Parameterising Feature Sensitive Cell Formation in Linsker Networks in the Audit
ory System
Lance Walton, David Bisset
This paper examines and extends the work of Linsker (1986) on self organising f
eature detectors. Linsker concentrates on the vi(cid:173) sual processing system
, but infers that the weak assumptions made will allow the model to be used in
the processing of other sensory information. This claim is examined here, with
special attention paid to the auditory system, where there is much lower connec
(cid:173) tivity and therefore more statistical variability. On-line training is
utilised, to obtain an idea of training times. These are then com(cid:173) par
ed to the time available to pre-natal mammals for the formation of feature sens
itive cells.
****
Learning Spatio-Temporal Planning from a Dynamic Programming Teacher: Feed-Forwa
rd Neurocontrol for Moving Obstacle Avoidance
Gerald Fahner, Rolf Eckmiller
Within a simple test-bed, application of feed-forward neurocontrol for short-
term planning of robot trajectories in a dynamic environ(cid:173) ment is stud
ied. The action network is embedded in a sensory(cid:173) motoric system
architecture that contains a separate world model. It is continuously fed
with short-term predicted spatio-temporal obstacle trajectories, and rece
ives robot state feedback. The ac(cid:173) tion net allows for external
switching between alternative plan(cid:173) ning tasks. It generates go
al-directed motor actions - subject to the robot's kinematic and dynami
c constraints - such that colli(cid:173) sions with moving obstacles are
avoided. Using supervised learn(cid:173) ing, we distribute examples of
the optimal planner mapping over a structure-level adapted parsimonious hi
gher order network. The training database is generated by a Dynamic P
rogramming algo(cid:173) rithm. Extensive simulations reveal, that the loca
l planner map(cid:173) ping is highly nonlinear, but can be effectively and
sparsely repre(cid:173) sented by the chosen powerful net model. Excellent
generalization occurs for unseen obstacle configurations. We also discuss th
e limi(cid:173) tations of feed-forward neurocontrol for growing planning horiz
ons.
****
Hidden Markov Models in Molecular Biology: New Algorithms and Applications
Pierre Baldi, Yves Chauvin, Tim Hunkapiller, Marcella McClure
Hidden Markov Models (HMMs) can be applied to several impor(cid:173) tant p
roblems in molecular biology. We introduce a new convergent learning algorithm
for HMMs that, unlike the classical Baum-Welch algorithm is smooth and can be
applied on-line or in batch mode, with or without the usual Viterbi most li
kely path approximation. Left-right HMMs with insertion and deletion states ar
e then trained to represent several protein families including immunoglobulins
and kinases. In all cases, the models derived capture all the important st
atistical properties of the families and can be used efficiently in a
number of important tasks such as multiple alignment, motif de(cid:173) tection,
and classification.
****
Statistical and Dynamical Interpretation of ISIH Data from Periodically Stimulat
ed Sensory Neurons
John K. Douglass, Frank Moss, André Longtin
We interpret the time interval data obtained from periodically stimulated senso
ry neurons in terms of two simple dynamical systems driven by noise with an emb
edded weak periodic function called the signal: 1) a bistable system defined by
two potential wells separated by a barrier, and 2) a Fit(cid:173) zHugh-Nagumo
system. The implementation is by analog simulation: elec(cid:173) tronic circuit
s which mimic the dynamics. For a given signal frequency, our simulators have o
nly two adjustable parameters, the signal and noise intensi(cid:173) ties. We sh
ow that experimental data obtained from the periodically stimu(cid:173) lated me
chanoreceptor in the crayfish tail fan can be accurately approximated by these

simulations. Finally, we discuss stochastic resonance in the two models.
****

Spiral Waves in Integrate-and-Fire Neural Networks
John Milton, Po Chu, Jack Cowan
The formation of propagating spiral waves is studied in a randomly connected neural network composed of integrate-and-fire neurons with recovery period and excitatory connections using computer simulations. Network activity is initiated by periodic stimulation at a single point. The results suggest that spiral waves can arise in such a network via a sub-critical Hopf bifurcation.
****

Adaptive Stimulus Representations: A Computational Theory of Hippocampal-Region Function
Mark Gluck, Catherine E. Myers
We present a theory of cortico-hippocampal interaction in discrimination learning. The hippocampal region is presumed to form new stimulus representations which facilitate learning by enhancing the discriminability of predictive stimuli and compressing stimulus-stimulus redundancies. The cortical and cerebellar regions, which are the sites of long-term memory. may acquire these new representations but are not assumed to be capable of forming new representations themselves. Instantiated as a connectionist model. this theory accounts for a wide range of trial-level classical conditioning phenomena in normal (intact) and hippocampal-Iesioned animals. It also makes several novel predictions which remain to be investigated empirically. The theory implies that the hippocampal region is involved in even the simplest learning tasks; although hippocampal-Iesioned animals may be able to use other strategies to learn these tasks. the theory predicts that they will show consistently different patterns of transfer and generalization when the task demands change.
****

Computing with Almost Optimal Size Neural Networks
Kai-Yeung Siu, Vwani Roychowdhury, Thomas Kailath
Artificial neural networks are comprised of an interconnected collection of certain nonlinear devices; examples of commonly used devices include linear threshold elements, sigmoidal elements and radial-basis elements. We employ results from harmonic analysis and the theory of rational ap(cid:173) proximation to obtain almost tight lower bounds on the size (i.e. number of elements) of neural networks. The class of neural networks to which our techniques can be applied is quite general; it includes any feedforward network in which each element can be piecewise approximated by a low degree rational function. For example, we prove that any depth-(d + 1) network of sigmoidal units or linear threshold elements computing the par(cid:173) ity function of n variables must have O(dnl/d-£) size, for any fixed i > O. In addition, we prove that this lower bound is almost tight by showing that the parity function can be computed with O(dnl/d) sigmoidal units or linear threshold elements in a depth-(d + 1) network. These almost tight bounds are the first known complexity results on the size of neural networks with depth more than two. Our lower bound techniques yield a unified approach to the complexity analysis of various models of neural networks with feedforward structures. Moreover, our results indicate that in the context of computing highly oscillating symmetric Boolean func-
****

Q-Learning with Hidden-Unit Restarting
Charles Anderson
Platt's resource-allocation network (RAN) (Platt, 1991a, 1991b) is modified for a reinforcement-learning paradigm and to "restart" existing hidden units rather than adding new units. After restart(cid:173) ing, units continue to learn via back-propagation. The resulting restart algorithm is tested in a Q-Iearning network that learns to solve an inverted pendulum problem. Solutions are found faster on average with the restart algorithm than without it.
****

Synaptic Weight Noise During MLP Learning Enhances Fault-Tolerance, Generalizati on and Learning Trajectory
Alan Murray, Peter Edwards
We analyse the effects of analog noise on the synaptic arithmetic during MultiLayer Perceptron training, by expanding the cost func(cid:173) tion to include noise-mediated penalty terms. Predictions are made in the light of these calculations which suggest that fault tolerance, generalisation ability and learning trajectory should be improved by such noise-injection. Extensive simulation experiments on two distinct classification problems substantiate the claims. The re(cid:173) sults appear to be perfectly general for all training schemes where weights are adjusted incrementally, and have wide-ranging implica(cid:173) tions for all applications, particularly those involving "inaccurate" analog neural VLSI.
****
Rational Parametrizations of Neural Networks
Uwe Helmke, Robert C. Williamson
A connection is drawn between rational functions, the realization theory of dynamical systems, and feedforward neural networks. This allows us to parametrize single hidden layer scalar neural networks with (almost) arbitrary analytic activation functions in terms of strictly proper rational functions. Hence, we can solve the uniqueness of parametrization problem for such networks.
****
Directional-Unit Boltzmann Machines
Richard Zemel, Christopher Williams, Michael C. Mozer
We present a general formulation for a network of stochastic di(cid:173) rectional units. This formulation is an extension of the Boltzmann machine in which the units are not binary, but take on values in a cyclic range, between 0 and 271' radians. The state of each unit in a Directional-Unit Boltzmann Machine (DUBM) is described by a complex variable, where the phase component specifies a direction; the weights are also complex variables. We associate a quadratic energy function, and corresponding probability, with each DUBM configuration. The conditional distribution of a unit's stochastic state is a circular version of the Gaussian probability distribution, known as the von Mises distribution. In a mean-field approxima(cid:173) tion to a stochastic DUBM, the phase component of a unit's state represents its mean direction, and the magnitude component spec(cid:173) ifies the degree of certainty associated with this direction. This combination of a value and a certainty provides additional repre(cid:173) sentational power in a unit. We describe a learning algorithm and simulations that demonstrate a mean-field DUBM'S ability to learn interesting mappings.
****
Stimulus Encoding by Multidimensional Receptive Fields in Single Cells and Cell Populations in V1 of Awake Monkey
Edward Stern, Ad Aertsen, Eilon Vaadia, Shaul Hochstein
Ad Aertsen
****
On the Use of Projection Pursuit Constraints for Training Neural Networks
Nathan Intrator
\Ve present a novel classifica t.ioll and regression met.hod that com(cid:173) bines exploratory projection pursuit. (unsupervised traiuing) with pro(cid:173) jection pursuit. regression (supervised t.raining), t.o yield a. nev,,' family of cost./complexity penalLy terms . Some improved generalization properties are demonstrat.ed on real \vorld problems.
****
Assessing and Improving Neural Network Predictions by the Bootstrap Algorithm
Gerhard Paass
The bootstrap algorithm is a computational intensive procedure to derive nonparametric confidence intervals of statistical estimators in situations where an analytic solution is intractable. It is ap(cid:173) plied to neura

