the important por
tions of a scene is crucial for good performance on the tasks. In this paper we
present a simple method of achieving spatial selective attention through the u
se of a saliency map. The saliency map indicates which regions of the input ret
ina are important for performing the task. The saliency map is cre(cid:173) ated
through predictive auto-encoding. The performance of this method is demonstrat
ed on two simple tasks which have multiple very strong distract(cid:173) ing fea
tures in the input retina. Architectural extensions and application directions
for this model are presented.
****

Factorial Learning and the EM Algorithm

Zoubin Ghahramani

Many real world learning problems are best characterized by an interaction of m
ultiple independent causes or factors. Discover(cid:173) ing such causal structu
re from the data is the focus of this paper. Based on Zemel and Hinton's cooper
ative vector quantizer (CVQ) architecture, an unsupervised learning algorithm i
s derived from the Expectation-Maximization (EM) framework. Due to the com(cid:
173) binatorial nature of the data generation process, the exact E-step is comp
utationally intractable. Two alternative methods for com(cid:173) puting the E-s
tep are proposed: Gibbs sampling and mean-field approximation, and some promisi
ng empirical results are presented.
****

Pairwise Neural Network Classifiers with Probabilistic Outputs

David Price, Stefan Knerr, Léon Personnaz, Gérard Dreyfus

Multi-class classification problems can be efficiently solved by partiti
oning the original problem into sub-problems involving only two classes:
for each pair of classes, a (potentially small) neural network is trained u
sing only the data of these two classes. We show how to combine the
outputs of the two-class neural networks in order to obtain posterior proba
bilities for the class decisions. The resulting probabilistic pairwise classi
fier is part of a handwriting recognition system which is currently applied t
o check reading. We present results on real world data bases and show that, fro
m a practical point of view, these results compare favorably to other neural ne
twork approaches.
****

Real-Time Control of a Tokamak Plasma Using Neural Networks

Chris M. Bishop, Paul S. Haynes, Mike E U Smith, Tom N. Todd, David L. Trotman,
Colin G. Windsor

This paper presents results from the first use of neural networks for
the real-time feedback control of high temperature plasmas in a tokamak fu
sion experiment. The tokamak is currently the prin(cid:173) cipal experime
ntal device for research into the magnetic confine(cid:173) ment approac
h to controlled fusion. In the tokamak, hydrogen plasmas, at temperatu
res of up to 100 Million K, are confined by strong magnetic fields.
Accurate control of the position and shape of the plasma boundary requir
es real-time feedback control of the magnetic field structure on a time-sc
ale of a few tens of mi(cid:173) croseconds. Software simulations have demons
trated that a neural network approach can give significantly better perf
ormance than the linear technique currently used on most tokamak experiment
s. The practical application of the neural network approach requires high-
speed hardware, for which a fully parallel implementation of the multil

ayer perceptron, using a hybrid of digital and analogue technology, has b
een developed.
****

An Actor/Critic Algorithm that is Equivalent to Q-Learning
Robert Crites, Andrew Barto
We prove the convergence of an actor/critic algorithm that is equiv(cid:173) ale
nt to Q-Iearning by construction. Its equivalence is achieved by encoding Q-
values within the policy and value function of the ac(cid:173) tor and crit
ic. The resultant actor/critic algorithm is novel in two ways: it updates th
e critic only when the most probable action is executed from any given
state, and it rewards the actor using cri(cid:173) teria that depend on the re
lative probability of the action that was executed.
****

Template-Based Algorithms for Connectionist Rule Extraction
Jay Alexander, Michael C. Mozer
Casting neural network weights in symbolic terms is crucial for interp
reting and explaining the behavior of a network. Additionally, in some do
mains, a symbolic description may lead to more robust generalization.
We present a principled approach to symbolic rule extraction based on
the notion of weight templates, parameterized regions of weight space cor
responding to specific symbolic expressions. With an appropriate choice of
representation, we show how template parameters may be efficiently iden
tified and instantiated to yield the optimal match to a unit's actual weig
hts. Depending on the requirements of the application domain, our method
can accommodate arbitrary disjunctions and conjunctions with O(k) comple
xity, simple n-of-m expressions with O(k!) complexity, or a more general cla
ss of recursive n-of-m expressions with O(k!) complexity, where k is th
e number of inputs to a unit. Our method of rule extraction offers se
veral benefits over alternative approaches in the literature, and simulation r
esults on a variety of problems demonstrate its effectiveness.
****

Reinforcement Learning with Soft State Aggregation
Satinder Singh, Tommi Jaakkola, Michael Jordan
It is widely accepted that the use of more compact representations than loo
kup tables is crucial to scaling reinforcement learning (RL) algorithms to real
-world problems. Unfortunately almost all of the theory of reinforcement lea
rning assumes lookup table representa(cid:173) tions. In this paper we add
ress the pressing issue of combining function approximation and RL, and
present 1) a function approx(cid:173) imator based on a simple extension
to state aggregation (a com(cid:173) monly used form of compact represe
ntation), namely soft state aggregation, 2) a theory of convergence for
RL with arbitrary, but fixed, soft state aggregation, 3) a novel intui
tive understanding of the effect of state aggregation on online RL, and 4) a
new heuristic adaptive state aggregation algorithm that finds improved compa
ct representations by exploiting the non-discrete nature of soft state a
ggregation. Preliminary empirical results are also presented.
****

A Connectionist Technique for Accelerated Textual Input: Letting a Network Do th
e Typing
Dean Pomerleau
Each year people spend a huge amount of time typing. The text people type typic
ally contains a tremendous amount of redundancy due to predictable word usage
patterns and the text's structure. This paper describes a neural netwo
rk system call AutoTypist that monitors a person's typing and predicts what wil
l be entered next. AutoTypist displays the most likely subsequent word to the
typist, who can accept it with a single keystroke, instead of typing it in its
entirety. The multi-layer perceptron at the heart of Auto'JYpist adapts its p
redictions of likely subsequent text to the user's word usage pattern, and to
the characteristics of the text currently being typed. Increases in typing spe
ed of 2-3% when typing English prose and 10-20% when typing C code have been de

monstrated using the system, suggesting a potential time savings of more than 2 0 hours per user per year. In addition to increasing typing speed, AutoTypist r educes the number of keystrokes a user must type by a similar amount (2-3% for English, 10- 20% for computer programs). This keystroke savings has the potent ial to significantly reduce the frequency and severity of repeated stress inju ries caused by typing, which are the most common injury suffered in today's of fice environment.
****

Advantage Updating Applied to a Differential Game
Mance E. Harmon, Leemon Baird, A. Harry Klopf
An application of reinforcement learning to a linear-quadratic, differential ga me is presented. The reinforcement learning system uses a recently devel oped algorithm, the residual gradient form of advantage updating. The game is a Markov Decision Process (MDP) with continuous time, states, and act ions, linear dynamics, and a quadratic cost function. The game consists of two players, a missile and a plane; the missile pursues the plane and the plane evades the missile. The reinforcement learning algorithm for optimal control is modified for differential games in order to find the minimax point, rather t han the maximum. Simulation results are compared to the optimal solution, demonstrating that the simulated reinforcement learning system converges to the optimal answer. The performance of both the residual gradient and n on-residual gradient forms of advantage updating and Q-learning are compared. The results show that advantage updating converges faster than Q-learning i n all simulations. The results also show advantage updating converges regardl ess of the time step duration; Q-learning is unable to converge as the time s tep duration ~rows small.
****

Pattern Playback in the 90s
Malcolm Slaney
Deciding the appropriate representation to use for modeling human audit ory processing is a critical issue in auditory science. While engi(cid:173) neers have successfully performed many single-speaker tasks with LPC and spe ctrogram methods, more difficult problems will need a richer representa tion. This paper describes a powerful auditory representation known as the corr elogram and shows how this non-linear representation can be converted back into sound, with no loss of perceptually impor(cid:173) tant information. The c orrelogram is interesting because it is a neuro(cid:173) physiologically p lausible representation of sound. This paper shows improved methods for spectrogram inversion (conventional pattern playback), inversion of a coch lear model, and inversion of the correlo(cid:173) gram representation.
****

An Analog Neural Network Inspired by Fractal Block Coding
Fernando Pineda, Andreas Andreou
We consider the problem of decoding block coded data, using a physical dynamica l system. We sketch out a decompression algorithm for fractal block codes and then show how to implement a recurrent neural network using physical ly simple but highly-nonlinear, analog circuit models of neurons and synaps es. The nonlinear system has many fixed points, but we have at our disposal a p rocedure to choose the parameters in such a way that only one solution, the d esired solution, is stable. As a partial proof of the concept, we present experimental data from a small system a 16-neuron analog CMOS chip fabricate d in a 2m analog p-well process. This chip operates in the subthreshold regime and, for each choice of parameters, converges to a unique stable state. Each s tate exhibits a qualitatively fractal shape.
****

Phase-Space Learning
Fu-Sheng Tsung, Garrison Cottrell
Existing recurrent net learning algorithms are inadequate. We in(cid:173) tro duce the conceptual framework of viewing recurrent training as matching vector fields of dynamical systems in phase space. Phase(cid:173) space reconstructi

on techniques make the hidden states explicit, reducing temporal learni ng to a feed-forward problem. In short, we propose viewing iterated p rediction [LF88] as the best way of training recurrent networks on de terministic signals. Using this framework, we can train multiple traject ories, insure their stabil(cid:173) ity, and design arbitrary dynamical sy stems.

****

An experimental comparison of recurrent neural networks

Bill Horne, C. Giles

Many different discrete-time recurrent neural network architec(cid:173) tu res have been proposed. However, there has been virtually no effort t o compare these arch:tectures experimentally. In this paper we review and c ategorize many of these architectures and compare how they perform on various classes of simple problems including grammatical inference and nonlinear sys tem identification.

****

Inferring Ground Truth from Subjective Labelling of Venus Images

Padhraic Smyth, Usama Fayyad, Michael Burl, Pietro Perona, Pierre Baldi

In remote sensing applications "ground-truth" data is often used as the basis for training pattern recognition algorithms to gener(cid:173) ate thematic maps or to detect objects of interest. In practical situation s, experts may visually examine the images and provide a subjective noisy esti mate of the truth. Calibrating the reliability and bias of expert labellers is a non-trivial problem. In this paper we discuss some of our recent work on this topic in the context of detecting small volcanoes in M agellan SAR images of Venus. Empirical results (using the Expectation-Maxim ization procedure) suggest that accounting for subjective noise can be quite signifi(cid:173) cant in terms of quantifying both human and algorithm de tection performance.

****

Learning Prototype Models for Tangent Distance

Trevor Hastie, Patrice Simard

Simard, LeCun & Denker (1993) showed that the performance of nearest-neigh bor classification schemes for handwritten character recognition can be improved by incorporating invariance to spe(cid:173) the so cific trans formations in the underlying distance metric - called tangent distance. The resulting classifier, however, can be prohibitively slow and memory in tensive due to the large amount of prototypes that need to be stored and used i n the distance compar(cid:173) isons. In this paper we develop rich models for representing large subsets of the prototypes. These models are either us ed singly per class, or as basic building blocks in conjunction with the K-means clustering algorithm.

