Learning to Predict Visibility and Invisibility from Occlusion Events
Jonathan Marshall, Richard Alley, Robert Hubbard
Visual occlusion events constitute a major source of depth information. This paper presents a self-organizing neural network that learns to detect, represent, and predict the visibility and invisibility relationships that arise during occlusion events, after a period of exposure to motion sequences containing occlusion and disocclusion events. The network develops two parallel opponent channels or "chains" of lateral excitatory connections for every resolvable motion trajectory. One channel, the "On" chain or "visible" chain, is activated when a moving stimulus is visible. The other channel, the "Off" chain or "invisible" chain, carries a persistent, amodal representation that predicts the motion of a formerly visible stimulus that becomes invisible due to occlusion. The learning rule uses disinhibition from the On chain to trigger learning in the Off chain. The On and Off chain neurons can learn separate associations with object depth or(cid:173) dering. The results are closely related to the recent discovery (Assad & Maunsell, 1995) of neurons in macaque monkey posterior parietal cortex that respond selectively to inferred motion of invisible stimuli.
****

Onset-based Sound Segmentation
Leslie Smith
A technique for segmenting sounds using processing based on mam(cid:173) malian early auditory processing is presented. The technique is based on features in sound which neuron spike recording suggests are detected in the cochlear nucleus. The sound signal is band(cid:173) passed and each signal processed to enhance onsets and offsets. The onset and offset signals are compressed, then clustered both in time and across frequency channels using a network of integrate(cid:173) and-fire neurons. Onsets and offsets are signalled by spikes, and the timing of these spikes used to segment the sound.
****

Beating a Defender in Robotic Soccer: Memory-Based Learning of a Continuous Function
Peter Stone, Manuela Veloso
Learning how to adjust to an opponent's position is critical to the success of having intelligent agents collaborating towards the achievement of specific tasks in unfriendly environments. This pa(cid:173) per describes our work on a Memory-based technique for to choose an action based on a continuous-valued state attribute indicating the position of an opponent. We investigate the question of how an agent performs in nondeterministic variations of the training situ(cid:173) ations. Our experiments indicate that when the random variations fall within some bound of the initial training, the agent performs better with some initial training rather than from a tabula-rasa.
****

A Neural Network Autoassociator for Induction Motor Failure Prediction
Thomas Petsche, Angelo Marcantonio, Christian Darken, Stephen Hanson, Gary Kuhn, N. Santoso
We present results on the use of neural network based autoassociators which act as novelty or anomaly detectors to detect imminent motor failures. The autoassociator is trained to reconstruct spectra obtained from the healthy motor. In laboratory tests, we have demonstrated that the trained autoassociator has a small reconstruction error on measurements recorded from healthy motors but a larger error on those recorded from a motor with a fault. We have designed and built a motor monitoring system using an autoassociator for anomaly detection and are in the process of testing the system at three industrial and commercial sites.
****

The Gamma MLP for Speech Phoneme Recognition
Steve Lawrence, Ah Tsoi, Andrew Back
We define a Gamma multi-layer perceptron (MLP) as an MLP with the usua

l synaptic weights replaced by gamma filters (as pro(cid:173) posed by de Vries and Principe (de Vries and Principe, 1992)) and associated gain terms thr oughout all layers. We derive gradient descent update equations and apply the model to the recognition of speech phonemes. We find that both t he inclusion of gamma filters in all layers, and the inclusion of syna ptic gains, improves the performance of the Gamma MLP. We compare the Gamma MLP with TDNN, Back-Tsoi FIR MLP, and Back-Tsoi I1R MLP architectu res, and a local approximation scheme. We find that the Gamma MLP results in an substantial reduction in error rates.

****

Laterally Interconnected Self-Organizing Maps in Hand-Written Digit Recognition
Yoonsuck Choe, Joseph Sirosh, Risto Miikkulainen

An application of laterally interconnected self-organizing maps (LISSOM) to handwritten digit recognition is presented. The lat(cid:173) eral connec tions learn the correlations of activity between units on the map. The resu lting excitatory connections focus the activity into local patches and the inhibitory connections decorrelate redun(cid:173) dant activity on the map. Th e map thus forms internal representa(cid:173) tions that are easy to recognize w ith e.g. a perceptron network. The recognition rate on a subset of NIST datab ase 3 is 4.0% higher with LISSOM than with a regular Self-Organizing Map (SOM) as the front end, and 15.8% higher than recognition of raw input bit maps directly. These results form a promising starting point for building pa ttern recognition systems with a LISSOM map as a front end.

****

Improved Silicon Cochlea using Compatible Lateral Bipolar Transistors
André van Schaik, Eric Fragnière, Eric Vittoz

Analog electronic cochlear models need exponentially scaled filters. CMO S Compatible Lateral Bipolar Transistors (CLBTs) can create exponentiall y scaled currents when biased using a resistive line with a voltage difference between both ends of the line. Since these CLBTs are independent of t he CMOS threshold voltage, current sources implemented with CLBTs are much better matched than current sources created with MOS transistors operated in weak inversion. Measurements from integrated test chips are shown to verify the improved matching.