l networks to estimate the predictive distribution for unseen inputs. The co nsistency of different bootstrap procedures and their convergence speed is discussed. A small scale simulation experiment shows the applicability of the bootstrap to practical problems and its potential use.
****
Self-Organizing Rules for Robust Principal Component Analysis
Lei Xu, Alan L. Yuille
In the presence of outliers, the existing self-organizing rules for Pri ncipal Component Analysis (PCA) perform poorly. Using sta(cid:173) tistic al physics techniques including the Gibbs distribution, binary decision fields and effective energies, we propose self-organizing PCA rules which ar e capable of resisting outliers while fulfilling various PCA-related task s such as obtaining the first principal com(cid:173) ponent vector, the first k principal component vectors, and directly finding the subspace spanned by the first k vector principal com(cid:173) ponent vectors without sol ving for each vector individually. Com(cid:173) parative experiments have shown that the proposed robust rules improve the performances of the existing PCA algorithms signifi(cid:173) cantly when outliers are present.
****
Intersecting regions: The Key to combinatorial structure in hidden unit space
Janet Wiles, Mark Ollila
Hidden units in multi-layer networks form a representation space in which each region can be identified with a class of equivalent outputs (Elman, 1989) or a logical state in a finite state machine (Cleeremans, Servan-Schreiber & McClell and, 1989; Giles, Sun, Chen, Lee, & Chen, 1990). We extend the analysis of the spatial structure of hidden unit space to a combinatorial task, based on bindin g features together in a visual scene. The logical structure requires a combina torial number of states to represent all valid scenes. On analysing our network s, we find that the high dimensionality of hidden unit space is exploited by us ing the intersection of neighboring regions to represent conjunctions of featur es. These results show how combinatorial structure can be based on the spatial nature of networks, and not just on their emulation of logical structure.
****
A Parallel Gradient Descent Method for Learning in Analog VLSI Neural Networks
J. Alspector, R. Meir, B. Yuhas, A. Jayakumar, D. Lippe
Typical methods for gradient descent in neural network learning involve calcula tion of derivatives based on a detailed knowledge of the network model. This re quires extensive, time consuming calculations for each pat(cid:173) tern present ation and high precision that makes it difficult to implement in VLSI. We prese nt here a perturbation technique that measures, not calculates, the gradient. S ince the technique uses the actual network as a measuring device, errors in mod eling neuron activation and synaptic weights do not cause errors in gradient de scent. The method is parallel in nature and easy to implement in VLSI. We descr ibe the theory of such an algorithm, an analysis of its domain of applicability , some simulations using it and an outline of a hardware implementation.
****
A Model of Feedback to the Lateral Geniculate Nucleus
Carlos Brody
Simplified models of the lateral geniculate nucles (LGN) and stri(cid:173) ate cortex illustrate the possibility that feedback to the LG N may be u sed for robust, low-level pattern analysis. The information fed back t o the LG N is rebroadcast to cortex using the LG N 's full fan-out, so th e cortex-LGN-cortex pathway mediates extensive cortico-cortical communication while keeping the number of neces(cid:173) sary connections small.
****
History-Dependent Attractor Neural Networks
Isaac Meilijson, Eytan Ruppin
We present a methodological framework enabling a detailed de(cid:173) sc ription of the performance of Hopfield-like attractor neural net(cid:173) w orks (ANN) in the first two iterations. Using the Bayesian ap(cid:173)

proach, we find that performance is improved when a history-based term is in cluded in the neuron's dynamics. A further enhancement of the network's perf ormance is achieved by judiciously choosing the censored neurons (those which become active in a given itera(cid:173) tion) on the basis of the magnitude of their post-synaptic poten(cid:173) tials. The contribution of biologically plausible, censored, history(cid:173) dependent dynamics is espec ially marked in conditions of low firing activity and sparse connectivity, two important characteristics of the mammalian cortex. In such networks , the performance at(cid:173) tained is higher than the performance of two 'independent' iter(cid:173) ations, which represents an upper bound on the performance of history-independent networks.
****
Single-Iteration Threshold Hamming Networks
Isaac Meilijson, Eytan Ruppin, Moshe Sipper
We analyze in detail the performance of a Hamming network clas(cid:173) si fying inputs that are distorted versions of one of its m stored memory patterns. The activation function of the memory neurons in the original Hamm ing network is replaced by a simple threshold function. The resulting Thres hold Hamming Network (THN) cor(cid:173) rectly classifies the input patte rn, with probability approaching 1, using only O(mln m) connections, in a single iteration. The THN drastically reduces the time and space complexity of Hamming Net(cid:173) work classifiers.
****
Reinforcement Learning Applied to Linear Quadratic Regulation
Steven Bradtke
Recent research on reinforcement learning has focused on algo(cid:173) r ithms based on the principles of Dynamic Programming (DP). One of the most promising areas of application for these algo(cid:173) rithms is the control of dynamical systems, and some impressive results have bee n achieved. However, there are significant gaps between practice and t heory. In particular, there are no con ver(cid:173) gence proofs for pro blems with continuous state and action spaces, or for systems involving no n-linear function approximators (such as multilayer perceptrons). This p aper presents research applying DP-based reinforcement learning theory t o Linear Quadratic Reg(cid:173) ulation (LQR), an important class of cont rol problems involving continuous state and action spaces and requiring a simple type of non-linear function approximator. We describe an algori thm based on Q-Iearning that is proven to converge to the optimal control ler for a large class of LQR problems. We also describe a slightly different algorithm that is only locally convergent to the optimal Q-f unction, demonstrating one of the possible pitfalls of using a non-linea r function approximator with DP-based learning.
****
The Computation of Stereo Disparity for Transparent and for Opaque Surfaces
Suthep Madarasmi, Daniel Kersten, Ting-Chuen Pong
The classical computational model for stereo vision incorporates a uniq ueness inhibition constraint to enforce a one-to-one feature match, there by sacrificing the ability to handle transparency. Crit(cid:173) ics of the m odel disregard the uniqueness constraint and argue that the smoothness co nstraint can provide the excitation support required for transparency compu tation. However, this modifica(cid:173) tion fails in neighborhoods with sparse features. We propose a Bayesian approach to stereo vision with priors favoring cohesive over transparent surfaces. The disparity and its segmentation into a multi-layer "depth planes" representation are simultaneou sly com(cid:173) puted. The smoothness constraint propagates support within ea ch layer, providing mutual excitation for non-neighboring transparent or part ially occluded regions. Test results for various random-dot and other ste reograms are presented.
****
Forecasting Demand for Electric Power

Jen-Lun Yuan, Terrence Fine
We are developing a forecaster for daily extremes of demand for electric power encountered in the service area of a large midwest(cid:173) ern utility and using this application as a testbed for approaches to input dimension reduction and decomposition of network train(cid:173) ing. Projection pursuit regression representations and the ability of algorithms like SIR to quickly find reasonable weighting vectors enable us to confront the vexing architecture selection problem by reducing high-dimensional gradient searchs to fitting single-input single-output (SISO) subnets. We introduce dimension reduction algorithms, to select features or relevant subsets of a set of many variables, based on minimizing an index of level-set dispersions (closely related to a projection index and to SIR), and combine them with backfitting to implement a neural network version of projection pursuit. The performance achieved by our approach, when trained on 1989, 1990 data and tested on 1991 data, is com(cid:173) parable to that achieved in our earlier study of backpropagation trained networks.
****
Visual Motion Computation in Analog VLSI Using Pulses
Rahul Sarpeshkar, Wyeth Bair, Christof Koch
The real time computation of motion from real images using a single chip with integrated sensors is a hard prob(cid:173) lem. We present two analog VLSI schemes that use pulse domain neuromorphic circuits to compute motion. Pulses of variable width, rather than graded potentials, represent a natural medium for evaluating temporal relationships. Both algorithms measure speed by timing a moving edge in the image. Our first model is inspired by Reichardt's algorithm in the fiy and yields a non-monotonic response vs. velocity curve. We present data from a chip that implements this model. Our second algorithm yields a monotonic response vs. velocity curve and is currently being translated into silicon.
****
Learning Control Under Extreme Uncertainty
Vijaykumar Gullapalli
A peg-in-hole insertion task is used as an example to illustrate the utility of direct associative reinforcement learning methods for learning control under real-world conditions of uncertainty and noise. Task complexity due to the use of an unchamfered hole and a clearance of less than 0.2mm is compounded by the presence of positional uncertainty of magnitude exceeding 10 to 50 times the clearance. Despite this extreme degree of uncertainty, our results indicate that direct reinforcement learning can be used to learn a robust reactive control strategy that results in skillful peg-in-hole insertions.
****
Efficient Pattern Recognition Using a New Transformation Distance
Patrice Simard, Yann LeCun, John Denker
Memory-based classification algorithms such as radial basis func(cid:173) tions or K-nearest neighbors typically rely on simple distances (Eu(cid:173) clidean, dot product ...), which are not particularly meaningful on pattern vectors. More complex, better suited distance measures are often expensive and rather ad-hoc (elastic matching, deformable templates). We propose a new distance measure which (a) can be made locally invariant to any set of transformations of the input and (b) can be computed efficiently. We tested the method on large handwritten character databases provided by the Post Office and the NIST. Using invariances with respect to translation, rota(cid:173) tion, scaling, shearing and line thickness, the method consistently outperformed all other systems tested on the same databases.
****
Planar Hidden Markov Modeling: From Speech to Optical Character Recognition
Esther Levin, Roberto Pieraccini

We propose in this paper a statistical model (planar hidden Markov model - PHMM) describing statistical properties of images. The model generalizes the single-dimensional HMM, used for speech processing, to the planar case. For this model to be useful an efficient segmentation algorithm, similar to the Viterbi algorithm for HMM, must exist We present conditions in terms of the PHMM parameters that are sufficient to guarantee that the plan ar segmentation problem can be solved in polynomial time, and describe an algorithm for that. This algorithm aligns optimally the image with the mo del, and therefore is insensitive to elastic distortions of images. Us ing this algorithm a joint optima1 segmentation and recognition of the image can be performed, thus overcoming the weakness of traditional OCR systems w here segmentation is performed independently before the recognition lead ing to unrecoverable recognition errors.
****
Global Regularization of Inverse Kinematics for Redundant Manipulators
David DeMers, Kenneth Kreutz-Delgado
The inverse kinematics problem for redundant manipulators is ill-posed and nonlinear. There are two fundamentally different issues which result in th e need for some form of regularization; the existence of multiple soluti on branches (global ill-posedness) and the existence of excess degrees of freed om (local ill(cid:173) posedness). For certain classes of manipulators, lear ning methods applied to input-output data generated from the forward functi on can be used to globally regularize the problem by partitioning the domain of the forward mapping into a finite set of regions over which the invers e problem is well-posed. Local regularization can be accomplished by an appropriate parameterization of the redundancy consistently over each region . As a result, the ill-posed problem can be transformed into a finite set of w ell-posed problems. Each can then be solved separately to construct approximate direct inverse functions.
****
Object-Based Analog VLSI Vision Circuits
Christof Koch, Binnal Mathur, Shih-Chii Liu, John Harris, Jin Luo, Massimo Sivil otti
We describe two successfully working, analog VLSI vision circuits that m ove beyond pixel-based early vision algorithms. One circuit, implementing the dynamic wires model, provides for dedicated lines of communication among grou ps of pixels that share a common property. The chip uses the dynamic w ires model to compute the arclength of visual contours. Another circuit labe ls all points inside a given contour with one voltage and all other with another volt(cid:173) age. Its behavior is very robust, since small break s in contours are automatically sealed, providing for Figure-Ground segrega tion in a noisy environment. Both chips are implemented using networks of resistors and switches and represent a step towards object level proc essing since a single voltage value encodes the property of an ensemble o f pixels.
****
Perceiving Complex Visual Scenes: An Oscillator Neural Network Model that Integr ates Selective Attention, Perceptual Organisation, and Invariant Recognition
Rainer Goebel
Which processes underly our ability to quickly recognize familiar objects withi n a complex visual input scene? In this paper an imple(cid:173) mented neural ne twork model is described that attempts to specify how selective visual attentio n, perceptual organisation, and invari(cid:173) ance transformations might work together in order to segment, select, and recognize objects out of complex inpu t scenes containing multi(cid:173) ple, possibly overlapping objects. Retinotopi cally organized feature maps serve as input for two main processing routes: pa thway' dealing with location information and the 'what-pathway' computing the s hape and attributes of objects. A location-based at(cid:173) tention mechanism o perates on an early stage of visual processing selecting a contigous region of the visual field for preferential proces(cid:173) sing. Additionally, location-b