****

Diffusion of Credit in Markovian Models

Yoshua Bengio, Paolo Frasconi

This paper studies the problem of diffusion in Markovian models, such as hidden Markov models (HMMs) and how it makes very difficult the task of learning of long-term dependencies in sequences. Using results from Markov c hain theory, we show that the problem of diffusion is reduced if the transitio n probabilities approach 0 or 1. Under this condition, standard HMMs have very limited modeling capabilities, but input/output HMMs can still perform inter esting computations.

****

The Ni1000: High Speed Parallel VLSI for Implementing Multilayer Perceptrons

Michael Perrone, Leon Cooper

In this paper we present a new version of the standard multilayer perce ptron (MLP) algorithm for the state-of-the-art in neural net(cid:173) work VLS I implementations: the Intel Ni1000. This new version of the MLP uses a fu ndamental property of high dimensional spaces which allows the 12-norm to be accurately approximated by the It -norm. This approach enables the

standard MLP to utilize the parallel architecture of the Ni1000 to achieve on the order of 40000, 256-dimensional classifications per second.

****

Model of a Biological Neuron as a Temporal Neural Network
Sean D. Murphy, Edward W. Kairiss

A biological neuron can be viewed as a device that maps a multidimen(cid:173) sional temporal event signal (dendritic postsynaptic activations) into a unidimensional temporal event signal (action potentials). We have designed a network, the Spatio-Temporal Event Mapping (STEM) architecture, which can learn to perform this mapping for arbitrary bio(cid:173) physical models of neurons. Such a network appropriately trained, called a STEM cell, can be used in place of a conventional compartmen(cid:173) tal model in simulations where only the transfer function is important, such as network simulations. The STEM cell offers advantages over compartmental models in terms of computational efficiency, analytical tractabili1ty, and as a framework for VLSI implementations of biologi(cid:173) cal neurons.

****

Non-linear Prediction of Acoustic Vectors Using Hierarchical Mixtures of Experts
Steve Waterhouse, Anthony Robinson

In this paper we consider speech coding as a problem of speech modelling. In particular, prediction of parameterised speech over short time segments is performed using the Hierarchical Mixture of Experts (HME) (Jordan & Jacobs 1994). The HME gives two ad(cid:173) vantages over traditional non-linear function approximators such as the Multi-Layer Percept ron (MLP); a statistical understand(cid:173) ing of the operation of the predictor and provision of information about the performance of the predictor in the form of likelihood information and local error bars. These two issues are examined on both toy and real world problems of regression and time series prediction. In the speech coding context, we extend the principle of combining local predictions via the HME to a Vector Quantiza(cid:173) tion scheme in which fixed local codebooks are combined on-line for each observation.

****

JPMAX: Learning to Recognize Moving Objects as a Model-fitting Problem
Suzanna Becker

Unsupervised learning procedures have been successful at low-level feature extraction and preprocessing of raw sensor data. So far, however, they have had limited success in learning higher-order representations, e.g., of objects in visual images. A promising ap(cid:173) proach is to maximize some measure of agreement between the outputs of two groups of units which receive inputs physically sep(cid:173) arated in space, time or modality, as in (Becker and Hinton, 1992; Becker, 1993; de Sa, 1993). Using the same approach, a much sim(cid:173) pler learning procedure is proposed here which discovers features in a single-layer network consisting of several populations of units, and can be applied to multi-layer networks trained one layer at a time. When trained with this algorithm on image sequences of moving geometric objects a two-layer network can learn to perform accurate position-invariant object classification.

****

The Electrotonic Transformation: a Tool for Relating Neuronal Form to Function
Nicholas T. Carnevale, Kenneth Y. Tsai, Brenda Claiborne, Thomas Brown

The spatial distribution and time course of electrical signals in neurons have important theoretical and practical consequences. Because it is difficult to infer how neuronal form affects electrical signaling, we have developed a quantitative yet intuitive approach to the analysis of electrotonus. This approach transforms the architecture of the cell from anatomical to electrotonic space, using the logarithm of voltage attenuation as the distance metric. We describe the theory behind this approach and illustrate its use.

****

Dynamic Modelling of Chaotic Time Series with Neural Networks
Jose C. Principe, Jyh-Ming Kuo

The auditory system of the barn owl contains several spatial maps. In young barn owls raised with optical prisms over their eyes, these auditory maps are shifted to stay in register with the visual map, suggesting that the visual input imposes a frame of reference on the auditory maps. However, the optic tectum, the first site of convergence of visual with auditory information, is not the site of plasticity for the shift of the auditory maps; the plasticity occurs instead in the inferior colliculus, which contains an auditory map and projects into the optic tectum. We explored a model of the owl remapping in which a global reinforcement signal whose delivery is controlled by visual foveation. A hebb learning rule gated by rein(cid:173) forcement learned to appropriately adjust auditory maps. In addi(cid:173) tion, reinforcement learning preferentially adjusted the weights in the inferior colliculus, as in the owl brain, even though the weights were allowed to change throughout the auditory system. This ob(cid:173) servation raises the possibility that the site of learning does not have to be genetically specified, but could be determined by how the learning procedure interacts with the network architecture.

****

Boltzmann Chains and Hidden Markov Models

Lawrence Saul, Michael Jordan

We propose a statistical mechanical framework for the modeling of discrete time series. Maximum likelihood estimation is done via Boltzmann learning in one-dimensional networks with tied weights. We call these networks Boltzmann chains and show that they contain hidden Markov models (HMMs) as a special case. Our framework also motivates new architectures that address partic(cid:173) ular shortcomings of HMMs. We look at two such architectures: parallel chains that model feature sets with disparate time scales, and looped networks that model long-term dependencies between hidden states. For these networks, we show how to implement the Boltzmann learning rule exactly, in polynomial time, without resort to simulated or mean-field annealing. The necessary com(cid:173) putations are done by exact decimation procedures from statistical mechanics.

****

Learning with Product Units

Laurens Leerink, C. Giles, Bill Horne, Marwan Jabri

The TNM staging system has been used since the early 1960's to predict breast cancer patient outcome. In an attempt to in(cid:173) crease prognostic accuracy, many putative prognostic factors have been identified. Because the TNM stage model can not accom(cid:173) modate these new factors, the proliferation of factors in breast cancer has lead to clinical confusion. What is required is a new computerized prognostic system that can test putative prognostic factors and integrate the predictive factors with the TNM vari(cid:173) ables in order to increase prognostic accuracy. Using the area un(cid:173) der the curve of the receiver operating characteristic, we compare the accuracy of the following predictive models in terms of five year breast cancer-specific survival: pTNM staging system, princi(cid:173) pal component analysis, classification and regression trees, logistic regression, cascade correlation neural network, conjugate gradient descent neural, probabilistic neural network, and backpropagation neural network. Several statistical models are significantly more ac-

****

Efficient Methods for Dealing with Missing Data in Supervised Learning

Volker Tresp, Ralph Neuneier, Subutai Ahmad

We present efficient algorithms for dealing with the problem of mis(cid:173) sing inputs (incomplete feature vectors) during training and recall. Our approach is based on the approximation of the input data dis(cid:173) tribution using Parzen windows. For recall, we obtain closed form solutions for arbitrary feedforward networks. For training, we show how the backpropagation step for an incomplete pattern can be approximated by a weighted averaged backpropagation step. The complexity of the solutions for training and recall is independent of the number of missing features. We verify our theoretical

results using one classification and one regression problem.
****
Predictive Coding with Neural Nets: Application to Text Compression
Jürgen Schmidhuber, Stefan Heil
To compress text files, a neural predictor network P is used to ap(cid:173) proximate the conditional probability distribution of possible "next characters", given n previous characters. P's outputs are fed into standard coding algorithms that generate short codes for characters with high predicted probability and long codes for highly unpre(cid:173) dictable characters. Tested on short German newspaper articles, our method outperforms widely used Lempel-Ziv algorithms (used in UNIX functions such as "compress" and "gzip").
****
Computational Structure of coordinate transformations: A generalization study
Zoubin Ghahramani, Daniel M. Wolpert, Michael Jordan
One of the fundamental properties that both neural networks and the central nervous system share is the ability to learn and gener(cid:173) alize from examples. While this property has been studied exten(cid:173) sively in the neural network literature it has not been thoroughly explored in human perceptual and motor learning. We have chosen a coordinate transformation system-the visuomotor map which transforms visual coordinates into motor coordinates-to study the generalization effects of learning new input-output pairs. Using a paradigm of computer controlled altered visual feedback, we have studied the generalization of the visuomotor map subsequent to both local and context-dependent remappings. A local remapping of one or two input-output pairs induced a significant global, yet decaying, change in the visuomotor map, suggesting a representa(cid:173) tion for the map composed of units with large functional receptive fields. Our study of context-dependent remappings indicated that a single point in visual space can be mapped to two different fin(cid:173) ger locations depending on a context variable-the starting point of the movement. Furthermore, as the context is varied there is a gradual shift between the two remappings, consistent with two visuomotor modules being learned and gated smoothly with the context.
****
Recognizing Handwritten Digits Using Mixtures of Linear Models
Geoffrey E. Hinton, Michael Revow, Peter Dayan
We construct a mixture of locally linear generative models of a col(cid:173) lection of pixel-based images of digits, and use them for recogni(cid:173) tion. Different models of a given digit are used to capture different styles of writing, and new images are classified by evaluating their log-likelihoods under each model. We use an EM-based algorithm in which the M-step is computationally straightforward principal components analysis (PCA). Incorporating tangent-plane informa(cid:173) tion [12] about expected local deformations only requires adding tangent vectors into the sample covariance matrices for the PCA, and it demonstrably improves performance.
****
A Critical Comparison of Models for Orientation and Ocular Dominance Columns in the Striate Cortex
E. Erwin, K. Obermayer, K. Schulten
More than ten of the most prominent models for the structure and for the activity dependent formation of orientation and ocu(cid:173) lar dominance columns in the striate cort(>x have been evaluated. We implemented those models on parallel machines, we extensively explored parameter space, and we quantitatively compared model predictions with experimental data which were recorded optically from macaque striate cortex. In our contribution we present a summary of our results to date. Briefly, we find that (i) despite apparent differences, many models are based on similar principles and, consequently, make similar pre(cid:173) dictions, (ii) certain "pattern models" as well as the developmental "correlation-based learning" models disagree with the experimen(cid:173) tal data, and (ii

i) of the models we have investigated, "competitive Hebbian" models and the rec ent model of Swindale provide the best match with experimental data.
****

Classifying with Gaussian Mixtures and Clusters
Nanda Kambhatla, Todd Leen
In this paper, we derive classifiers which are winner-take-all (WTA) approximat ions to a Bayes classifier with Gaussian mixtures for class conditional densities. The derived classifiers include clustering based algorithms like L VQ and k-Means. We propose a constrained rank Gaussian mixtures model and derive a WTA algorithm for it. Our experiments with two speech classification tasks indicate that the constrained rank model and the WTA approximations impro ve the performance over the unconstrained models.
****