ased attention plays an important role for invariant object recognition controlling appropriate normalization processes within the what-pathway. Object recognition is supported through the segmentation of the visual field into distinct entities. In order to represent different segmented entities at the same time, the model uses an oscillatory binding mechanism. Connections between the where-pathway and the what-pathway lead to a flexible coope(cid:173) ration between different functional subsystems producing an overall behavior which is consistent with a variety of psychophysical data.
****
Second order derivatives for network pruning: Optimal Brain Surgeon
Babak Hassibi, David Stork
We investigate the use of information from all second order derivatives of the error function to perfonn network pruning (i.e., removing unimportant weights from a trained network) in order to improve generalization, simplify networks, reduce hardware or storage requirements, increase the speed of further training, and in some cases enable rule extraction. Our method, Optimal Brain Surgeon (OBS), is Significantly better than magnitude-based methods and Optimal Brain Damage [Le Cun, Denker and Sol1a, 1990], which often remove the wrong weights. OBS permits the pruning of more weights than other methods (for the same error on the training set), and thus yields better generalization on test data. Crucial to OBS is a recursion relation for calculating the inverse Hessian matrix H-I from training data and structural information of the net. OBS permits a 90%, a 76%, and a 62% reduction in weights over backpropagation with weighL decay on three benchmark MONK's problems [Thrun et aI., 1991]. Of OBS, Optimal Brain Damage, and magnitude-based methods, only OBS deletes the correct weights from a trained XOR network in every case. Finally, whereas Sejnowski and Rosenberg [1987J used 18,000 weights in their NETtalk network, we used OBS to prune a network to just 1560 weights, yielding better generalization.
****
Automatic Learning Rate Maximization by On-Line Estimation of the Hessian's Eigenvectors
Yann LeCun, Patrice Simard, Barak Pearlmutter
We propose a very simple, and well principled way of computing the optimal step size in gradient descent algorithms. The on-line version is very efficient computationally, and is applicable to large backpropagation networks trained on large data sets. The main ingredient is a technique for estimating the principal eigenvalue(s) and eigenvector(s) of the objective function's second derivative ma(cid:173) trix (Hessian), which does not require to even calculate the Hes(cid:173) sian. Several other applications of this technique are proposed for speeding up learning, or for eliminating useless parameters.
****
Computation of Heading Direction from Optic Flow in Visual Cortex
Markus Lappe, Josef Rauschecker
We have designed a neural network which detects the direction of ego(cid:173) motion from optic flow in the presence of eye movements (Lappe and Rauschecker, 1993). The performance of the network is consistent with human psychophysical data, and its output neurons show great similarity to "triple component" cells in area MSTd of monkey visual cortex. We now show that by using assumptions about the kind of eye movements that the obsenrer is likely to perform, our model can generate various other cell types found in MSTd as well.
****
Statistical Modeling of Cell Assemblies Activities in Associative Cortex of Behaving Monkeys
Itay Gat, Naftali Tishby
So far there has been no general method for relating extracellular electrophysiological measured activity of neurons in the associative cortex to underlying network or "cognitive" states. We propose to model such data using a multivariate Poisson Hidden Markov Model. We demonstrate the application of this approach for tem(cid:173) poral segmentation of the firing patterns, and for characterization of the cortical responses to external stimuli. Using such a statisti(cid:1

73) cal model we can significantly discriminate two behavioral modes of the mon key, and characterize them by the different firing pat(cid:173) terns, as well a s by the level of coherency of their multi-unit firing activity. Our study uti lized measurements carried out on behaving Rhesus monkeys by M. Abeles, E. Vaad ia, and H. Bergman, of the Hadassa Medical School of the Hebrew University.
****

Harmonic Grammars for Formal Languages

Paul Smolensky

Basic connectionist principles imply that grammars should take the form of syst ems of parallel soft constraints defining an optimization problem the solutio ns to which are the well-formed structures in the language. Such Harm onic Grammars have been successfully applied to a number of problems in the theory of natural languages. Here it is shown that formal languages too can be specified by Harmonic Grammars, rather than by conventional s erial re-write rule systems.
****

Silicon Auditory Processors as Computer Peripherals

John Lazzaro, John Wawrzynek, M. Mahowald, Massimo Sivilotti, Dave Gillespie

Sever

Present address: f\1. Mahowald, f\1H.C ,\natol1lical Ncmophamacology Unit, Mans field TId, Oxfc)('d OXI :1'rl£ Ellgland . mam~vax.oxford.ac.uk t Present addres s: f\lass Siviloui. '1'(1111)(,1' H,csearrh, 180 Nort.h Vinedo Avenue, Pasaden a, CA 9Il07. mass~tanner. corn :I: Present address: Dave Gill,>spiE', SYllapf,i cs, :l()!)8 Orchard Parkway, San Jose CA, 95131. daveg~synaptics. com

****

Information, Prediction, and Query by Committee

Yoav Freund, H. Sebastian Seung, Eli Shamir, Naftali Tishby

We analyze the "query by committee" algorithm, a method for fil(cid:173) tering informative queries from a random stream of inputs. We show tha t if the two-member committee algorithm achieves infor(cid:173) mation gain with positive lower bound, then the prediction error decreases exponent ially with the number of queries. We show that, in particular, this exponen tial decrease holds for query learning of thresholded smooth functions.
****

Improving Convergence in Hierarchical Matching Networks for Object Recognition

Joachim Utans, Gene Gindi

We are interested in the use of analog neural networks for recog(cid:1 73) nizing visual objects. Objects are described by the set of parts t hey are composed of and their structural relationship. Struc(cid:173) tu ral models are stored in a database and the recognition prob(cid:173) lem reduces to matching data to models in a structurally consis(cid:17 3) tent way. The object recognition problem is in general very diffi(ci d:173) cult in that it involves coupled problems of grouping, segmentation and matching. We limit the problem here to the simultaneous la(cid:173) bellin g of the parts of a single object and the determination of analog par ameters. This coupled problem reduces to a weighted match problem in w hich an optimizing neural network must min(cid:173) imize E(M, p) = LO'i MO'i WO'i(p), where the {MO'd are binary match variables for data part s i to model parts a and {Wai(P)} are weights dependent on parameters p . In this work we show that by first solving for estimates p without sol ving for M ai , we may obtain good initial parameter estimates that yield b etter solutions for M and p.
****

A dynamical model of priming and repetition blindness

Daphne Bavelier, Michael Jordan

We describe a model of visual word recognition that accounts for several aspect s of the temporal processing of sequences of briefly presented words. The model

utilizes a new representation for writ(cid:173) ten words, based on dynamic time warping and multidimensional scaling. The visual input passes through cascaded perceptual, com(cid:173) parison, and detection stages. We describe how these dynamical processes can account for several aspects of word recognition, in(cid:173) cluding repetition priming and repetition blindness.
****
Unsupervised Discrimination of Clustered Data via Optimization of Binary Information Gain
Nicol Schraudolph, Terrence J. Sejnowski
We present the information-theoretic derivation of a learning algorithm that clusters unlabelled data with linear discriminants. In contrast to methods that try to preserve information about the input patterns, we maximize the information gained from observing the output of robust binary discriminators implemented with sigmoid nodes. We deri ve a local weight adaptation rule via gradient ascent in this objective, demonstrate its dynamics on some simple data sets, relate our approach to previous work and suggest directions in which it may be extended.
****
On-Line Estimation of the Optimal Value Function: HJB- Estimators
James Peterson
In this paper, we discuss on-line estimation strategies that model the optimal value function of a typical optimal control problem. We present a general strategy that uses local corridor solutions obtained via dynamic programming to provide local optimal con(cid:173) trol sequence training data for a neural architecture model of the optimal value function.
****
Using Aperiodic Reinforcement for Directed Self-Organization During Development
P. Montague, P. Dayan, S.J. Nowlan, A Pouget, T.J. Sejnowski
We present a local learning rule in which Hebbian learning is conditional on an incorrect prediction of a reinforcement signal. We propose a biological interpretation of such a framework and display its utility through examples in which the reinforcement signal is cast as the delivery of a neuromodulator to its target. Three exam pIes are presented which illustrate how this framework can be applied to the development of the oculomotor system.
****
A Connectionist Symbol Manipulator That Discovers the Structure of Context-Free Languages
Michael C. Mozer, Sreerupa Das
We present a neural net architecture that can discover hierarchical and re(cid:173) cursive structure in symbol strings. To detect structure at multiple levels, the architecture has the capability of reducing symbols substrings to single symbols, and makes use of an external stack memory. In terms of formal languages, the architecture can learn to parse strings in an LR(O) context(cid:173) free grammar. Given training sets of positive and negative exemplars, the architecture has been trained to recognize many different grammars. The architecture has only one layer of modifiable weights, allowing for a straightforward interpretation of its behavior.
****
Analogy-- Watershed or Waterloo? Structural alignment and the development of connectionist models of analogy
Dedre Gentner, Arthur Markman
Neural network models have been criticized for their inability to make use of compositional representations. In this paper, we describe a series of psychological phenomena that demonstrate the role of structured representations in cognition. These findings suggest that people compare relational representations via a process of structural alignment. This process will have to be captured by any model of cognition, symbolic or subsymbolic.
****
A Recurrent Neural Network for Generation of Occular Saccades
Lina L.E. Massone

This paper presents a neural network able to control saccadic movements. The in
put to the network is a specification of a stimulation site on the collicular m
otor map. The output is the time course of the eye position in the orbit (horiz
ontal and vertical angles). The units in the network exhibit a one-to-one corre
spondance with neurons in the intermediate layer of the superior colliculus (co
llicular motor map), in the brainstem and with oculomotor neurons. Simulations
 carried out with this network demonstrate its ability to reproduce in a straig
htforward fashion many experimental observations.
****
Network Structuring and Training Using Rule-based Knowledge
Volker Tresp, Jürgen Hollatz, Subutai Ahmad
We demonstrate in this paper how certain forms of rule-based knowledge can be u
sed to prestructure a neural network of nor(cid:173) malized basis functions and
 give a probabilistic interpretation of the network architecture. We describe s
everal ways to assure that rule-based knowledge is preserved during training an
d present a method for complexity reduction that tries to minimize the num(cid:
173) ber of rules and the number of conjuncts. After training the refined rules
 are extracted and analyzed.
****
Hybrid Circuits of Interacting Computer Model and Biological Neurons
Sylvie Masson, Gwendal Le Masson, Eve Marder, L. Abbott
We demonstrate the use of a digital signal processing board to construct hybrid
 networks consisting of computer model neurons connected to a biological neural
  network. This system operates in real time. and the synaptic connecti
ons are realistic effective conductances. Therefore. the synapses made
from the computer model neuron are integrated correctly by the postsyna
ptic biological neuron. This method provides us with the ability to add
additional. completely known elements to a biological network and study t
heir effect on network activity. Moreover. by changing the parameters
of the model neuron. it is possible to assess the role of individual cond
uctances in the activity of the neuron. and in the network in which it pa
rticipates.
****
A Knowledge-Based Model of Geometry Learning
Geoffrey Towell, Richard Lehrer
We propose a model of the development of geometric reasoning in children that e
xplicitly involves learning. The model uses a neural network that is initialized
  with an understanding of geometry similar to that of second-grade children. T
hrough the presentation of a series of examples, the model is shown to develop
an understanding of geometry similar to that of fifth-grade children who were t
rained using similar materials.
****
Integration of Visual and Somatosensory Information for Preshaping Hand in Grasp
ing Movements
Yoji Uno, Naohiro Fukumura, Ryoji Suzuki, Mitsuo Kawato
The primate brain must solve two important problems in grasping move(cid:173) me
nts. The first problem concerns the recognition of grasped objects: spec
ifically, how does the brain integrate visual and motor information on a
grasped object? The second problem concerns hand shape planning: specifically,
how does the brain design the hand configuration suited to the shape of the ob
ject and the manipulation task? A neural network model that solves these probl
ems has been developed. The operations of the net(cid:173) work are divided int
o a learning phase and an optimization phase. In the learning phase, internal
representations, which depend on the grasped ob(cid:173) jects and the task, ar
e acquired by integrating visual and somatosensory information. In the optimi
zation phase, the most suitable hand shape for grasping an object is determine
d by using a relaxation computation of the network.
****
Kohonen Feature Maps and Growing Cell Structures - a Performance Comparison
Bernd Fritzke

A performance comparison of two self-organizing networks, the Ko(cid:173) honen Feature Map and the recently proposed Growing Cell Struc(cid:173) tures is made. For this purpose several performance criteria for self-organizing networks are proposed and motivated. The models are tested with three ex ample problems of increasing difficulty. The Kohonen Feature Map demonstrat es slightly superior results only for the simplest problem. For the other m ore difficult and also more realistic problems the Growing Cell Structures exhi bit significantly better performance by every criterion . Additional adva ntages of the new model are that all parameters are constant over time a nd that size as well as structure of the network are determined auto(c id:173) matically.
****
A Neural Model of Descending Gain Control in the Electrosensory System
Mark Nelson
In the electrosensory system of weakly electric fish, descending pathwa ys to a first-order sensory nucleus have been shown to influ(cid:173) ence the g ain of its output neurons. The underlying neural mecha(cid:173) nisms that sub serve this descending gain control capability are not yet fully understood. We suggest that one possible gain control mechanism could involve the re gulation of total membrane conduc(cid:173) tance of the output neurons. In thi s paper, a neural model based on this idea is used to demonstrate how act ivity levels on descend(cid:173) ing pathways could control both the gain and baseline excitation of a target neuron .
****
Memory-Based Reinforcement Learning: Efficient Computation with Prioritized Swee ping
Andrew Moore, Christopher Atkeson
We present a new algorithm, Prioritized Sweeping, for efficient prediction a nd control of stochastic Markov systems. Incremental learning methods such as Temporal Differencing and Q-Iearning have fast real time perfor(cid:173) manc e. Classical methods are slower, but more accurate, because they make full use of the observations. Prioritized Sweeping aims for the best of bo th worlds. It uses all previous experiences both to prioritize impor(cid:1 73) tant dynamic programming sweeps and to guide the exploration of state(cid:1 73) space. We compare Prioritized Sweeping with other reinforcement learning s chemes for a number of different stochastic optimal control problems. It suc cessfully solves large state-space real time problems with which other m ethods have difficulty.
****
Learning Fuzzy Rule-Based Neural Networks for Control
Charles Higgins, Rodney Goodman
A three-step method for function approximation with a fuzzy sys(cid:173) t em is proposed. First, the membership functions and an initial rule r epresentation are learned; second, the rules are compressed as much as possible using information theory; and finally, a com(cid:173) putational network is constructed to compute the function value. This system is a pplied to two control examples: learning the truck and trailer backer-upper control system, and learning a cruise con(cid:173) trol system for a radio-c ontrolled model car.
****
Neural Network On-Line Learning Control of Spacecraft Smart Structures
Christopher Bowman
The overall goal is to reduce spacecraft weight. volume, and cost by on(cid:173) line adaptive non-linear control of flexible structural components. The object ive of this effort is to develop an adaptive Neural Network (NN) controller fo r the Ball C-Side 1m x 3m antenna with embedded actuators and the RAMS sensor system. A traditional optimal controller for the major modes is provided pe rturbations by the NN to compensate for unknown residual modes. On-line tr aining of recurrent and feed-forward NN architectures have achieved adaptiv e vibration control with unknown modal variations and noisy measurements

. On-line training feedback to each actuator NN output is computed via Newton's method to reduce the difference between desired and achieved antenna positions.
****
Hidden Markov Model Induction by Bayesian Model Merging
Andreas Stolcke, Stephen Omohundro
This paper describes a technique for learning both the number of states and the topology of Hidden Markov Models from examples. The induction process starts with the most specific model consistent with the training data and generalizes by successively merging states. Both the choice of states to merge and the stopping criterion are guided by the Bayesian posterior probability. We compare our algorithm with the Baum-Welch method of estimating fixed-size models, and find that it can induce minimal HMMs from data in cases where fixed estimation does not converge or requires redundant parameters to converge.
****
Statistical Mechanics of Learning in a Large Committee Machine
Holm Schwarze, John Hertz
We use statistical mechanics to study generalization in large com(cid:173) mittee machines. For an architecture with nonoverlapping recep(cid:173) tive fields a replica calculation yields the generalization error in the limit of a large number of hidden units. For continuous weights the generalization error falls off asymptotically inversely proportional to Q, the number of training examples per weight. For binary weights we find a discontinuous transition from poor to perfect generalization followed by a wide region of metastability. Broken replica symmetry is found within this region at low temperatures. For a fully connected architecture the generalization error is cal(cid:173) culated within the annealed approximation. For both binary and continuous weights we find transitions from a symmetric state to one with specialized hidden units, accompanied by discontinuous drops in the generalization error.
****
An Information-Theoretic Approach to Deciphering the Hippocampal Code
William Skaggs, Bruce McNaughton, Katalin Gothard
Information theory is used to derive a simple formula for the amount of information conveyed by the firing rate of a neuron about any experimentally measured variable or combination of variables (e.g. running speed, head direction, location of the animal, etc.). The derivation treats the cell as a communication channel whose input is the measured variable and whose output is the cell's spike train. Applying the formula, we find systematic differences in the information content of hippocampal "place cells" in different ex(cid:173) perimental conditions.
****
Using hippocampal 'place cells' for navigation, exploiting phase coding
Neil Burgess, John O'Keefe, Michael Recce
A model of the hippocampus as a central element in rat naviga(cid:173) tion is presented. Simulations show both the behaviour of single cells and the resultant navigation of the rat. These are compared with single unit recordings and behavioural data. The firing of CAl place cells is simulated as the (artificial) rat moves in an en(cid:173) vironment. This is the input for a neuronal network whose output, at each theta (0) cycle, is the next direction of travel for the rat. Cells are characterised by the number of spikes fired and the time of firing with respect to hippocampal 0 rhythm. 'Learning' occurs in 'on-off' synapses that are switched on by simultaneous pre- and post-synaptic activity. The simulated rat navigates successfully to goals encountered one or more times during exploration in open fields. One minute of random exploration of a 1m2 environment allows navigation to a newly-presented goal from novel starting po(cid:173) sitions. A limited number of obstacles can be successfully avoided.
****
Weight Space Probability Densities in Stochastic Learning: I. Dynamics and Equilibria

Todd Leen, John Moody
The ensemble dynamics of stochastic learning algorithms can be studied using theoretical techniques from statistical physics. We develop the eq uations of motion for the weight space probability densities for stochast ic learning algorithms. We discuss equilibria in the diffusion approximati on and provide expressions for special cases of the LMS algorithm. The equi librium densities are not in general thermal (Gibbs) distributions in the obj ective function be(cid:173) ing minimized, but rather depend upon an effective p otential that includes diffusion effects. Finally we present an exact a nalytical expression for the time evolution of the density for a learning algo(cid:173) rithm with weight updates proportional to the sign of the gradient.
****
Discriminability-Based Transfer between Neural Networks
L. Y. Pratt
Previously, we have introduced the idea of neural network transfer, where learn ing on a target problem is sped up by using the weights obtained from a network trained for a related source task. Here, we present a new algorithm. called Di scriminability-Based Transfer (DBT), which uses an information measure to estim ate the utility of hyperplanes defined by source weights in the target network, and rescales transferred weight magnitudes accordingly. Several experiments d emonstrate that target networks initialized via DBT learn significantly faster than networks initialized randomly.
****
Learning Sequential Tasks by Incrementally Adding Higher Orders
Mark Ring
An incremental, higher-order, non-recurrent network combines two properties fou nd to be useful for learning sequential tasks: higher(cid:173) order connections and incremental introduction of new units. The network adds higher orders when needed by adding new units that dynamically modify connection weights. Since t he new units mod(cid:173) ify the weights at the next time-step with information from the previous step, temporal tasks can be learned without the use of feed back, thereby greatly simplifying training. Furthermore, a the(cid:173) oretical ly unlimited number of units can be added to reach into the arbitrarily distant past. Experiments with the Reber gram(cid:173) mar have demonstrated speedups o f two orders of magnitude over recurrent networks.
****
Topography and Ocular Dominance with Positive Correlations
Geoffrey Goodhill
A new computational model that addresses the formation of both topog-
raphy and ocular dominance is presented. This is motivated by exper-
imental evidence that these phenomena may be subserved by the same
mechanisms. An important aspect of this model is that ocular domi-
nance segregation can occur when input activity is both distributed, and
positively correlated between the eyes. This allows investigation of the
dependence of the pattern of ocular dominance stripes on the degree of
correlation between the eyes: it is found that increasing correlation leads
to narrower stripes. Experiments are suggested to test whether such be-
haviour occurs in the natural system.
****
A Hybrid Neural Net System for State-of-the-Art Continuous Speech Recognition
G. Zavaliagkos, Y. Zhao, R. Schwartz, J. Makhoul
Untill recently, state-of-the-art, large-vocabulary, continuous speech recognit ion (CSR) has employed Hidden Markov Modeling (HMM) to model speech sounds. In an attempt to improve over HMM we developed a hybrid system that integrates HM M technology with neu(cid:173) ral networks. We present the concept of a "Segmen tal Neural Net" (SNN) for phonetic modeling in CSR. By taking into account all the frames of a phonetic segment simultaneously, the SNN overcomes the well-kn own conditional-independence limitation of HMMs. In several speaker-independent experiments with the DARPA Resource Manage(cid:173) ment corpus, the hybrid sys tem showed a consistent improvement in performance over the baseline HMM system