Anatomical origin and computational role of diversity in the response properties of cortical neurons
Kalanit Spector, Shimon Edelman, Rafael Malach
The maximization of diversity of neuronal response properties has been r ecently suggested as an organizing principle for the formation of such prominent features of the functional architecture of the brain as the corti(cid :173) cal columns and the associated patchy projection patterns (Malach, 1994). We show that (1) maximal diversity is attained when the ratio of dendritic an d axonal arbor sizes is equal to one, as found in many cortical are as and across species (Lund et al., 1993; Malach, 1994), and (2) that maxi(cid:173) mization of diversity leads to better performance in systems of receptive fields implementing steerable/shiftable filters, and in matc hing spatially distributed signals, a problem that arises in many high-le vel visual tasks.
****

Synchrony and Desynchrony in Neural Oscillator Networks
Deliang Wang, David Terman
An novel class of locally excitatory, globally inhibitory oscillator networks i s proposed. The model of each oscillator corresponds to a standard relaxation o scillator with two time scales. The network exhibits a mechanism of selective g ating, whereby an oscillator jumping up to its active phase rapidly recruits th e oscillators stimulated by the same pattern, while preventing others from jump ing up. We show analytically that with the selective gating mechanism the netwo rk rapidly achieves both synchronization within blocks of oscillators that are stimulated by connected regions and desynchronization between different blocks . Computer simulations demonstrate the network's promising ability for segmenti ng multiple input patterns in real time. This model lays a physical foundation for the oscillatory correlation theory of feature binding, and may provide an e ffective computational framework for scene segmentation and figure/ground segre gation.
****

A Computational Model of Prefrontal Cortex Function
Todd Braver, Jonathan D. Cohen, David Servan-Schreiber
Accumulating data from neurophysiology and neuropsychology have suggested two information processing roles for prefrontal cor(cid:173) tex (PFC): 1) short-term active memory; and 2) inhibition. We present a new behav ioral task and a computational model which were developed in parallel. The task was developed to probe both of these prefrontal functions sim ultaneously, and produces a rich set of behavioral data that act as c onstraints on the model. The model is implemented in continuous-time, thus providing a natural framework in which to study the temporal dynamics of processing in the task. We show how the model can be used to examine the be(c id:173) havioral consequences of neuromodulation in PFC . Specifically, we use the model to make novel and testable predictions regarding the behavioral per formance of schizophrenics, who are hypothesized to suffer from reduced d opaminergic tone in this brain area.
****

Combining Estimators Using Non-Constant Weighting Functions
Volker Tresp, Michiaki Taniguchi
This paper discusses the linearly weighted combination of estima(cid:173) tors in which the weighting functions are dependent on the input . We show that the weighting functions can be derived either by evaluating the input dependent variance of each estimator or by estimating how likely it is that a given estimator has seen data in the region of the input space close to the input pattern. The lat(cid:173) ter solution is closely related to the mixture of experts approach and we show how learning rules for the mixture of experts can be derived from the theory about learning with missing features. The presented approaches are modular since the weighting functions can easily be modified (no retraining) if more estimators are ad(cid:173) ded. Furthermore, it is easy to incorporate estimators which were not derived from data such as expert systems or algorithms.
****

Stochastic Dynamics of Three-State Neural Networks
Toru Ohira, Jack Cowan
We present here an analysis of the stochastic neurodynamics of a neural network composed of three-state neurons described by a master equation. An outer-product representation of the mas(cid:173) ter equation is employed. In this representation, an extension of the analysis from two to three-state neurons is easily performed. We apply this formalism with approximation schemes to a sim(cid:173) ple three-state network and compare the results with Monte Carlo simulations.
****

On the Computational Utility of Consciousness
Donald Mathis, Michael C. Mozer
We propose a computational framework for understanding and modeling human consciousness. This framework integrates many existing theoretical perspectives, yet is sufficiently concrete to allow simulation experiments. We do not attempt to explain qualia (sub(cid:173) jective experience), but instead ask what differences exist within the cognitive information processing system when a person is con(cid:173) scious of mentally-represented information versus when that infor(cid:173) mation is unconscious. The central idea we explore is that the con(cid:173) tents of consciousness correspond to temporally persistent states in a network of computational modules. Three simulations are de(cid:173) scribed illustrating that the behavior of persistent states in the models corresponds roughly to the behavior of conscious states people experience when performing similar tasks. Our simulations show that periodic settling to persistent (i.e., conscious) states im(cid:173) proves performance by cleaning up inaccuracies and noise, forcing decisions, and helping keep the system on track toward a solution.
****

Ocular Dominance and Patterned Lateral Connections in a Self-Organizing Model of the Primary Visual Cortex
Joseph Sirosh, Risto Miikkulainen
A neural network model for the self-organization of ocular dominance and lateral connections from binocular input is presented. The self-organizing process results in a network where (1) afferent weights of each neuron or(cid:173) ganize into smooth hill-shaped receptive fields primarily on one of the reti(cid:173) nas, (2) neurons with common eye preference form connected, intertwined patches, and (3) lateral connections primarily link regions of the same eye preference. Similar self-organization of cortical structures has been ob(cid:173) served experimentally in strabismic kittens. The model shows how pat(cid:173) terned lateral connections in the cortex may develop based on correlated activity and explains why lateral connection patterns follow receptive field properties such as ocular dominance.
****

Effects of Noise on Convergence and Generalization in Recurrent Networks
Kam Jim, Bill Horne, C. Giles

We introduce and study methods of inserting synaptic noise into dynamically-dri ven recurrent neural networks and show that ap(cid:173) plying a controlled amou nt of noise during training may improve convergence and generalization. In addi tion, we analyze the effects of each noise parameter (additive vs. multiplicati ve, cumulative vs. non-cumulative, per time step vs. per string) and predict th at best overall performance can be achieved by injecting additive noise at eac h time step. Extensive simulations on learning the dual parity grammar from tem poral strings substantiate these predictions.

****

An Integrated Architecture of Adaptive Neural Network Control for Dynamic System s

Ke Liu, Robert Tokar, Brain McVey

In this study, an integrated neural network control architecture for n onlinear dynamic systems is presented. Most of the recent emphasis in the neural network control field has no error feedback as the control input, whic h rises the lack of adaptation problem. The integrated architecture in this paper combines feed forward control and error feedback adaptive c ontrol using neural networks. The paper reveals the different internal functionality of these two kinds of neural network controllers for certai n input styles, e.g., state feedback and error feedback. With error fee dback, neural network controllers learn the slopes or the gains with respect to the error feedback, producing an error driven adaptive cont rol systems. The results demonstrate that the two kinds of control sch eme can be combined to realize their individual advantages. Testing wit h disturbances added to the plant shows good tracking and adaptation with the integrated neural control architecture.

****

Implementation of Neural Hardware with the Neural VLSI of URAN in Applications w ith Reduced Representations

Il Han, Ki-Chul Kim, Hwang-Soo Lee

implement Korean

****

Estimating Conditional Probability Densities for Periodic Variables

Chris M. Bishop, Claire Legleye

Most of the common techniques for estimating conditional prob(cid:173) ability d ensities are inappropriate for applications involving peri(cid:173) odic variabl es. In this paper we introduce three novel techniques for tackling such problem s, and investigate their performance us(cid:173) ing synthetic data. We then app ly these techniques to the problem of extracting the distribution of wind vecto r directions from radar scatterometer data gathered by a remote-sensing satelli te.

****

Analysis of Unstandardized Contributions in Cross Connected Networks

Thomas Shultz, Yuriko Oshima-Takane, Yoshio Takane

Understanding knowledge representations in neural nets has been a diffi cult problem. Principal components analysis (PCA) of contributions (produ cts of sending activations and connection weights) has yielded valuable insight s into knowledge representations, but much of this work has focused on the corr elation matrix of contributions. The present work shows that analyzing the variance-covariance matrix of contributions yields more valid insights by tak ing account of weights.

****

A Rigorous Analysis of Linsker-type Hebbian Learning

J. Feng, H. Pan, V. P. Roychowdhury

We propose a novel rigorous approach for the analysis of Linsker's unsupervised Hebbian learning network. The behavior of this model is determined by the unde rlying nonlinear dynamics which are parameterized by a set of parameters origin ating from the Heb(cid:173) bian rule and the arbor density of the synapses. The se parameters determine the presence or absence of a specific receptive field (also referred to as a 'connection pattern') as a saturated fixed point attract

or of the model. In this paper, we perform a qualitative analysis of the underl ying nonlinear dynamics over the parameter space, determine the effects of the system parameters on the emer(cid:173) gence of various receptive fields, and pr edict precisely within which parameter regime the network will have the potenti al to develop a specially designated connection pattern. In particular, this ap (cid:173) proach exposes, for the first time, the crucial role played by the sy naptic density functions, and provides a complete precise picture of the parame ter space that defines the relationships among the different receptive fields. Our theoretical predictions are confirmed by numerical simulations.

****

Associative Decorrelation Dynamics: A Theory of Self-Organization and Optimizati on in Feedback Networks

Dawei Dong

This paper outlines a dynamic theory of development and adap(cid:173) t ation in neural networks with feedback connections. Given in(cid:173) pu t ensemble, the connections change in strength according to an associative learning rule and approach a stable state where the neuronal outputs are decorrelated . We apply this theory to pri(cid:173) mary visual c ortex and examine the implications of the dynamical decorrelation of the a ctivities of orientation selective cells by the intracortical connections. The theory gives a unified and quantita(cid:173) tive explanation of the ps ychophysical experiments on orientation contrast and orientation adaptation. Using only one parameter, we achieve good agreements between the theoretic al predictions and the experimental data.

****

Visual Speech Recognition with Stochastic Networks

Javier Movellan

This paper presents ongoing work on a speaker independent visual speech recognition system. The work presented here builds on previous research effo rts in this area and explores the potential use of simple hidden Mark ov models for limited vocabulary, speaker independent visual speech reco gnition. The task at hand is recognition of the first four English di gits, a task with possible applications in car-phone images were model ed as mixtures of independent dialing. The Gaussian distributions, and the temporal dependencies were captured with standard left-to-right hidden Markov models. The results indicate that simple hidden Markov models may be used to successfully recognize relatively unprocessed image sequence s. The system achieved performance levels equivalent to untrained humans when asked to recognize the fIrst four English digits.

****

Finding Structure in Reinforcement Learning

Sebastian Thrun, Anton Schwartz

Reinforcement learning addresses the problem of learning to select actions in o rder to maximize one's performance in unknown environments. To scale reinforce ment learning to complex real-world tasks, such as typically studied in AI, one must ultimately be able to discover the structure in the world, in order to ab stract away the myriad of details and to operate in more tractable problem spac es. This paper presents the SKILLS algorithm. SKILLS discovers skills, which a re partially defined action policies that arise in the context of multiple, rel ated tasks. Skills collapse whole action sequences into single operators. The y are learned by minimizing the com(cid:173) pactness of action policies, using a description length argument on their representation. Empirical results in s imple grid navigation tasks illustrate the successful discovery of structur e in reinforcement learning.