.
****
A Formal Model of the Insect Olfactory Macroglomerulus: Simulations and Analytic Results
Christiane Linster, David Marsan, Claudine Masson, Michel Kerszberg, Gérard Dreyfus, Léon Personnaz
It is known from biological data that the response patterns of interneurons in the olfactory macroglomerulus (MGC) of insects are of central importance for the coding of the olfactory signal. We propose an analytically tractable model of the MGC which allows us to relate the distribution of response patterns to the architecture of the network.
****
Using Prior Knowledge in a NNPDA to Learn Context-Free Languages
Sreerupa Das, C. Giles, Guo-Zheng Sun
Although considerable interest has been shown in language inference and automata induction using recurrent neural networks, success of these models has mostly been limited to regular languages. We have previ(cid:173) ously demonstrated that Neural Network Pushdown Automaton (NNPDA) model is capable of learning deterministic context-free languages (e.g., anbn and parenthesis languages) from examples. However, the learning task is computationally intensive. In this paper we discus some ways in which a priori knowledge about the task and data could be used for efficient learning. We also observe that such knowledge is often an experimental prerequisite for learning nontrivial languages (eg. anbncbmam).
****
A Method for Learning From Hints
Yaser Abu-Mostafa
We address the problem of learning an unknown function by pu tting together several pieces of information (hints) that we know about the function. We introduce a method that generalizes learn(cid:173) ing from examples to learning from hints. A canonical representa(cid:173) tion of hints is defined and illustrated for new types of hints. All the hints are represented to the learning process by examples, and examples of the function are treated on equal footing with the rest of the hints. During learning, examples from different hints are selected for processing according to a given schedule. We present two types of schedules; fixed schedules that specify the relative em(cid:173) phasis of each hint, and adaptive schedules that are based on how well each hint has been learned so far. Our learning method is compatible with any descent technique that we may choose to use.
****
Holographic Recurrent Networks
Tony A. Plate
Holographic Recurrent Networks (HRNs) are recurrent networks which incorporate associative memory techniques for storing se(cid:173) quential structure. HRNs can be easily and quickly trained using gradient descent techniques to generate sequences of discrete out(cid:173) puts and trajectories through continuous spaee. The performance of HRNs is found to be superior to that of ordinary recurrent net(cid:173) works on these sequence generation tasks.
****
Generalization Abilities of Cascade Network Architecture
E. Littmann, H. Ritter
In [5], a new incremental cascade network architecture has been presented. This paper discusses the properties of such cascade networks and investigates their generalization abilities under the particular constraint of small data sets. The evaluation is done for cascade networks consisting of local linear maps using the Mackey(cid:173) Glass time series prediction task as a benchmark. Our results in(cid:173) dicate that to bring the potential of large networks to bear on the problem of extracting information from small data sets without run(cid:173) ning the risk of overjitting, deeply cascaded network architectures are more favorable than shallow broad architectures that contain the same number of nodes
.

****
Explanation-Based Neural Network Learning for Robot Control
Tom M. Mitchell, Sebastian B. Thrun
How can artificial neural nets generalize better from fewer examples? In order to generalize successfully, neural network learning methods typically require l arge training data sets. We introduce a neural network learning method that gen eralizes rationally from many fewer data points, relying instead on prior knowl edge encoded in previously learned neural networks. For example, in robot contr ol learning tasks reported here, previously learned networks that model the eff ects of robot actions are used to guide subsequent learning of robot control fu nctions. For each observed training example of the target function (e.g. the ro bot control policy), the learner explains the observed example in terms of its prior knowledge, then analyzes this explanation to infer additional information about the shape, or slope, of the target function. This shape knowledge is used to bias generalization when learning the target function. Results are presente d applying this approach to a simulated robot task based on reinforcement learn ing.
****
Some Solutions to the Missing Feature Problem in Vision
Subutai Ahmad, Volker Tresp
In visual processing the ability to deal with missing and noisy informa(cid:173) tion is crucial. Occlusions and unreliable feature detectors often lead to si tuations where little or no direct information about features is availa(cid:1 73) ble. However the available information is usually sufficient to hig hly constrain the outputs. We discuss Bayesian techniques for extractin g class probabilities given partial data. The optimal solution involves inte(ci d:173) grating over the missing dimensions weighted by the local probabil ity densities. We show how to obtain closed-form approximations to the Bayesian solution using Gaussian basis function networks. The frame(cid: 173) work extends naturally to the case of noisy features. Simulations on a complex task (3D hand gesture recognition) validate the theory. When bot h integration and weighting by input densities are used, performance decreases gracefully with the number of missing or noisy features. Per(cid:173) formance is substantially degraded if either step is omitted.
****
A Note on Learning Vector Quantization
Virginia de, Dana Ballard
Vector Quantization is useful for data compression. Competitive Learn(cid:173) ing which minimizes reconstruction error is an appropriate algorithm for vector quantization of unlabelled data. Vector quantization of labelled data for cla ssification has a different objective, to minimize the number of misclassificat ions, and a different algorithm is appropriate. We show that a variant of Koh onen's LVQ2.1 algorithm can be seen as a multi(cid:173) class extension o f an algorithm which in a restricted 2 class case can be proven to conver ge to the Bayes optimal classification boundary. We compare the performance of the LVQ2.1 algorithm to that of a modified version having a decreasing window and normalized step size, on a ten class vowel classification problem.
****
Diffusion Approximations for the Constant Learning Rate Backpropagation Algorith m and Resistence to Local Minima
William Finnoff
In this paper we discuss the asymptotic properties of the most com(cid:173) monl y used variant of the backpropagation algorithm in which net(cid:173) work weigh ts are trained by means of a local gradient descent on ex(cid:173) amples drawn randomly from a fixed training set, and the learning rate TJ of the gradient up dates is held constant (simple backpropa(cid:173) gation). Using stochastic appr oximation results, we show that for TJ ~ 0 this training process approaches a b atch training and pro(cid:173) vide results on the rate of convergence. Further, we show that for small TJ one can approximate simple back propagation by the s um of a batch training process and a Gaussian diffusion which is the unique so

lution to a linear stochastic differential equation. Using this approximation we indicate the reasons why simple backprop(cid:173) agation is less likely to get stuck in local minima than the batch training process and demonstrate this empirically on a number of examples.
****
Analog Cochlear Model for Multiresolution Speech Analysis
Weimin Liu, Andreas Andreou, Moise Goldstein
This paper discusses the parameterization of speech by an analog cochlear model. The tradeoff between time and frequency resolution is viewed as the fundamental difference between conventional spectrographic analysis and cochlear signal processing for broadband, rapid-changing signals. The model's response exhibits a wavelet-like analysis in the scale domain that preserves good temporal resolution; the frequency of each spectral compo(cid:173) nent in a broadband signal can be accurately determined from the inter(cid:173) peak intervals in the instantaneous firing rates of auditory fibers. Such properties of the cochlear model are demonstrated with natural speech and synthetic complex signals.
****
On Learning µ-Perceptron Networks with Binary Weights
Mostefa Golea, Mario Marchand, Thomas Hancock
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Boundary Hunting Radial Basis Function Classifier which Allocates Centers Constructively
Eric Chang, Richard P. Lippmann
A new boundary hunting radial basis function (BH-RBF) classifier which allocates RBF centers constructively near class boundaries is described. This classifier creates complex decision boundaries only in regions where confusions occur and corresponding RBF outputs are similar. A predicted square error measure is used to determine how many centers to add and to determine when to stop adding centers. Two experiments are presented which demonstrate the advantages of the BH(cid:173) RBF classifier. One uses artificial data with two classes and two input features where each class contains four clusters but only one cluster is near a decision region boundary. The other uses a large seismic database with seven classes and 14 input features. In both experiments the BH(cid:173) RBF classifier provides a lower error rate with fewer centers than are required by more conventional RBF, Gaussian mixture, or MLP classifiers.
****
Learning to See Where and What: Training a Net to Make Saccades and Recognize Handwritten Characters
Gale Martin, Mosfeq Rashid, David Chapman, James Pittman
to integrated segmentation and This paper describes an approach recognition of hand-printed characters. The approach, called Saccade, integrates ballistic and corrective saccades (eye movements) with character recognition. A single backpropagation net is trained to make a classification decision on a character centered in its input window, as well as to estimate the distance of the current and next character from the center of the input window. The net learns to accurately estimate these distances regardless of variations in character width, spacing between characters, writing style and other factors. During testing, the system uses the net~xtracted classification and distance information, along with a set of jumping rules, to jump from character to character.
****
How Oscillatory Neuronal Responses Reflect Bistability and Switching of the Hidden Assembly Dynamics
K. Pawelzik, H.-U. Bauer, J. Deppisch, T. Geisel

A switching between apparently coherent (oscillatory) and stochastic episodes o f activity has been observed in responses from cat and monkey visual cortex. We describe the dynamics of these phenomena in two paral(cid:173) lel approaches, a phenomenological and a rather microscopic one. On the one hand we analyze neu ronal responses in terms of a hidden state model (HSM). The parameters of this model are extracted directly from exper(cid:173) imental spike trains. They char acterize the underlying dynamics as well as the coupling of individual neurons to the network. This phenomenolog(cid:173) ical model thus provides a new framew ork for the experimental analysis of network dynamics. The application of this method to multi unit ac(cid:173) tivities from the visual cortex of the cat subs tantiates the existence of oscillatory and stochastic states and quantifies the switching behaviour in the assembly dynamics. On the other hand we start from the single spiking neuron and derive a master equation for the time evolution o f the assembly state which we represent by a phase density. This phase density dynamics (PDD) exhibits costability of two attractors, a limit cycle, and a fi xed point when synaptic interaction is nonlinear. External fluctuations can swi tch the bistable system from one state to the other. Finally we show, that the two approaches are mutually consistent and therefore both explain the detailed time structure in the data.
****

Summed Weight Neuron Perturbation: An O(N) Improvement Over Weight Perturbation
Barry Flower, Marwan Jabri
The algorithm presented performs gradient descent on the weight space of an Art ificial Neural Network (ANN), using a finite difference to approximate the grad ient The method is novel in that it achieves a com(cid:173) putational complexit y similar to that of Node Perturbation, O(N3), but does not require access to t he activity of hidden or internal neurons. This is possible due to a stochastic relation between perturbations at the weights and the neurons of an ANN. The a lgorithm is also similar to Weight Perturbation in that it is optimal in terms of hardware require(cid:173) ments when used for the training ofVLSI implementat ions of ANN's.
****

Extended Regularization Methods for Nonconvergent Model Selection
W. Finnoff, F. Hergert, H. G. Zimmermann
Many techniques for model selection in the field of neural networks correspond to well established statistical methods. The method of 'stopped training', on t he other hand, in which an oversized network is trained until the error on a fu rther validation set of ex(cid:173) amples deteriorates, then training is stoppe d, is a true innovation, since model selection doesn't require convergence of t he training process. In this paper we show that this performance can be signif icantly enhanced by extending the 'non convergent model selection method' of s topped training to include dynamic topology modifications (dynamic weight pruni ng) and modified complexity penalty term methods in which the weighting of the penalty term is adjusted during the training process.
****

Synchronization and Grammatical Inference in an Oscillating Elman Net
Bill Baird, Todd Troyer, Frank Eeckman
We have designed an architecture to span the gap between bio(cid:173) physics an d cognitive science to address and explore issues of how a discrete symbol proc essing system can arise from the continuum, and how complex dynamics like oscil lation and synchronization can then be employed in its operation and affect its learning. We show how a discrete-time recurrent "Elman" network architecture c an be constructed from recurrently connected oscillatory associative memory mo dules described by continuous nonlinear ordinary dif(cid:173) ferential equation s. The modules can learn connection weights be(cid:173) tween themselves which w ill cause the system to evolve under a clocked "machine cycle" by a sequence of transitions of attractors within the modules, much as a digital computer evolv es by transi(cid:173) tions of its binary flip-flop attractors. The architecture thus em(cid:173) ploys the principle of "computing with attractors" used by mac ro(cid:173) scopic systems for reliable computation in the presence of noise. We

have specifically constructed a system which functions as a finite state auto maton that recognizes or generates the infinite set of six symbol strings that are defined by a Reber grammar. It is a symbol processing system, but with anal og input and oscillatory subsym(cid:173) bolic representations. The time steps (machine cycles) of the sys(cid:173) tem are implemented by rhythmic variation (c locking) of a bifurca(cid:173) tion parameter. This holds input and "context" mo dules clamped at their attractors while 'hidden and output modules change state , then clamps hidden and output states while context modules are released to l oad those states as the new context for the next cycle of input. Superior noise immunity has been demonstrated for systems with dynamic attractors over system s with static attractors, and synchronization ("binding") between coupled oscil latory attractors in different modules has been shown to be important for effec ting reliable transitions.
****
Attractor Neural Networks with Local Inhibition: from Statistical Physics to a D igitial Programmable Integrated Circuit
E. Pasero, R. Zecchina
Networks with local inhibition are shown to have enhanced compu(cid:173) tationa l performance with respect to the classical Hopfield-like net(cid:173) works. In particular the critical capacity of the network is increased as well as its ca pability to store correlated patterns. Chaotic dy(cid:173) namic behaviour (expo nentially long transients) of the devices in(cid:173) dicates the overloading of the associative memory. An implementa(cid:173) tion based on a programmable log ic device is here presented. A 16 neurons circuit is implemented whit a XILINK 4020 device. The peculiarity of this solution is the possibility to change part s of the project (weights, transfer function or the whole architecture) with a simple software download of the configuration into the XILINK chip.
****
Metamorphosis Networks: An Alternative to Constructive Models
Brian Bonnlander, Michael C. Mozer
Given a set oft raining examples, determining the appropriate num(cid:173) ber of free parameters is a challenging problem. Constructive learning algor ithms attempt to solve this problem automatically by adding hidden units, an d therefore free parameters, during learn(cid:173) ing. We explore an a lternative class of algorithms-called meta(cid:173) morphosis algorithms-in which the number of units is fixed, but the number of free parameters g radually increases during learning. The architecture we investigate is compose d of RBF units on a lat(cid:173) tice, which imposes flexible constraints o n the parameters of the network. Virtues of this approach include varia ble subset selec(cid:173) tion, robust parameter selection, multiresolutio n processing, and interpolation of sparse training data.
****
A Hybrid Linear/Nonlinear Approach to Channel Equalization Problems
Wei-Tsih Lee, John Pearson
Channel equalization problem is an important problem in high-speed communicatio ns. The sequences of symbols transmitted are distorted by neighboring symbols. Traditionally, the channel equalization problem is considered as a channel-inve rsion operation. One problem of this approach is that there is no direct corres pondence between error proba(cid:173) bility and residual error produced by the channel inversion operation. In this paper, the optimal equalizer design is for mulated as a classification problem. The optimal classifier can be constructed by Bayes decision rule. In general it is nonlinear. An efficient hybrid linear/ nonlinear equalizer approach has been proposed to train the equalizer. The erro r probability of new linear/nonlinear equalizer has been shown to be bet(cid:17 3) ter than a linear equalizer in an experimental channel.
****
Physiologically Based Speech Synthesis
Makoto Hirayama, Eric Vatikiotis-Bateson, Kiyoshi Honda, Yasuharu Koike, Mitsuo Kawato
This study demonstrates a paradigm for modeling speech produc(cid:173) tion

based on neural networks. Using physiological data from speech utterances, a neural network learns the forward dynamics relating motor commands to muscles and the ensuing articulator behavior that allows articulator trajectories to be generated from motor commands constrained by phoneme input strings and global performance parameters. From these movement trajectories, a sec(cid:173) ond neural network generates PARCOR parameters that are then used to synthesize the speech acoustics.
****
Weight Space Probability Densities in Stochastic Learning: II. Transients and Basin Hopping Times
Genevieve Orr, Todd Leen
In stochastic learning, weights are random variables whose time evolution is governed by a Markov process. At each time-step, n, the weights can be described by a probability density function pew, n). We summarize the theory of the time evolution of P, and give graphical examples of the time evolution that contrast the behavior of stochastic learning with true gradient descent (batch learning). Finally, we use the formalism to obtain predictions of the time required for noise-induced hopping between basins of different optima. We compare the theoretical predictions with simulations of large ensembles of networks for simple problems in supervised and unsupervised learning.
****
Performance Through Consistency: MS-TDNN's for Large Vocabulary Continuous Speech Recognition
Joe Tebelskis, Alex Waibel
Connectionist Rpeech recognition systems are often handicapped by an inconsistency between training and testing criteria. This prob(cid:173) lem is addressed by the Multi-State Time Delay Neural Network (MS-TDNN), a hierarchical phonf'mp and word classifier which uses DTW to modulate its connectivit.y pattern, and which is directly trained on word-level targets. The consistent use of word accu(cid:173) racy as a criterion during bot.h t.raining and testing leads to very high system performance, even wif II limited training dat.a. Until now, the MS-TDN N has been appli('d primarily to small vocabu(cid:173) lary recognition and word spotting tasks. In this papf'f we apply the architecture to large vocabulary continuous speech recognition, and demonstrate that our MS-TDNN outperforms all ot,hf'r sys(cid:173) tems that have been tested on tht' eMU Conference Registration database.
****
Generic Analog Neural Computation - The EPSILON Chip
Stephen Churcher, Donald Baxter, Alister Hamilton, Alan Murray, H. Reekie
An analog CMOS VLSI neural processing chip has been designed and fabri(cid:173) cated. The device employs "pulse-stream" neural state signalljng, and is capa(cid:173) ble of computing some 360 million synaptic connections per secood. In addi(cid:173) tion to basic characterisation results. the performance of the chip in solving "real-world" problems is also demonstrated.
****
Analog VLSI Implementation of Multi-dimensional Gradient Descent
David B. Kirk, Douglas Kerns, Kurt Fleischer, Alan Barr
We describe an analog VLSI implementation of a multi-dimensional gradient estimation and descent technique for minimizing an on(cid:173) chip scalar function fO. The implementation uses noise injec(cid:173) tion and multiplicative correlation to estimate derivatives, as in [Anderson, Kerns 92]. One intended application of this technique is setting circuit parameters on-chip automatically, rather than manually [Kirk 91]. Gradient descent optimization may be used to adjust synapse weights for a backpropagation or other on-chip learning implementation. The approach combines the features of continuous multi-dimensional gradient descent and the potential for an annealing style of optimization. We present data measured from our analog VLSI implementation.
****

Improving Performance in Neural Networks Using a Boosting Algorithm
Harris Drucker, Robert Schapire, Patrice Simard
A boosting algorithm converts a learning machine with error rate less than 50% to one with an arbitrarily low error rate. However, the algorithm discussed her e depends on having a large supply of independent training samples. We show how to circumvent this problem and generate an ensemble of learning machines whose performance in optical character recognition problems is dramatically improve d over that of a single network. We report the effect of boosting on four datab ases (all handwritten) consisting of 12,000 digits from segmented ZIP codes fro m the United State Postal Service (USPS) and the following from the National In stitute of Standards and Testing (NIST): 220,000 digits, 45,000 upper case alph as, and 45,000 lower case alphas. We use two performance measures: the raw erro r rate (no rejects) and the reject rate required to achieve a 1% error rate on the patterns not rejected. Boosting improved performance in some cases by a fa ctor of three.
****
An Object-Oriented Framework for the Simulation of Neural Nets
A. Linden, Th. Sudbrak, Ch. Tietz, F. Weber
The field of software simulators for neural networks has been ex(cid:173) pandin g very rapidly in the last years but their importance is still being underestim ated. They must provide increasing levels of as(cid:173) sistance for the design , simulation and analysis of neural networks. With our object-oriented framewor k (SESAME) we intend to show that very high degrees of transparency, manageabil ity and flexibil(cid:173) ity for complex experiments can be obtained. SESAME's basic de(cid:173) sign philosophy is inspired by the natural way in which resear chers explain their computational models. Experiments are performed with netwo rks of building blocks, which can be extended very eas(cid:173) ily. Mechanisms have been integrated to facilitate the construction and analysis of very comple x architectures. Among these mech(cid:173) anisms are t.he automatic configurati on of building blocks for an experiment and multiple inheritance at run-time.
****
A Practice Strategy for Robot Learning Control
Terence Sanger
"Trajectory Extension Learning" is a new technique for Learning Control in Robots which assumes that there exists some parameter of the desired traject ory that can be smoothly varied from a region of easy solvability of the dynam ics to a region of desired behavior which may have more difficult dynamics. By gradually varying the parameter, practice movements remain near the desired path while a Neural Network learns to approximate the inverse dynamics. For example, the average speed of motion might be varied, and the in(cid:173) vers e dynamics can be "bootstrapped" from slow movements with simpler dynamics to fast movements. This provides an example of the more general concept of a "Practice Strategy" in which a se(cid:173) quence of intermediate tasks is used to simplify learning a complex task. I show an example of the appli cation of this idea to a real 2-joint direct drive robot arm.
****
Modeling Consistency in a Speaker Independent Continuous Speech Recognition Syst em
Yochai Konig, Nelson Morgan, Chuck Wooters, Victor Abrash, Michael Cohen, Horaci o Franco
We would like to incorporate speaker-dependent consistencies, such as gender , in an otherwise speaker-independent speech recognition system. In this paper we discuss a Gender Dependent Neural Network (GDNN) which can be tuned for eac h gender, while sharing most of the speaker independent parameters. We use a classification network to help generate gender-dependent phonetic probabilities for a statistical (HMM) recogni(cid:173) tion system. The gender classificatio n net predicts the gender with high accuracy, 98.3% on a Resource Management t est set. However, the in(cid:173) tegration of the GDNN into our hybrid HMM-ne ural network recognizer provided an improvement in the recognition score that i s not statistically significant on a Resource Management test set.