****

Active Learning with Statistical Models

David Cohn, Zoubin Ghahramani, Michael Jordan

For many types of learners one can compute the statistically "op(cid:173) timal" way to select data. We review how these techniques have been used with feedfor ward neural networks [MacKay, 1992; Cohn, 1994] . We then show how the same pri

nciples may be used to select data for two alternative, statistically-based lea rning architectures: mixtures of Gaussians and locally weighted regression. Whi le the techniques for neural networks are expensive and approximate, the techn iques for mixtures of Gaussians and locally weighted regres(cid:173) sion are bo th efficient and accurate.

****

From Data Distributions to Regularization in Invariant Learning

Todd Leen

Ideally pattern recognition machines provide constant output when the inputs ar e transformed under a group 9 of desired invariances. These invariances can be achieved by enhancing the training data to include examples of inputs transfo rmed by elements of g, while leaving the corresponding targets unchanged. Alternatively the cost function for training can include a regularizat ion term that penalizes changes in the output when the input is transformed un(cid:173) der the group.

****

An Input Output HMM Architecture

Yoshua Bengio, Paolo Frasconi

We introduce a recurrent architecture having a modular structure and we formulate a training procedure based on the EM algorithm. The resulting model has similarities to hidden Markov models, but supports recurrent networks p rocessing style and allows to exploit the supervised learning paradigm while using maximum likelihood estimation.

****

Grouping Components of Three-Dimensional Moving Objects in Area MST of Visual Co rtex

Richard Zemel, Terrence J. Sejnowski

Many cells in the dorsal part of the medial superior temporal (MST) area of visual cortex respond selectively to spiral flow patterns-speci fic combinations of expansion/ contraction and ro(cid:173) tation motions. Previous investigators have suggested that these cells may represent sel f-motion. Spiral patterns can also be gener(cid:173) ated by the relative m otion of the observer and a particular object. An MST cell may then accou nt for some portion of the complex flow field, and the set of active cells could encode the entire flow; in this manner, MST effectively segments moving objects. Such a grouping operation is essential in inte rpreting scenes containing several independent moving objects and observe r motion. We de(cid:173) scribe a model based on the hypothesis that the selective tuning of MST cells reflects the grouping of object compone nts undergo(cid:173) ing coherent motion. Inputs to the model were gene rated from sequences of ray-traced images that simulated realistic motion s it(cid:173) uations, combining observer motion, eye movements, and indepen(ci d:173) dent object motion. The input representation was modeled after r esponse properties of neurons in area MT, which provides the pri(cid:173) mary input to area MST. After applying an unsupervised learning algorithm, the units became tuned to patterns signaling coherent motion. The results match many of the known properties of MST cells and are consistent with recent studies indicating that these cells process 3-D object motion information.

****

Higher Order Statistical Decorrelation without Information Loss

Gustavo Deco, Wilfried Brauer

A neural network learning paradigm based on information theory is pro(cid:173) p osed as a way to perform in an unsupervised fashion, redundancy reduction among the elements of the output layer without loss of infor(cid:173) mation from the sensory input. The model developed performs nonlin(cid:173) ear decorrelation u p to higher orders of the cumulant tensors and results in probabilistic ally in dependent components of the output layer. This means that we don't need to assu me Gaussian distribution neither at the input nor at the output. The theory pre sented is related to the unsuper(cid:173) vised-learning theory of Barlow, which

proposes redundancy reduction as the goal of cognition. When nonlinear units are used nonlinear princi(cid:173) pal component analysis is obtained. In this case nonlinear manifolds can be reduced to minimum dimension manifolds. If such units are used the network performs a generalized principal component analysis in the sense that non-Gaussian distributions can be linearly decorrelated and higher orders of the correlation tensors are also taken into account. The basic structure of the architecture involves a general transfOlmation that is volume conserving and therefore the entropy, yielding a map without loss of infoIIDation. Minimization of the mutual infoIIDation among the output neurons eliminates the redundancy between the outputs and results in statistical decorrelation of the extracted features. This is known as factorialleaming.

****

Sample Size Requirements for Feedforward Neural Networks

Michael Turmon, Terrence L. Fine

We estimate the number of training samples required to ensure that the performance of a neural network on its training data matches that obtained when fresh data is applied to the network. Existing estimates are higher by orders of magnitude than practice indicates. This work seeks to narrow the gap between theory and practice by transforming the problem into determining the distribution of the supremum of a random field in the space of weight vectors, which in turn is attacked by application of a recent technique called the Poisson clumping heuristic.

****

Generalisation in Feedforward Networks

Adam Kowalczyk, Herman Ferrá

We discuss a model of consistent learning with an additional re(cid:173) striction on the probability distribution of training samples, the target concept and hypothesis class. We show that the model pro(cid:173) vides a significant improvement on the upper bounds of sample complexity, i.e. the minimal number of random training samples allowing a selection of the hypothesis with a predefined accuracy and confidence. Further, we show that the model has the poten(cid:173) tial for providing a finite sample complexity even in the case of infinite VC-dimension as well as for a sample complexity below VC-dimension. This is achieved by linking sample complexity to an "average" number of implement able dichotomies of a training sample rather than the maximal size of a shattered sample, i.e. VC-dimension.

****

The Use of Dynamic Writing Information in a Connectionist On-Line Cursive Handwriting Recognition System

Stefan Manke, Michael Finke, Alex Waibel

In this paper we present NPen ++, a connectionist system for writer independent, large vocabulary on-line cursive handwriting recognition. This system combines a robust input representation, which preserves the dynamic writing information, with a neural network architecture, a so called Multi-State Time Delay Neural Network (MS-TDNN), which integrates rec.ognition and segmen(cid:173) tation in a single framework. Our preprocessing transforms the original coordinate sequence into a (still temporal) sequence offea(cid:173) ture vectors, which combine strictly local features, like curvature or writing direction, with a bitmap-like representation of the co(cid:173) ordinate's proximity. The MS-TDNN architecture is well suited for handling temporal sequences as provided by this input rep(cid:173) resentation. Our system is tested both on writer dependent and writer independent tasks with vocabulary sizes ranging from 400 up to 20,000 words. For example, on a 20,000 word vocabulary we achieve word recognition rates up to 88.9% (writer dependent) and 84.1 % (writer independent) without using any language models.

****

Direct Multi-Step Time Series Prediction Using TD(λ)

Peter T. Kazlas, Andreas Weigend

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.

****

Capacity and Information Efficiency of a Brain-like Associative Net

Bruce Graham, David Willshaw

We have determined the capacity and information efficiency of an associative ne t configured in a brain-like way with partial connec(cid:173) tivity and noisy i nput cues. Recall theory was used to calculate the capacity when pattern recall is achieved using a winners-take(cid:173) all strategy. Transforming the dendri tic sum according to input activity and unit usage can greatly increase the cap acity of the associative net under these conditions. For moderately sparse pat(cid:173) terns, maximum information efficiency is achieved with very low connec tivity levels (~ 10%). This corresponds to the level of con(cid:173) nectivity c ommonly seen in the brain and invites speculation that the brain is connected i n the most information efficient way.

****

SARDNET: A Self-Organizing Feature Map for Sequences

Daniel L. James, Risto Miikkulainen

A self-organizing neural network for sequence classification called SARD NET is described and analyzed experimentally. SARDNET extends the Kohonen Fe ature Map architecture with activation re(cid:173) tention and decay in ord er to create unique distributed response patterns for different sequences. SARDNET yields extremely dense yet descriptive representations of sequential input in very few train(cid:173) ing iterations. The network has proven s uccessful on mapping ar(cid:173) bitrary sequences of binary and real numbers , as well as phonemic representations of English words. Potential appli cations include isolated spoken word recognition and cognitive science models of sequence processing.

****

Deterministic Annealing Variant of the EM Algorithm

Naonori Ueda, Ryohei Nakano

We present a deterministic annealing variant of the EM algorithm for maxim um likelihood parameter estimation problems. In our approach, the EM pr ocess is reformulated as the problem of min(cid:173) imizing the thermodyn amic free energy by using the principle of maximum entropy and statisti cal mechanics analogy. Unlike simu(cid:173) lated annealing approaches, this minimization is deterministically performed. Moreover, the derived algor ithm, unlike the conven(cid:173) tional EM algorithm, can obtain better esti mates free of the initial parameter values.

****

A Non-linear Information Maximisation Algorithm that Performs Blind Separation

Anthony Bell, Terrence J. Sejnowski

A new learning algorithm is derived which performs online stochas(cid:173) tic g radient ascent in the mutual information between outputs and inputs of a netw ork. In the absence of a priori knowledge about the 'signal' and 'nois e' components of the input, propagation of information depends on calibr ating network non-linearities to the detailed higher-order moments of the input density functions. By incidentally minimising mutual information be tween outputs, as well as maximising their individual entropies, the n etwork 'fac(cid:173) torises' the input into independent components. As an example application, we have achieved near-perfect separation of ten digi(cid:173) tally mixed speech signals. Our simulations lead us to believe t hat our network performs better at blind separation than the Herault(cid:173) J utten network, reflecting the fact that it is derived rigorously from the mu tual information objective.

****

Pulsestream Synapses with Non-Volatile Analogue Amorphous-Silicon Memories

A. Holmes, Alan Murray, Stephen Churcher, J. Hajto, M. Rose

This paper presents results from the first use of neural networks for

the real-time feedback control of high temperature plasmas in a tokamak fusion experiment. The tokamak is currently the prin(cid:173) cipal experimental device for research into the magnetic confine(cid:173) ment approach to controlled fusion. In the tokamak, hydrogen plasmas, at temperatures of up to 100 Million K, are confined by strong magnetic fields. Accurate control of the position and shape of the plasma boundary requires real-time feedback control of the magnetic field structure on a time-scale of a few tens of mi(cid:173) croseconds. Software simulations have demonstrated that a neural network approach can give significantly better performance than the linear technique currently used on most tokamak experiments. The practical application of the neural network approach requires high-speed hardware, for which a fully parallel implementation of the multilayer perceptron, using a hybrid of digital and analogue technology, has been developed.

****

Dynamic Cell Structures

Jörg Bruske, Gerald Sommer

Dynamic Cell Structures (DCS) represent a family of artificial neural architectures suited both for unsupervised and supervised learning. They belong to the recently [Martinetz94] introduced class of Topology Representing Networks (TRN) which build perlectly topology pre(cid:173) serving feature maps. DCS empI'oy a modified Kohonen learning rule in conjunction with competitive Hebbian learning. The Kohonen type learning rule serves to adjust the synaptic weight vectors while Hebbian learning establishes a dynamic lateral connection structure between the units reflecting the topology of the feature manifold. In case of super(cid:173) vised learning, i.e. function approximation, each neural unit implements a Radial Basis Function, and an additional layer of linear output units adjusts according to a delta-rule. DCS is the first RBF-based approxima(cid:173) tion scheme attempting to concurrently learn and utilize a perfectly to(cid:173) pology preserving map for improved performance. Simulations on a selection of CMU-Benchmarks indicate that the DCS idea applied to the Growing Cell Structure algorithm [Fritzke93] leads to an efficient and elegant algorithm that can beat conventional models on similar tasks.