****
Learning to categorize objects using temporal coherence
Suzanna Becker
The invariance of an objects' identity as it transformed over time provides a powerful cue for perceptual learning. We present an un(cid:173) supervised learning procedure which maximizes the mutual infor(cid:173) mation between the representations adopted by a feed-forward net(cid:173) work at consecutive time steps. We demonstrate that the network can learn, entirely unsupervised, to classify an ensemble of several patterns by observing pattern trajectories, even though there are abrupt transitions from one object to another between trajecto(cid:173) ries. The same learning procedure should be widely applicable to a variety of perceptual learning tasks.
****
Learning Curves, Model Selection and Complexity of Neural Networks
Noboru Murata, Shuji Yoshizawa, Shun-ichi Amari
Learning curves show how a neural network is improved as the number of t.raiuing examples increases and how it is related to the network complexity. The present paper clarifies asymptotic properties and their relation of t.wo learning curves, one concerning the predictive loss or generalization loss and the other the training loss. The result gives a natural definition of the complexity of a neural network. Moreover, it provides a new criterion of model selection.
****
Optimal Depth Neural Networks for Multiplication and Related Problems
Kai-Yeung Siu, Vwani Roychowdhury
An artificial neural network (ANN) is commonly modeled by a threshold circuit, a network of interconnected processing units called linear threshold gates. The depth of a network represents the number of unit delays or the time for parallel computation. The SIze of a circuit is the number of gates and measures the amount of hardware . It was known that traditional logic circuits consisting of only unbounded fan-in AND, OR, NOT gates would require at least O(log n/log log n) depth to compute common arithmetic functions such as the product or the quotient of two n-bit numbers, unless we allow the size (and fan-in) to increase exponentially (in n). We show in this paper that ANNs can be much more powerful than traditional logic circuits. In particular, we prove that that iterated addition can be com(cid:173) puted by depth-2 ANN, and multiplication and division can be computed by depth-3 ANNs with polynomial size and polynomially bounded integer weights, respectively. Moreover, it follows from known lower bound re(cid:173) sults that these ANNs are optimal in depth. We also indicate that these techniques can be applied to construct polynomial-size depth-3 ANN for powering, and depth-4 ANN for mUltiple product.
****
Time Warping Invariant Neural Networks
Guo-Zheng Sun, Hsing-Hen Chen, Yee-Chun Lee
We proposed a model of Time Warping Invariant Neural Networks (TWINN)
****
Deriving Receptive Fields Using an Optimal Encoding Criterion
Ralph Linsker
An information-theoretic optimization principle ('infomax') has previously been used for unsupervised learning of statistical reg(cid:173) ularities in an input ensemble. The principle states that the input(cid:173) output mapping implemented by a processing stage should be cho(cid:173) sen so as to maximize the average mutual information between input and output patterns, subject to constraints and in the pres(cid:173) ence of processing noise. In the present work I show how infomax, when applied to a class of nonlinear input-output mappings, can under certain conditions generate optimal filters that have addi(cid:173) tional useful properties: (1) Output activity (for each input pat(cid:173) tern) tends to be concentrated among a relatively small number (2) The filters are sensitive to higher-order statistical of nodes. structure (beyond pairwise correlations

). If the input features are localized, the filters' receptive fields t end to be localized as well. (3) Multiresolution sets of filters with subsampling at low spatial frequencies - related to pyramid coding and wavele t representations - emerge as favored solutions for certain types of input en sembles.
****
Connected Letter Recognition with a Multi-State Time Delay Neural Network
Hermann Hild, Alex Waibel
The Multi-State Time Delay Neural Network (MS-TDNN) inte(cid:173) grates a nonlinear time alignment procedure (DTW) and the high(cid:173) accuracy pho neme spotting capabilities of a TDNN into a connec(cid:173) tionist speech rec ognition system with word-level classification and error backpropagation. We present an MS-TDNN for recognizing continuously spelled letters, a tas k characterized by a small but highly confusable vocabulary. Our MS-TDNN achieves 98.5/92.0% word accuracy on speaker dependent/independent tasks , outper(cid:173) forming previously reported results on the same databases. W e pro(cid:173) pose training techniques aimed at improving sentence level per for(cid:173) mance, including free alignment across word boundaries, word du(cid:173) ration modeling and error backpropagation on the sentence rather than the word level. Architectures integrating submodules special(cid:173) iz ed on a subset of speakers achieved further improvements.
****
Input Reconstruction Reliability Estimation
Dean A. Pomerleau
This paper describes a technique called Input Reconstruction Reliability Estimat ion (IRRE) for determining the response reliability of a restricted class of mu lti-layer perceptrons (MLPs). The technique uses a network's ability to accurat ely encode the input pattern in its internal representation as a measure of its reliability. The more accurately a network is able to reconstruct the input pa ttern from its internal representation, the more reliable the network is consid ered to be. IRRE is provides a good estimate of the reliability of MLPs trained for autonomous driving. Results are presented in which the reliability estimat es provided by IRRE are used to select between networks trained for different d riving situations.
****
Mapping Between Neural and Physical Activities of the Lobster Gastric Mill
Kenji Doya, Mary Boyle, Allen Selverston
A computer model of the musculoskeletal system of the lobster gastric mill was constructed in order to provide a behavioral in(cid:173) terpretation of the rhy thmic patterns obtained from isolated stom(cid:173) atogastric ganglion. The mod el was based on Hill's muscle model and quasi-static approximation of the skele tal dynamics and could simulate the change of chewing patterns by the effect of neuromod(cid:173) ulators.
****
A Fast Stochastic Error-Descent Algorithm for Supervised Learning and Optimizati on
Gert Cauwenberghs
A parallel stochastic algorithm is investigated for error-descent learni ng and optimization in deterministic networks of arbitrary topology. No explicit information about internal network struc(cid:173) ture is needed. The method is based on the model-free distributed learning mechanism of Dembo and Kailath. A modified parameter update rule is proposed by which each indiv idual parameter vector perturbation contributes a decrease in error. A substa ntially faster learning speed is hence allowed. Furthermore, the modified algo(cid:173) rithm supports learning time-varying features in dynamical net(cid:173) works. We analyze the convergence and scaling properties of the algorithm, and present simulation results for dynamic trajectory learning in recurrent networks.
****
Nets with Unreliable Hidden Nodes Learn Error-Correcting Codes

Stephen Judd, Paul Munro
In a multi-layered neural network, anyone of the hidden layers can be viewed as computing a distributed representation of the input. Several "encoder" experiments have shown that when the representation space is small it can be fully used. But computing with such a representation requires completely dependable nodes. In the case where the hidden nodes are noisy and unreliable, we find that error correcting schemes emerge simply by using noisy units during training; random errors in(cid:173) jected during backpropagation result in spreading representations apart. Average and minimum distances increase with misfire probability, as predicted by coding-theoretic considerations. Furthennore, the effect of this noise is to protect the machine against permanent node failure, thereby potentially extending the useful lifetime of the machine.
****
Unsmearing Visual Motion: Development of Long-Range Horizontal Intrinsic Connections
Kevin Martin, Jonathan Marshall
Human VlSlon systems integrate information nonlocally, across long spatial ranges. For example, a moving stimulus appears smeared when viewed briefly (30 ms), yet sharp when viewed for a longer exposure (100 ms) (Burr, 1980). This suggests that visual systems combine information along a trajectory that matches the motion of the stimulus. Our self-organizing neural network model shows how developmental exposure to moving stimuli can direct the formation of horizontal trajectory-specific motion integration pathways that unsmear representations of moving stimuli. These results account for Burr's data and can potentially also model ot.her phenomena, such as visual inertia.
****
Remote Sensing Image Analysis via a Texture Classification Neural Network
Hayit K. Greenspan, Rodney Goodman
In this work we apply a texture classification network to remote sensing im(cid:173) age analysis. The goal is to extract the characteristics of the area depicted in the input image, thus achieving a segmented map of the region. We have recently proposed a combined neural network and rule-based framework for texture recognition. The framework uses unsupervised and supervised learning, and provides probability estimates for the output classes. We describe the texture classification network and extend it to demonstrate its application to the Landsat and Aerial image analysis domain .
****
Filter Selection Model for Generating Visual Motion Signals
Steven Nowlan, Terrence J. Sejnowski
Neurons in area MT of primate visual cortex encode the velocity of moving objects. We present a model of how MT cells aggregate responses from VI to form such a velocity representation. Two different sets of units, with local receptive fields, receive inputs from motion energy filters. One set of units forms estimates of local motion, while the second set computes the utility of these estimates. Outputs from this second set of units "gate" the outputs from the first set through a gain control mechanism. This active process for selecting only a subset of local motion responses to integrate into more global responses distinguishes our model from previous models of velocity estimation. The model yields accurate velocity estimates in synthetic images containing multiple moving targets of varying size, luminance, and spatial frequency profile and deals well with a number of transparency phenomena.
****
Non-Linear Dimensionality Reduction
David DeMers, Garrison Cottrell
A method for creating a non-linear encoder-decoder for multidimensional data with compact representations is presented. The commonly used technique of autoassociation is extended to allow non-linear representations, and an objec(cid:173) tive function which penalizes activations of individual hidden units is shown to result in minimum dimensional encodings with respect to all

owable error in reconstruction.
****
Feudal Reinforcement Learning
Peter Dayan, Geoffrey E. Hinton
One way to speed up reinforcement learning is to enable learning to happen simu
ltaneously at multiple resolutions in space and time. This paper shows how to c
reate a Q-Iearning managerial hierarchy in which high level managers learn how
to set tasks to their sub(cid:173) managers who, in turn, learn how to satisfy t
hem. Sub-managers need not initially understand their managers' commands. T
hey simply learn to maximise their reinforcement in the context of the current
command. We illustrate the system using a simple maze task .. As the system
learns how to get around, satisfying commands at the multiple levels, it expl
ores more efficiently than standard, flat, Q-Iearning and builds a more compre
hensive map.
****
A Neural Network that Learns to Interpret Myocardial Planar Thallium Scintigrams
Charles Rosenberg, Jacob Erel, Henri Atlan
The planar thallium-201 myocardial perfusion scintigram is a widely used diagno
stic technique for detecting and estimating the risk of coronary artery disease
. Neural networks learned to interpret 100 thallium scinti(cid:173) grams as det
ermined by individual expert ratings. Standard error back(cid:173) propagation w
as compared to standard LMS, and LMS combined with one layer of RBF units. Usin
g the "leave-one-out" method, generaliza(cid:173) tion was tested on all 100 cas
es. Training time was determined automati(cid:173) cally from cross-validation p
erfonnance. Best perfonnance was attained by the RBF/LMS network with three hid
den units per view and compares favorably with human experts.
****
Learning Cellular Automaton Dynamics with Neural Networks
N. Wulff, J A. Hertz
We have trained networks of E - II units with short-range connec(cid:173) tions
to simulate simple cellular automata that exhibit complex or chaotic behaviour.
Three levels of learning are possible (in decreas(cid:173) ing order of difficu
lty): learning the underlying automaton rule, learning asymptotic dynamical beh
aviour, and learning to extrap(cid:173) olate the training history. The levels o
f learning achieved with and without weight sharing for different automata prov
ide new insight into their dynamics.
****
Probability Estimation from a Database Using a Gibbs Energy Model
John Miller, Rodney Goodman
We present an algorithm for creating a neural network which pro(cid:173) duces a
ccurate probability estimates as outputs. The network im(cid:173) plements a Gib
bs probability distribution model of the training database. This model is creat
ed by a new transformation relating the joint probabilities of attributes in th
e database to the weights (Gibbs potentials) of the distributed network model.
The theory of this transformation is presented together with experimental re(ci
d:173) sults. One advantage of this approach is the network weights are prescri
bed without iterative gradient descent. Used as a classifier the network tied o
r outperformed published results on a variety of databases.
****
Word Space
Hinrich Schütze
Representations for semantic information about words are neces(cid:173) sa
ry for many applications of neural networks in natural language processing.
This paper describes an efficient, corpus-based method for inducing distri
buted semantic representations for a large num(cid:173) ber of words (50,0
00) from lexical coccurrence statistics by means of a large-scale linear
regression. The representations are success(cid:173) fully applied to word
sense disambiguation using a nearest neighbor method .
****
An Analog VLSI Chip for Radial Basis Functions

Janeen Anderson, John Platt, David B. Kirk
We have designed, fabricated, and tested an analog VLSI chip which computes radial basis functions in parallel. We have de(cid:173) veloped a synapse circuit that approximates a quadratic function. We aggregate these circuits to form radial basis functions. These radial basis functions are then averaged together using a follower aggregator.
****
Predicting Complex Behavior in Sparse Asymmetric Networks
Ali Minai, William Levy
Recurrent networks of threshold elements have been studied inten(cid:173) sively as associative memories and pattern-recognition devices. While most research has concentrated on fully-connected symmetric net(cid:173) works. which relax to stable fixed points. asymmetric networks show richer dynamical behavior. and can be used as sequence generators or flexible pattern-recognition devices. In this paper. we approach the problem of predicting the complex global behavior of a class of ran(cid:173) dom asymmetric networks in terms of network parameters. These net(cid:173) works can show fixed-point. cyclical or effectively aperiodic behavior. depending on parameter values. and our approach can be used to set parameters. as necessary. to obtain a desired complexity of dynamics. The approach also provides qualitative insight into why the system behaves as it does and suggests possible applications.
****
Destabilization and Route to Chaos in Neural Networks with Random Connectivity
Bernard Doyon, Bruno Cessac, Mathias Quoy, Manuel Samuelides
The occurence of chaos in recurrent neural networks is supposed to depend on the architecture and on the synaptic coupling strength. It is studied here for a randomly diluted architecture. By normalizing the variance of synaptic weights, we produce a bifurcation parameter, dependent on this variance and on the slope of the transfer function but independent of the connectivity, that allows a sustained activity and the occurence of chaos when reaching a critical value. Even for weak connectivity and small size, we find numerical results in accordance with the theoretical ones previously established for fully connected infinite sized networks. Moreover the route towards chaos is numerically checked to be a quasi-periodic one, whatever the type of the first bifurcation is (Hopf bifurcation, pitchfork or flip).
****
Recognition-based Segmentation of On-Line Hand-printed Words
M. Schenkel, H. Weissman, I. Guyon, C. Nohl, D. Henderson
This paper reports on the performance of two methods for recognition-based segmentation of strings of on-line hand-printed capital Latin characters. The input strings consist of a time(cid:173) ordered sequence of X-Y coordinates, punctuated by pen-lifts. The methods were designed to work in "run-on mode" where there is no constraint on the spacing between characters. While both methods use a neural network recognition engine and a graph-algorithmic post-processor, their approaches to segmentation are quite differ(cid:173) ent. The first method, which we call IN SEC (for input segmen(cid:173) tation), uses a combination of heuristics to identify particular pen(cid:173) lifts as tentative segmentation points. The second method, which we call OUTSEC (for output segmentation), relies on the empiri(cid:173) cally trained recognition engine for both recognizing characters and identifying relevant segmentation points.
****
Some Estimates of Necessary Number of Connections and Hidden Units for Feed-Forward Networks
Adam Kowalczyk
The feed-forward networks with fixed hidden units (FllU-networks) are compared against the category of remaining feed-forward net(cid:173) works with variable hidden units (VHU-networks). Two broad classes of tasks on a finite domain X C R n are considered: ap(cid:173) proximation of every function from an open subset of functions on X and representation of every dichotomy of X. For the first t

ask it is found that both network categories require the same minimal number of synaptic weights. For the second task and X in gen(cid:173) eral position it is shown that VHU-networks with threshold logic hidden units can have approximately lin times fewer hidden units than any FHU-network must have.
****

The Power of Approximating: a Comparison of Activation Functions
Bhaskar DasGupta, Georg Schnitger
We compare activation functions in terms of the approximation power of their feedforward nets. We consider the case of analog as well as boolean input.
****

Neural Network Model Selection Using Asymptotic Jackknife Estimator and Cross-Validation Method
Yong Liu
Two theorems and a lemma are presented about the use of jackknife es(cid:173) timator and the cross-validation method for model selection. Theorem 1 gives the asymptotic form for the jackknife estimator. Combined with the model selection criterion, this asymptotic form can be used to obtain the fit of a model. The model selection criterion we used is the negative of the average predictive likehood, the choice of which is based on the idea of the cross-validation method. Lemma 1 provides a formula for further explo(cid:173) ration of the asymptotics of the model selection criterion. Theorem 2 gives an asymptotic form of the model selection criterion for the regression case, when the parameters optimization criterion has a penalty term. Theorem 2 also proves the asymptotic equivalence of Moody's model selection cri(cid:173) terion (Moody, 1992) and the cross-validation method, when the distance measure between response y and regression function takes the form of a squared difference.
****

Context-Dependent Multiple Distribution Phonetic Modeling with MLPs
Michael Cohen, Horacio Franco, Nelson Morgan, David Rumelhart, Victor Abrash
A number of hybrid multilayer perceptron (MLP)/hidden Markov model (HMM:) speech recognition systems have been developed in recent years (Morgan and Bourlard. 1990). In this paper. we present a new MLP architecture and training algorithm which allows the modeling of context-dependent phonetic classes in a hybrid MLP/HMM: framework. The new training procedure smooths MLPs trained at different degrees of context dependence in order to obtain a robust estimate of the cootext-dependent probabilities. Tests with the DARPA Resomce Management database have shown substantial advantages of the context-dependent MLPs over earlier cootext (cid:173) independent MLPs. and have shown substantial advantages of this hybrid approach over a pure HMM approach.
****

On the Use of Evidence in Neural Networks
David Wolpert
The Bayesian "evidence" approximation has recently been employed to determine the noise and weight-penalty terms used in back-propagation. This paper shows that for neural nets it is far easier to use the exact result than it is to use the evidence approximation. Moreover, unlike the evi(cid:173) dence approximation, the exact result neither has to be re-calculated for every new data set, nor requires the running of computer code (the exact result is closed form). In addition, it turns out that the evidence proce(cid:173) dure's MAP estimate for neural nets is, in toto, approximation error. An(cid:173) other advantage of the exact analysis is that it does not lead one to incor(cid:173) rect intuition, like the claim that using evidence one can "evaluate differ(cid:173) ent priors in light of the data". This paper also discusses sufficiency conditions for the evidence approximation to hold, why it can sometimes give "reasonable" results, etc.
****

Automatic Capacity Tuning of Very Large VC-Dimension Classifiers
I. Guyon, B. Boser, V. Vapnik

Large VC-dimension classifiers can learn difficult tasks, but are usually impractical because they generalize well only if they are trained with huge quantities of data. In this paper we show that even high-order polynomial classifiers in high dimensional spaces can be trained with a small amount of training data and yet generalize better than classifiers with a smaller VC-dimension. This is achieved with a maximum margin algorithm (the Generalized Portrait). The technique is applicable to a wide variety of classifiers, including Perceptrons, polynomial classifiers (sigma-pi unit net(cid:173) works) and Radial Basis Functions. The effective number of parameters is adjusted automatically by the training algorithm to match the complexity of the problem. It is shown to equal the number of those training patterns which are closest patterns to the decision boundary (supporting patterns). Bounds on the generalization error and the speed of convergence of the al(cid:173) gorithm are given. Experimental results on handwritten digit recognition demonstrate good generalization compared to other algorithms.
****
Interposing an ontogenetic model between Genetic Algorithms and Neural Networks
Richard Belew
The relationships between learning, development and evolution in Nature is taken seriously, to suggest a model of the developmental process whereby the genotypes manipulated by the Genetic Algo(cid:173) rithm (GA) might be expressed to form phenotypic neural networks (NNet) that then go on to learn. ONTOL is a grammar for gener(cid:173) ating polynomial NN ets for time-series prediction. Genomes corre(cid:173) spond to an ordered sequence of ONTOL productions and define a grammar that is expressed to generate a NNet. The NNet's weights are then modified by learning, and the individual's prediction error is used to determine GA fitness. A new gene doubling operator appears critical to the formation of new genetic alternatives in the preliminary but encouraging results presented.
****
Bayesian Learning via Stochastic Dynamics
Radford Neal
The attempt to find a single "optimal" weight vector in conven(cid:173) tional network training can lead to overfitting and poor generaliza(cid:173) tion. Bayesian methods avoid this, without the need for a valida(cid:173) tion set, by averaging the outputs of many networks with weights sampled from the posterior distribution given the training data. This sample can be obtained by simulating a stochastic dynamical system that has the posterior as its stationary distribution.
****
Biologically Plausible Local Learning Rules for the Adaptation of the Vestibulo-Ocular Reflex
Olivier Coenen, Terrence J. Sejnowski, Stephen Lisberger
The vestibulo-ocular reflex (VOR) is a compensatory eye movement that stabilizes images on the retina during head turns. Its magnitude, or gain, can be modified by visual experience during head movements. Possible learning mechanisms for this adaptation have been explored in a model of the oculomotor system based on anatomical and physiological con(cid:173) straints. The local correlational learning rules in our model reproduce the adaptation and behavior of the VOR under certain parameter conditions. From these conditions, predictions for the time course of adaptation at the learning sites are made.
****
Combining Neural and Symbolic Learning to Revise Probabilistic Rule Bases
J. Mahoney, Raymond Mooney
a system for revising probabilis(cid:173)
****
Transient Signal Detection with Neural Networks: The Search for the Desired Signal
José Príncipe, Abir Zahalka
Matched filtering has been one of the most powerful techniques employed for transient detection. Here we will show that a dynamic neural network out

performs the conventional approach. When the artificial neural network (ANN) is trained with supervised learning schemes there is a need to supply the desired signal for all time, although we are only interested in detecting the transient. In this paper we also show the effects on the detection agreement of different strategies to construct the desired signal. The extension of the Bayes decision rule (011 desired signal), optimal in static classification, performs worse than desired signals constructed by random noise or prediction during the background.
****

Information Theoretic Analysis of Connection Structure from Spike Trains

Satoru Shiono, Satoshi Yamada, Michio Nakashima, Kenji Matsumoto

We have attempted to use information theoretic quantities for ana(cid:173) lyzing neuronal connection structure from spike trains. Two point mu tual information and its maximum value, channel capacity, be(cid:173) tween a pair of neurons were found to be useful for sensitive de(cid:173) tection of crosscorrelation and for estimation of synaptic strength, respectively. Three point mutual information among three neurons could give their interconnection structure. Therefore, our informa(cid:173) tion theoretic analysis was shown to be a very powerful technique for deducing neuronal connection structure. Some concrete exam(cid:173) ples of its application to simulated spike trains are presented.
****