****

Single Transistor Learning Synapses

Paul Hasler, Chris Diorio, Bradley Minch, Carver Mead

We describe single-transistor silicon synapses that compute, learn, and provide non-volatile memory retention. The single transistor synapses simultaneously perform long term weight storage, com(cid:173) pute the product of the input and the weight value, and update the weight value according to a Hebbian or a backpropagation learning rule. Memory is accomplished via charge storage on polysilicon floating gates, providing long-term retention without refresh. The synapses efficiently use the physics of silicon to perform weight up(cid:173) dates; the weight value is increased using tunneling and the weight value decreases using hot electron injection. The small size and low power operation of single transistor synapses allows the devel(cid:173) opment of dense synaptic arrays. We describe the design, fabri(cid:173) cation, characterization, and modeling of an array of single tran(cid:173) sistor synapses. When the steady state source current is used as the representation of the weight value, both the incrementing and decrementing functions are proportional to a power of the source current. The synaptic array was fabricated in the standard 21'm double - poly, analog process available from MOSIS.

****

Comparing the prediction accuracy of artificial neural networks and other statistical models for breast cancer survival

Harry B. Burke, David B. Rosen, Philip H. Goodman

The TNM staging system has been used since the early 1960's to predict breast cancer patient outcome. In an attempt to in(cid:173) crease prognostic accuracy, many putative prognostic factors have been identified.

Because the TNM stage model can not accom(cid:173) modate these new factors, the proliferation of factors in breast cancer has lead to cli nical confusion. What is required is a new computerized prognostic syst em that can test putative prognostic factors and integrate the predictiv e factors with the TNM vari(cid:173) ables in order to increase prognosti c accuracy. Using the area un(cid:173) der the curve of the receiver operat ing characteristic, we compare the accuracy of the following predictive models in terms of five year breast cancer-specific survival: pTNM stagin g system, princi(cid:173) pal component analysis, classification and regressio n trees, logistic regression, cascade correlation neural network, conjugat e gradient descent neural, probabilistic neural network, and backpropagation neural network. Several statistical models are significantly more ac-

****

Learning direction in global motion: two classes of psychophysically-motivated m odels

V. Sundareswaran, Lucia Vaina

Perceptual learning is defined as fast improvement in performance and ret ention of the learned ability over a period of time. In a set of psy(c id:173) chophysical experiments we demonstrated that perceptual learning oc(cid:173) curs for the discrimination of direction in stochastic motion stimuli. Here we model this learning using two approaches: a clustering model that learns to accommodate the motion noise, and an averaging model that learns to ignore the noise. Simulations of the models show performance sim ilar to the psychophysical results.

****

On-line Learning of Dichotomies

N. Barkai, H. Seung, H. Sompolinsky

The performance of on-line algorithms for learning dichotomies is studied. In on -line learn(cid:173) ing, the number of examples P is equivalent to the learning time, since each example is presented only once. The learning curve, or genera lization error as a function of P, depends on the schedule at which the learnin g rate is lowered. For a target that is a perceptron rule, the learning curve o f the perceptron algorithm can decrease as fast as p- 1 , if the sched(cid:173) ule is optimized. If the target is not realizable by a perceptron, the perceptro n algorithm does not generally converge to the solution with lowest generalizat ion error. For the case of unrealizability due to a simple output noise, we pro pose a new on-line algorithm for a perceptron yielding a learning curve that ca n approach the optimal generalization error as fast as p-l/2. We then generaliz e the perceptron algorithm to any class of thresholded smooth functions learnin g a target from that class. For "well-behaved" input distributions, if this alg orithm converges to the optimal solution, its learning curve can decrease as fas t as p-l.

****

Asymptotics of Gradient-based Neural Network Training Algorithms

Sayandev Mukherjee, Terrence L. Fine

We study the asymptotic properties of the sequence of iterates of weight-vector estimates obtained by training a multilayer feed for(cid:173) ward neural netwo rk with a basic gradient-descent method using a fixed learning constant and no batch-processing. In the one(cid:173) dimensional case, an exact analysis establ ishes the existence of a limiting distribution that is not Gaussian in general. For the gen(cid:173) eral case and small learning constant, a linearization app roximation permits the application of results from the theory of random ma(cid:173) trices to again establish the existence of a limiting distribution. We stu dy the first few moments of this distribution to compare and contrast the resul ts of our analysis with those of techniques of stochastic approximation.

****

Convergence Properties of the K-Means Algorithms

Léon Bottou, Yoshua Bengio

This paper studies the convergence properties of the well known K-Means cluster ing algorithm. The K-Means algorithm can be de(cid:173) scribed either as a grad

ient descent algorithm or by slightly extend(cid:173) ing the mathematics of the EM algorithm to this hard threshold case. We show that the K-Means algorithm a ctually minimizes the quantization error using the very fast Newton algorithm.
****

Using Voice Transformations to Create Additional Training Talkers for Word Spott ing
Eric Chang, Richard P. Lippmann
Speech recognizers provide good performance for most users but the err or rate often increases dramatically for a small percentage of talkers who a re "different" from those talkers used for training. One expensive solution t o this problem is to gather more training data in an attempt to sample these outlier users. A second solution, explored in this paper, is to artificially enlarge the number of training talkers by transforming the speech of existing t raining talkers. This approach is similar to enlarging the training set for OC R digit recognition by warping the training digit images, but is more dif ficult because continuous speech has a much larger number of dimensions (e.g. linguistic, phonetic, style, temporal, spectral) that differ across t alkers. We explored the use of simple linear spectral warping to enlarge a 48-t alker training data base used for word spotting. The average detection rat e overall was increased by 2.9 percentage points (from 68.3% to 71.2%) for male speakers and 2.5 percentage points (from 64.8% to 67.3%) for female speakers. This increase is small but similar to that obtained b y doubling the amount of training data.
****

Forward dynamic models in human motor control: Psychophysical evidence
Daniel M. Wolpert, Zoubin Ghahramani, Michael Jordan
Based on computational principles, with as yet no direct experi(cid:173) mental validation, it has been proposed that the central nervous sys tem (CNS) uses an internal model to simulate the dynamic be(cid:173) havior of the motor system in planning, control and learning (Sut(cid:173) ton and Ba rto, 1981; Ito, 1984; Kawato et aI., 1987; Jordan and Rumelhart, 1992 ; Miall et aI., 1993). We present experimental re(cid:173) sults and simul ations based on a novel approach that investigates the temporal propagation of errors in the sensorimotor integration process. Our results provide dire ct support for the existence of an internal model.
****

Direction Selectivity In Primary Visual Cortex Using Massive Intracortical Conne ctions
Humbert Suarez, Christof Koch, Rodney Douglas
Almost all models of orientation and direction selectivity in visual cortex are based on feedforward connection schemes, where genicu(cid:173) late input provi des all excitation to both pyramidal and inhibitory neurons. The latter neurons then suppress the response of the for(cid:173) mer for non-optimal stimuli. How ever, anatomical studies show that up to 90 % of the excitatory synaptic input onto any corti(cid:173) cal cell is provided by other cortical cells. The massiv e excitatory feedback nature of cortical circuits is embedded in the canonical microcircuit of Douglas &. Martin (1991). We here investigate ana(cid:173) lyti cally and through biologically realistic simulations the function(cid:173) ing o f a detailed model of this circuitry, operating in a hysteretic mode. In the mo del, weak geniculate input is dramatically ampli(cid:173) fied by intracortical excitation, while inhibition has a dual role: (i) to prevent the early genicula te-induced excitation in the null di(cid:173) rection and (ii) to restrain excit ation and ensure that the neurons fire only when the stimulus is in their recep tive-field. Among the
****

Bayesian Query Construction for Neural Network Models
Gerhard Paass, Jörg Kindermann
If data collection is costly, there is much to be gained by actively se(cid:173) lecting particularly informative data points in a sequential way. In a Baye sian decision-theoretic framework we develop a query selec(cid:173) tion cri

terion which explicitly takes into account the intended use of the model p redictions. By Markov Chain Monte Carlo methods the necessary quantities c an be approximated to a desired preci(cid:173) sion. As the number of d ata points grows, the model complexity is modified by a Bayesian mode l selection strategy. The proper(cid:173) ties of two versions of the crit erion ate demonstrated in numerical experiments.
****

A Silicon Axon
Bradley Minch, Paul Hasler, Chris Diorio, Carver Mead
We present a silicon model of an axon which shows promise as a buildi ng block for pulse-based neural computations involving cor(cid:173) relation s of pulses across both space and time. The circuit shares a number of f eatures with its biological counterpart including an excitation threshol d, a brief refractory period after pulse comple(cid:173) tion, pulse ampli tude restoration, and pulse width restoration. We provide a simple explanatio n of circuit operation and present data from a chip fabricated in a standa rd 2Jlm CMOS process through the MOS Implementation Service (MOSIS). We emphasize the ne(cid:173) cessity of the restoration of the width of the pulse in time for stable propagation in axons.
****

Plasticity-Mediated Competitive Learning
Nicol Schraudolph, Terrence J. Sejnowski
Differentiation between the nodes of a competitive learning net(cid:173) work is conventionally achieved through competition on the ba(cid:173) sis of neural ac tivity. Simple inhibitory mechanisms are limited to sparse representations, while decorrelation and factorization schemes that support distributed rep resentations are computation(cid:173) ally unattractive. By letting neural plas ticity mediate the compet(cid:173) itive interaction instead, we obtain diffuse, nonadaptive alterna(cid:173) tives for fully distributed representations. We use this technique to Simplify and improve our binary information gain optimiz a(cid:173) tion algorithm for feature extraction (Schraudolph and Sejnowski, 19 93); the same approach could be used to improve other learning algorithms.
****

Active Learning for Function Approximation
Kah Sung, Partha Niyogi
We develop a principled strategy to sample a function optimally for function a pproximation tasks within a Bayesian framework. Using ideas from optimal ex periment design, we introduce an objective function (incorporating both bi as and variance) to measure the de(cid:173) gree of approximation, and the po tential utility of the data points towards optimizing this objective. We show how the general strat(cid:173) egy can be used to derive precise algorith ms to select data for two cases: learning unit step functions and polyn omial functions. In particular, we investigate whether such active algorithms can learn the target with fewer examples. We obtain theoretical and empir(cid:173) ical results to suggest that this is the case.
****

Patterns of damage in neural networks: The effects of lesion area, shape and num ber
Eytan Ruppin, James Reggia
Current understanding of the effects of damage on neural networks is rudimenta ry, even though such understanding could lead to im(cid:173) portant insights concerning neurological and psychiatric disorders. Motivated by this consid eration, we present a simple analytical framework for estimating the func tional damage resulting from fo(cid:173) cal structural lesions to a neura l network. The effects of focal le(cid:173) sions of varying area, shape a nd number on the retrieval capacities of a spatially-organized associative memo ry. Although our analyti(cid:173) cal results are based on some approximations, they correspond well with simulation results. This study sheds light on some important features characterizing the clinical manifestations of multi-inf arct dementia, including the strong association between the number of infarc

ts and the prevalence of dementia after stroke, and the 'mul(cid:173) tiplicati ve' interaction that has been postulated to occur between Alzheimer's diseas e and multi-infarct dementia.

****

A Study of Parallel Perturbative Gradient Descent

D. Lippe, Joshua Alspector

We have continued our study of a parallel perturbative learning method [Alspector et al., 1993] and implications for its implemen(cid:173) tation in analog VLSI. Our new results indicate that, in most cases, a single parallel perturbation (per pattern presentation) of the func(cid:173) tion parameters (weights in a neural network) is theoretically the best course. This is not true, however, for certain problems and may not generally be tru e when faced with issues of implemen(cid:173) tation such as limited pre cision. In these cases, multiple parallel perturbations may be best as ind icated in our previous results.

****

A Neural Model of Delusions and Hallucinations in Schizophrenia

Eytan Ruppin, James Reggia, David Horn

We implement and study a computational model of Stevens' [19921 theory of th e pathogenesis of schizophrenia. This theory hypoth(cid:173) esizes that the onset of schizophrenia is associated with reactive synaptic regenerat ion occurring in brain regions receiving degener(cid:173) ating temporal lobe p rojections. Concentrating on one such area, the frontal cortex, we mode l a frontal module as an associative memory neural network whose input synapses represent incoming temporal projections. We analyze how, in t he face of weakened external input projections, compensatory strengthening of internal synaptic connections and increased noise levels can maintain mem(cid :173) ory capacities (which are generally preserved in schizophrenia) . However, These compensatory changes adversely lead to sponta(cid:173) neo us, biased retrieval of stored memories, which corresponds to the occurr ence of schizophrenic delusions and hallucinations with(cid:173) out any ap parent external trigger, and for their tendency to con(cid:173) centrate on just few central themes. Our results explain why these symptoms tend t o wane as schizophrenia progresses, and why de(cid:173) layed therapeutical intervention leads to a much slower response.

****

Correlation and Interpolation Networks for Real-time Expression Analysis/Synthes is

Trevor Darrell, Irfan Essa, Alex Pentland

We describe a framework for real-time tracking of facial expressions th at uses neurally-inspired correlation and interpolation methods. A distr ibuted view-based representation is used to characterize facial state, and is computed using a replicated correlation network. The ensemble response of the set of view correlation scores is input to a network based interpolation metho d, which maps perceptual state to motor control states for a simulated 3-D face model. Activation levels of the motor state correspond to muscle acti vations in an anatomically derived model. By integrating fast and robust 2-D processing with 3-D models, we obtain a system that is able to quickly track an d interpret complex facial motions in real-time.

****

Neural Network Ensembles, Cross Validation, and Active Learning

Anders Krogh, Jesper Vedelsby

Learning of continuous valued functions using neural network en(cid:173) sembles (committees) can give improved accuracy, reliable estima(cid:173) tio n of the generalization error, and active learning. The ambiguity is define d as the variation of the output of ensemble members aver(cid:173) aged over u nlabeled data, so it quantifies the disagreement among the networks. It is discussed how to use the ambiguity in combina(cid:173) tion with cross-valid ation to give a reliable estimate of the ensemble generalization error, and how this type of ensemble cross-validation can sometimes improve performance. It

is shown how to estimate the optimal weights of the ensemble members using u nlabeled data. By a generalization of query by committee, it is finally show n how the ambiguity can be used to select new training data to be labeled in a n active learning scheme.
****
Extracting Rules from Artificial Neural Networks with Distributed Representation s
Sebastian Thrun
Although artificial neural networks have been applied in a variety of real-world scenarios with remarkable success, they have often been criticized for exhib iting a low degree of human comprehensibility. Techniques that compile compact sets of symbolic rules out of artificial neural networks offer a promisin g perspective to overcome this obvious deficiency of neural network represe ntations. This paper presents an approach to the extraction of if-then ru les from artificial neu(cid:173) Its key mechanism is validity interval analysis, which is a generic ral networks. tool for extracting symbol ic knowledge by propagating rule-like knowledge through Backpropagation-styl e neural networks. Empirical studies in a robot arm domain illus(cid:173) trat e the appropriateness of the proposed method for extracting rules from networks with real-valued and distributed representations.
****
A model of the hippocampus combining self-organization and associative memory fu nction
Michael Hasselmo, Eric Schnell, Joshua Berke, Edi Barkai
A model of the hippocampus is presented which forms rapid self -orga(cid:173) ni zed representations of input arriving via the perforant path, performs recall o f previous associations in region CA3, and performs comparison of this recall w ith afferent input in region CA 1. This comparison drives feedback regulation of cholinergic modulation to set appropriate dynamics for learning of new repre sentations in region CA3 and CA 1. The network responds to novel patterns with increased cholinergic mod(cid:173) ulation, allowing storage of new self-organiz ed representations, but responds to familiar patterns with a decrease in acetyl choline, allowing recall based on previous representations. This requires sel ectivity of the cholinergic suppression of synaptic transmission in stratum rad iatum of regions CA3 and CAl, which has been demonstrated experimentally.
****
Glove-TalkII: Mapping Hand Gestures to Speech Using Neural Networks
Sidney Fels, Geoffrey E. Hinton
Glove-TaikII is a system which translates hand gestures to speech through an adaptive interface. Hand gestures are mapped contin(cid:173) uously t o 10 control parameters of a parallel formant speech syn(cid:173) thesize r. The mapping allows the hand to act as an artificial vocal tract that produces speech in real time. This gives an unlimited vocabulary in ad dition to direct control of fundamental frequency and volume. Currently, t he best version of Glove-TalkII uses sev(cid:173) eral input devices (in cluding a CyberGlove, a ContactGlove, a 3- space tracker, and a foot-ped al), a parallel formant speech synthe(cid:173) sizer and 3 neural networks . The gesture-to-speech task is divided into vowel and consonant produc tion by using a gating network to weight the outputs of a vowel and a consonant neural network. The gating network and the consonant network are trained with examples from the user. The vowel network implement s a fixed, user-defined relationship between hand-position and vowel so und and does not require any training examples from the user. Volume, fundam ental frequency and stop consonants are produced with a fixed mapping from the input devices. One subject has trained to speak intelligibly wi th Glove-TalkII. He speaks slowly with speech quality similar to a text-to- speech synthesizer but with far more natural-sounding pitch variations.
****
Learning in large linear perceptrons and why the thermodynamic limit is relevant to the real world

Peter Sollich
We present a new method for obtaining the response function 9 and its average G from which most of the properties of learning and generalization in linear perceptrons can be derived. We first rederive the known results for the 'thermodynamic limit' of infinite perceptron size N and show explicitly that 9 is self-averaging in this limit. We then discuss extensions of our method to more gen(cid:173) eral learning scenarios with anisotropic teacher space priors, input distributions, and weight decay terms. Finally, we use our method to calculate the finite N corrections of order 1/ N to G and discuss the corresponding finite size effects on generalization and learning dynamics. An important spin-off is the observation that results obtained in the thermodynamic limit are often directly relevant to systems of fairly modest, 'real-world' sizes.

****

Learning Saccadic Eye Movements Using Multiscale Spatial Filters
Rajesh Rao, Dana Ballard
We describe a framework for learning saccadic eye movements using a photometric representation of target points in natural scenes. The rep(cid:173) resentation takes the form of a high-dimensional vector comprised of the responses of spatial filters at different orientations and scales. We first demonstrate the use of this response vector in the task of locating pre(cid:173) viously foveated points in a scene and subsequently use this property in a multisaccade strategy to derive an adaptive motor map for delivering accurate saccades.

****

A Charge-Based CMOS Parallel Analog Vector Quantizer
Gert Cauwenberghs, Volnei Pedroni
We present an analog VLSI chip for parallel analog vector quantiza(cid:173) tion. The MOSIS 2.0 J..Lm double-poly CMOS Tiny chip contains an array of 16 x 16 charge-based distance estimation cells, implementing a mean absolute difference (MAD) metric operating on a 16-input analog vector field and 16 analog template vectors. The distance cell includ(cid:173) ing dynamic template storage measures 60 x 78 J..Lm2 • Additionally, the chip features a winner-take-all (WTA) output circuit of linear com(cid:173) plexity, with global positive feedback for fast and decisive settling of a single winner output. Experimental results on the complete 16 x 16 VQ system demonstrate correct operation with 34 dB analog input dynamic range and 3 J..Lsec cycle time at 0.7 mW power dissipation.

****

Boosting the Performance of RBF Networks with Dynamic Decay Adjustment
Michael Berthold, Jay Diamond
Radial Basis Function (RBF) Networks, also known as networks of locally-tuned processing units (see [6]) are well known for their ease of use. Most algorithms used to train these types of net(cid:173) works, however, require a fixed architecture, in which the number of units in the hidden layer must be determined before training starts. The RCE training algorithm, introduced by Reilly, Cooper and Elbaum (see [8]), and its probabilistic extension, the P-RCE algorithm, take advantage of a growing structure in which hidden units are only introduced when necessary. The nature of these al(cid:173) gorithms allows training to reach stability much faster than is the case for gradient-descent based methods. Unfortunately P-RCE networks do not adjust the standard deviation of their prototypes individually, using only one global value for this parameter. This paper introduces the Dynamic Decay Adjustment (DDA) al(cid:173) gorithm which utilizes the constructive nature of the P-RCE al(cid:173) gorithm together with independent adaptation of each prototype's decay factor. In addition, this radial adjustment is class dependent and distinguishes between different neighbours. It is shown that networks trained with the presented algorithm perform substan(cid:173) tially better than common RBF networks.

****

An Alternative Model for Mixtures of Experts

Lei Xu, Michael Jordan, Geoffrey E. Hinton
We propose an alternative model for mixtures of experts which uses a different parametric form for the gating network. The modified model is trained by the EM algorithm. In comparison with earlier models-trained by either EM or gradient ascent-there is no need to select a learning stepsize. We repor t simulation experiments which show that the new architecture yields fa ster convergence. We also apply the new model to two problem domains: pie cewise nonlinear function approximation and the combination of multiple previo usly trained classifiers.
****

Catastrophic Interference in Human Motor Learning
Tom Brashers-Krug, Reza Shadmehr, Emanuel Todorov
Biological sensorimotor systems are not static maps that transform input (sens ory information) into output (motor behavior). Evi(cid:173) dence from m any lines of research suggests that their representa(cid:173) tions are pl astic, experience-dependent entities. While this plastic(cid:173) ity is essen tial for flexible behavior, it presents the nervous system with difficult or ganizational challenges. If the sensorimotor system adapts itself to perform w ell under one set of circumstances, will it then perform poorly when placed in an environment with different demands (negative transfer)? Will a later experience-dependent change undo the benefits of previous learning (cata strophic inter(cid:173) ference)? We explore the first question in a sepa rate paper in this volume (Shadmehr et al. 1995). Here we present psych ophysical and computational results that explore the question of catastrophic interference in the context of a dynamic motor learning task. Un(cid:173) der some conditions, subjects show evidence of catastrophic inter(cid:173) fere nce. Under other conditions, however, subjects appear to be immune to its effects. These results suggest that motor learning can undergo a process of consolidation. Modular neural networks are well suited for th e demands of learning multiple input/output mappings. By incorporating the n otion of fast- and slow-changing connections into a modular architecture, w e were able to account for the psychophysical results.
****

On the Computational Complexity of Networks of Spiking Neurons
Wolfgang Maass
We investigate the computational power of a formal model for net(cid:173) works of spiking neurons, both for the assumption of an unlimited timing precision, and for the case of a limited timing precision. We also prove upper and lowe r bounds for the number of examples that are needed to train such networks.
****

New Algorithms for 2D and 3D Point Matching: Pose Estimation and Correspondence
Steven Gold, Chien-Ping Lu, Anand Rangarajan, Suguna Pappu, Eric Mjolsness
A fundamental open problem in computer vision-determining pose and corr espondence between two sets of points in space(cid:173) is solved with a novel, robust and easily implementable algorithm. The technique works on noisy point sets that may be of unequal sizes and may differ by non-ri gid transformations. A 2D varia(cid:173) tion calculates the pose betwee n point sets related by an affine transformation-translation, rotation, s cale and shear. A 3D to 3D variation calculates translation and rotation. An objective describ(cid:173) ing the problem is derived from Mean field theo ry. The objective is minimized with clocked (EM-like) dynamics. Experiment s with both handwritten and synthetic data provide empirical evidence fo r the method.
****

PCA-Pyramids for Image Compression
Horst Bischof, Kurt Hornik
This paper presents a new method for image compression by neural networks. Fi rst, we show that we can use neural networks in a py(cid:173) ramidal frame work, yielding the so-called PCA pyramids. Then we present an image compressio n method based on the PCA pyramid, which is similar to the Laplace pyrami

d and wavelet transform. Some experimental results with real images are re
ported. Finally, we present a method to combine the quantization step
with the learning of the PCA pyramid.
****

Morphogenesis of the Lateral Geniculate Nucleus: How Singularities Affect Global
Structure
Svilen Tzonev, Klaus Schulten, Joseph Malpeli
The macaque lateral geniculate nucleus (LGN) exhibits an intricate lamination p
attern, which changes midway through the nucleus at a point coincident with sma
ll gaps due to the blind spot in the retina. We present a three-dimensional m
odel of morphogenesis in which local cell interactions cause a wave of devel
opment of neuronal re(cid:173) ceptive fields to propagate through the nucleus
and establish two distinct lamination patterns. We examine the interactions
between the wave and the localized singularities due to the gaps, and find
that the gaps induce the change in lamination pattern. We explore critical fa
ctors which determine general LGN organization.
****

Reinforcement Learning Predicts the Site of Plasticity for Auditory Remapping in
the Barn Owl
Alexandre Pouget, Cedric Deffayet, Terrence J. Sejnowski
The auditory system of the barn owl contains several spatial maps. In young ba
rn owls raised with optical prisms over their eyes, these auditory maps are
shifted to stay in register with the visual map, suggesting that the v
isual input imposes a frame of reference on the auditory maps. However
, the optic tectum, the first site of convergence of visual with audi
tory information, is not the site of plasticity for the shift of the aud
itory maps; the plasticity occurs instead in the inferior colliculus, wh
ich contains an auditory map and projects into the optic tectum. We explor
ed a model of the owl remapping in which a global reinforcement signal whose
delivery is controlled by visual foveation. A hebb learning rule gated by
rein(cid:173) forcement learned to appropriately adjust auditory maps. In a
ddi(cid:173) tion, reinforcement learning preferentially adjusted the weig
hts in the inferior colliculus, as in the owl brain, even though the weights
were allowed to change throughout the auditory system. This ob(cid:173)
servation raises the possibility that the site of learning does not ha
ve to be genetically specified, but could be determined by how the lea
rning procedure interacts with the network architecture.
****

A Lagrangian Formulation For Optical Backpropagation Training In Kerr-Type Optic
al Networks
James Steck, Steven Skinner, Alvaro Cruz-Cabrara, Elizabeth Behrman
A training method based on a form of continuous spatially distributed optical e
rror back-propagation is presented for an all optical network composed of nondi
screte neurons and weighted interconnections. The all optical network is feed-f
orward and is composed of thin layers of a Kerr(cid:173) type self focusing/defo
cusing nonlinear optical material. The training method is derived from a Lagran
gian formulation of the constrained minimization of the network error at the ou
tput. This leads to a formulation that describes training as a calculation of t
he distributed error of the optical signal at the output which is then reflecte
d back through the device to assign a spatially distributed error to the intern
al layers. This error is then used to modify the internal weighting values. Res
ults from several computer simulations of the training are presented, and a sim
ple optical table demonstration of the network is discussed.
****

Instance-Based State Identification for Reinforcement Learning
R. Andrew McCallum
This paper presents instance-based state identification, an approach to reinfor
cement learning and hidden state that builds disambiguat(cid:173) ing amounts of
short-term memory on-line, and also learns with an order of magnitude fewer tr
aining steps than several previous ap(cid:173) proaches. Inspired by a key simil

arity between learning with hidden state and learning in continuous geometrical spaces, this approach uses instance-based (or "memory-based") learning, a method that has worked well in continuous spaces.

****

Nonlinear Image Interpolation using Manifold Learning
Christoph Bregler, Stephen Omohundro
The problem of interpolating between specified images in an image sequence is a simple, but important task in model-based vision. We describe an approach based on the abstract task of "manifold learning" and present results on both synthetic and real image se(cid:173) quences. This problem arose in the development of a combined lip-reading and speech recognition system.

****

A Growing Neural Gas Network Learns Topologies
Bernd Fritzke
An incremental network model is introduced which is able to learn the important topological relations in a given set of input vectors by means of a simple Hebb-like learning rule. In contrast to previous approaches like the "neural gas" method of Martinetz and Schulten (1991, 1994), this model has no parameters which change over time and is able to continue learning, adding units and connections, until a performance criterion has been met. Applications of the model include vector quantization, clustering, and interpolation.

****

Transformation Invariant Autoassociation with Application to Handwritten Character Recognition
Holger Schwenk, Maurice Milgram
When training neural networks by the classical backpropagation algo(cid:173) rithm the whole problem to learn must be expressed by a set of inputs and desired outputs. However, we often have high-level knowledge about the learning problem. In optical character recognition (OCR), for in(cid:173) stance, we know that the classification should be invariant under a set of transformations like rotation or translation. We propose a new modular classification system based on several autoassociative multilayer percep(cid:173) trons which allows the efficient incorporation of such knowledge. Results are reported on the NIST database of upper case handwritten letters and compared to other approaches to the invariance problem.

****

Learning to Play the Game of Chess
Sebastian Thrun
This paper presents NeuroChess, a program which learns to play chess from the final outcome of games. NeuroChess learns chess board evaluation functions, represented by artificial neural networks. It integrates inductive neural network learning, temporal differencing, and a variant of explanation-based learning. Performance results illustrate some of the strengths and weaknesses of this approach.

****

Interior Point Implementations of Alternating Minimization Training
Michael Lemmon, Peter Szymanski
This paper presents an alternating minimization (AM) algorithm used in the training of radial basis function and linear regressor networks. The algorithm is a modification of a small-step interior point method used in solving primal linear programs. The algo(cid:173) rithm has a convergence rate of O(fo,L) iterations where n is a measure of the network size and L is a measure of the resulting solution's accuracy. Two results are presented that specify how aggressively the two steps of the AM may be pursued to ensure convergence of each step of the alternating minimization.

****

A Convolutional Neural Network Hand Tracker
Steven Nowlan, John Platt
We describe a system that can track a hand in a sequence of video frames an

d recognize hand gestures in a user-independent manner. The system locates the hand in each video frame and determines if the hand is open or close d. The tracking system is able to track the hand to within ±10 pixels o f its correct location in 99.7% of the frames from a test set containin g video sequences from 18 dif(cid:173) ferent individuals captured in 18 di fferent room environments. The gesture recognition network correctly determ ines if the hand being tracked is open or closed in 99.1 % of the frames in this test set . The system has been designed to operate in real time wit h existing hardware.

****

Temporal Dynamics of Generalization in Neural Networks

Changfeng Wang, Santosh Venkatesh

This paper presents a rigorous characterization of how a general nonlin ear learning machine generalizes during the training process when it is tr ained on a random sample using a gradient descent algorithm based on reduction of training error. It is shown, in particular, that best gener alization performance occurs, in general, before the global minimum of the training error is achieved. The different roles played by the complexit y of the machine class and the complexity of the specific machine in the class during learning are also precisely demarcated.

****

Learning Stochastic Perceptrons Under k-Blocking Distributions

Mario Marchand, Saeed Hadjifaradji

We present a statistical method that PAC learns the class of stochast ic perceptrons with arbitrary monotonic activation func(cid:173) tion and weights Wi E {-I, 0, + I} when the probability distribution that generates the input examples is member of a family that we call k-blocking distributi ons. Such distributions represent an impor(cid:173) tant step beyond the case w here each input variable is statistically independent since the 2k-blocking f amily contains all the Markov distributions of order k. By stochastic per cept ron we mean a per(cid:173) ceptron which, upon presentation of input ve ctor x, outputs 1 with probability fCLJi WiXi - B). Because the same algorit hm works for any monotonic (nondecreasing or nonincreasing) activation f unc(cid:173) tion f on Boolean domain, it handles the well studied cas es of sigmolds and the "usual" radial basis functions.

****

Recurrent Networks: Second Order Properties and Pruning

Morten Pedersen, Lars Hansen

Second order properties of cost functions for recurrent networks are inv estigated. We analyze a layered fully recurrent architecture, the virtue o f this architecture is that it features the conventional feedforward arch itecture as a special case. A detailed description of recursive computation o f the full Hessian of the network cost func(cid:173) tion is provided. We di scuss the possibility of invoking simplifying approximations of the Hessian an d show how weight decays iron the cost function and thereby greatly assist trai ning. We present tenta(cid:173) tive pruning results, using Hassibi et al.' s Optimal Brain Surgeon, demonstrating that recurrent networks can const ruct an efficient internal memory.

****

Unsupervised Classification of 3D Objects from 2D Views

Satoshi Suzuki, Hiroshi Ando

This paper presents an unsupervised learning scheme for categorizing 3D object s from their 2D projected images. The scheme exploits an auto-associati ve network's ability to encode each view of a single object into a representati on that indicates its view direction. We propose two models that employ differ ent classification mechanisms; the first model selects an auto-associative netw ork whose recovered view best matches the input view, and the second model is b ased on a modular architecture whose additional network classifies the view s by splitting the input space nonlinearly. We demonstrate the effectiv eness of the proposed classification models through simulations using 3D wire

-frame objects.
****

Hyperparameters Evidence and Generalisation for an Unrealisable Rule
Glenn Marion, David Saad
Using a statistical mechanical formalism we calculate the evidence, generalisation error and consistency measure for a linear percep(cid:173) tron trained and tested on a set of examples generated by a non linear teacher. The teacher is said to be unrealisable because the student can never model it without error. Our model allows us to interpolate between the known case of a linear teacher, and an un(cid:173) realisable, nonlinear teacher. A comparison of the hyperparameters which maximise the evidence with those that optimise the perfor(cid:173) mance measures reveals that, in the non-linear case, the evidence procedure is a misleading guide to optimising performance. Finally, we explore the extent to which the evidence procedure is unreliable and find that, despite being sub-optimal, in some circumstances it might be a useful method for fixing the hyperparameters.
****

Adaptive Elastic Input Field for Recognition Improvement
Minoru Asogawa
For machines to perform classification tasks, such as speech and character recognition, appropriately handling deformed patterns is a key to achieving high performance. The authors presents a new type of classification system, an Adaptive Input Field Neu(cid:173) ral Network (AIFNN), which includes a simple pre-trained neural network and an elastic input field attached to an input layer. By using an iterative method, AIFNN can determine an optimal affine translation for an elastic input field to compensate for the original deformations. The convergence of the AIFNN algorithm is shown. AIFNN is applied for handwritten numerals recognition. Conse(cid:173) quently, 10.83% of originally misclassified patterns are correctly categorized and total performance is improved, without modifying the neural network.
****

A Model for Chemosensory Reception
Rainer Malaka, Thomas Ragg, Martin Hammer
A new model for chemosensory reception is presented. It models reacti(cid:173) ons between odor molecules and receptor proteins and the activation of second messenger by receptor proteins. The mathematical formulation of the reaction kinetics is transformed into an artificial neural network (ANN). The resulting feed-forward network provides a powerful means for parameter fitting by applying learning algorithms. The weights of the network corresponding to chemical parameters can be trained by presen(cid:173) ting experimental data. We demonstrate the simulation capabilities of the model with experimental data from honey bee chemosensory neurons. It can be shown that our model is sufficient to rebuild the observed data and that simpler models are not able to do this task.
****

A Real Time Clustering CMOS Neural Engine
Teresa Serrano-Gotarredona, Bernabé Linares-Barranco, José Huertas
We describe an analog VLSI implementation of the ARTI algorithm (Carpenter, 1987). A prototype chip has been fabricated in a standard low cost 1.5~m double-metal single-poly CMOS process. It has a die area of lcm2 and is mounted in a 12O-pins PGA package. The chip realizes a modified version of the original ARTI architecture. Such modification has been shown to preserve all computational properties of the original algorithm (Serrano, 1994a), while being more appropriate for VLSI realizations. The chip implements an ARTI network with 100 F 1 nodes and 18 F2 nodes. It can therefore cluster 100 binary pixels input patterns into up to 18 different categories. Modular expansibility of the system is possible by assembling an NxM array of chips without any extra interfacing circuitry, resulting in an F 1 layer with l00xN nodes, and an F2 layer with 18xM nodes. Pattern classification is performed in less than 1.8~s, which means an equivalent computing power of 2.2x109

connections and connection-updates per second. Although internally the chip is analog in nature, it interfaces to the outside world through digital signals, t hus having a true asynchrounous digital behavior. Experimental chip test resu lts are available, which have been obtained through test equipments for digital chips.

****

Learning from queries for maximum information gain in imperfectly learnable prob lems

Peter Sollich, David Saad

In supervised learning, learning from queries rather than from random examples can improve generalization performance signif(cid:173) icantly. W e study the performance of query learning for problems where the student cannot learn the teacher perfectly, which occur frequently in practice . As a prototypical scenario of this kind, we consider a linear percept ron student learning a binary perceptron teacher. Two kinds of queries for maximum information gain, i.e., minimum entropy, are investigated: Minimu m student space en(cid:173) tropy (MSSE) queries, which are appropriate if the teacher space is unknown, and minimum teacher space entropy (MTS E) queries, which can be used if the teacher space is assumed to be known, b ut a student of a simpler form has deliberately been chosen. We find that for MSSE queries, the structure of the student space deter(cid:173) mi nes the efficacy of query learning, whereas MTSE queries lead to a hi gher generalization error than random examples, due to a lack of feedbac k about the progress of the student in the way queries are selected.

****

Factorial Learning by Clustering Features

Joshua Tenenbaum, Emanuel V. Todorov

We introduce a novel algorithm for factorial learning, motivated by segmentatio n problems in computational vision, in which the underlying factors correspond to clusters of highly correlated input features. The algorithm derives from a n ew kind of competitive clustering model, in which the cluster generators compet e to ex(cid:173) plain each feature of the data set and cooperate to explain eac h input example, rather than competing for examples and cooper(cid:173) ating o n features, as in traditional clustering algorithms. A natu(cid:173) ral extensi on of the algorithm recovers hierarchical models of data generated from multipl e unknown categories, each with a differ(cid:173) ent, multiple causal structure . Several simulations demonstrate the power of this approach.

****

Generalization in Reinforcement Learning: Safely Approximating the Value Functio n

Justin Boyan, Andrew Moore

A straightforward approach to the curse of dimensionality in re(cid:173) inforcement learning and dynamic programming is to replace the lookup table with a generalizing function approximator such as a neu(cid:173) ral net. Although this has been successful in the domain of backgam(cid:173) mon, ther e is no guarantee of convergence. In this paper, we show that the comb ination of dynamic programming and function approx(cid:173) imation is not ro bust, and in even very benign cases, may produce an entirely wrong po licy. We then introduce Grow-Support, a new algorithm which is safe from divergence yet can still reap the benefits of successful generalization .

****

A Rapid Graph-based Method for Arbitrary Transformation-Invariant Pattern Classi fication

Alessandro Sperduti, David Stork

We present a graph-based method for rapid, accurate search through prot otypes for transformation-invariant pattern classifica(cid:173) tion. Our meth od has in theory the same recognition accuracy as other recent methods based on ''tangent distance" [Simard et al., 1994], since it uses the same categ orization rule. Nevertheless ours is significantly faster during classific ation because far fewer tan(cid:173) gent distances need be computed. C

rucial to the success of our system are 1) a novel graph architecture in which transformation constraints and geometric relationships among prototypes are en(cid:173) coded during learning, and 2) an improved graph search criterion, used during classification. These architectural insights are applica(cid:173) ble to a wide range of problem domains. Here we demonstrate that on a handwriting recognition task, a basic implementation of our system requires less than half the computation of the Euclidean sorting method.

****

A solvable connectionist model of immediate recall of ordered lists

Neil Burgess

A model of short-term memory for serially ordered lists of verbal stimuli is proposed as an implementation of the 'articulatory loop' thought to mediate this type of memory (Baddeley, 1986). The model predicts the presence of a repeatable time-varying 'context' signal coding the timing of items' presentation in addition to a store of phonological information and a process of serial rehearsal. Items are associated with context nodes and phonemes by Hebbian connections showing both short and long term plasticity. Items are activated by phonemic input during presentation and reactivated by context and phonemic feedback during output. Serial selection of items occurs via a winner-take-all interaction amongst items, with the winner subsequently receiving decaying inhibition. An approximate analysis of error probabilities due to Gaussian noise during output is presented. The model provides an explanatory account of the probability of error as a function of serial position, list length, word length, phonemic similarity, temporal grouping, item and list familiarity, and is proposed as the starting point for a model of rehearsal and vocabulary acquisition.

****

H∞ Optimal Training Algorithms and their Relation to Backpropagation

Babak Hassibi, Thomas Kailath

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.

****

Using a neural net to instantiate a deformable model

Christopher Williams, Michael Revow, Geoffrey E. Hinton

Deformable models are an attractive approach to recognizing non(cid:173) rigid objects which have considerable within class variability. How(cid:173) ever, there are severe search problems associated with fitting the models to data. We show that by using neural networks to provide better starting points, the search time can be significantly reduced. The method is demonstrated on a character recognition task.

****

Grammar Learning by a Self-Organizing Network

Michiro Negishi

This paper presents the design and simulation results of a self(cid:173) organizing neural network which induces a grammar from exam(cid:173) ple sentences. Input sentences are generated from a simple phrase structure grammar including number agreement, verb transitiv(cid:173) ity, and recursive noun phrase construction rules. The network induces a grammar explicitly in the form of symbol categorization rules and phrase structure rules.

****

Coarse-to-Fine Image Search Using Neural Networks

Clay Spence, John Pearson, Jim Bergen

The efficiency of image search can be greatly improved by using a coarse-to-fine search strategy with a multi-resolution image representa(cid:173) tion. However, if the resolution is so low that the objects have few dis(cid:173) tinguishing features, search becomes difficult. We show that the performance of search at such low resolutions can be improved by using context information, i.e., objects visible at low-resolution which are not the objects of

interest but are associated with them. The networks can be given explicit conte xt information as inputs, or they can learn to detect the context objects, in w hich case the user does not have to be aware of their existence. We also use In tegrated Feature Pyramids, which repre(cid:173) sent high-frequency information at low resolutions. The use of multi(cid:173) resolution search techniques a llows us to combine information about the appearance of the objects on many sca les in an efficient way. A natural fOlm of exemplar selection also arises fro m these techniques. We illus(cid:173) trate these ideas by training hierarchic al systems of neural networks to find clusters of buildings in aerial photogra phs of farmland.

****

ICEG Morphology Classification using an Analogue VLSI Neural Network

Richard Coggins, Marwan Jabri, Barry Flower, Stephen Pickard

An analogue VLSI neural network has been designed and tested to perfor m cardiac morphology classification tasks. Analogue tech(cid:173) niques were c hosen to meet the strict power and area requirements of an Implantable Cardiove rter Defibrillator (ICD) system. The ro(cid:173) bustness of the neural networ k architecture reduces the impact of noise, drift and offsets inherent i n analogue approaches. The net(cid:173) work is a 10:6:3 multi-layer percept ron with on chip digital weight storage, a bucket brigade input to feed the Intracardiac Electro(cid:173) gram (ICEG) to the network and has a winner take all circuit at the output. The network was trained in loop and included a commercial ICD in the signal processing path. The syst em has suc(cid:173) cessfully distinguished arrhythmia for different patients wi th better than 90% true positive and true negative detections for dangerous rhythms which cannot be detected by present ICDs. The chip was implemented in 1.2um CMOS and consumes less than 200n W max(cid:173) imum average power in a n area of 2.2 x 2.2mm2.

****