****

On the Computational Power of Noisy Spiking Neurons
Wolfgang Maass

It has remained unknown whether one can in principle carry out reliabl e digital computations with networks of biologically realistic models for ne urons. This article presents rigorous constructions for simulating in re al-time arbitrary given boolean circuits and fi(cid:173) nite automata with ar bitrarily high reliability by networks of noisy spiking neurons. In addition we show that with the help of "shunting inhibition" even networks of very unreliable spiking neurons can simulate in real-time any McCulloc h-Pitts neuron (or "threshold gate"), and therefore any multilayer perc eptron (or "threshold circuit") in a reliable manner. These constructio ns provide a possible explana(cid:173) tion for the fact that biological neural systems can carry out quite complex computations within 100 msec. It turns out that the assumption that these constructions require about the s hape of the EPSP's and the behaviour of the noise are surprisingly weak.

****

Parallel Optimization of Motion Controllers via Policy Iteration
Jefferson Coelho, R. Sitaraman, Roderic Grupen

This paper describes a policy iteration algorithm for optimizing the performan ce of a harmonic function-based controller with respect to a user-defined index. Value functions are represented as poten(cid:173) tial distribution s over the problem domain, being control policies represented as gradi ent fields over the same domain. All interme(cid:173) diate policies are intrinsically safe, i.e. collisions are not promoted during the adaptati on process. The algorithm has efficient imple(cid:173) mentation in parallel

SIMD architectures. One potential applica(cid:173) tion - travel distance minimization - illustrates its usefulness.

****

Model Matching and SFMD Computation

Steven Rehfuss, Dan Hammerstrom

In systems that process sensory data there is frequently a model matchin g stage where class hypotheses are combined to recognize a complex entity. We introduce a new model of parallelism, the Single Function Multiple Data (SFM D) model, appropriate to this stage. SFMD functionality can be added with small hardware expense to certain existing SIMD architectures, and as an incremental addition to the programming model. Adding SFMD to an SIMD machine will n ot only allow faster model matching, but also increase its flexibility as a general purpose machine and its scope in performing the initial stages o f sensory processing.

****

Learning with ensembles: How overfitting can be useful

Peter Sollich, Anders Krogh

We study the characteristics of learning with ensembles. Solving exactly the simple model of an ensemble of linear students, we find surprising ly rich behaviour. For learning in large ensembles, it is advantageous to use under-regularized students, which actu(cid:173) ally over-fit th e training data. Globally optimal performance can be obtained by choosing the training set sizes of the students ap(cid:173) propriately. For smaller ensembles, optimization of the ensemble weights can yield significant impro vements in ensemble generaliza(cid:173) tion performance, in particular if the individual students are sub(cid:173) ject to noise in the training process. Choosing students with a wide range of regularization parameters makes this im provement robust against changes in the unknown level of noise in the training data.

****

Analog VLSI Processor Implementing the Continuous Wavelet Transform

R. Edwards, Gert Cauwenberghs

We present an integrated analog processor for real-time wavelet decom(cid:173) p osition and reconstruction of continuous temporal signals covering the audio fr equency range. The processor performs complex harmonic modu(cid:173) lation and Gaussian lowpass filtering in 16 parallel channels, each clocked at a different rate, producing a multiresolution mapping on a logarithmic frequency scale. Ou r implementation uses mixed-mode analog and dig(cid:173) ital circuits, oversamp ling techniques, and switched-capacitor filters to achieve a wide linear dynami c range while maintaining compact circuit size and low power consumption. We in clude experimental results on the processor and characterize its components sep arately from measurements on a single-channel test chip.

****

A Novel Channel Selection System in Cochlear Implants Using Artificial Neural Ne twork

Marwan Jabri, Raymond Wang

State-of-the-art speech processors in cochlear implants perform channel s election using a spectral maxima strategy. This strategy can lead to conf usions when high frequency features are needed to discriminate between s ounds. We present in this paper a novel channel selection strategy based u pon pattern recognition which al(cid:173) lows "smart" channel selections to be made. The proposed strategy is implemented using multi-layer perceptrons trained on a multi(cid:173) speaker labelled speech database. The input to t he network are the energy coefficients of N energy channels. The output of th e system are the indices of the M selected channels. We compare the per formance of our proposed system to that of spectral maxima strategy, and show that our strategy can produce significantly better results.

****

Rapid Quality Estimation of Neural Network Input Representations

Kevin Cherkauer, Jude Shavlik

The choice of an input representation for a neural network can have a profound impact on its accuracy in classifying novel instances. However, neural networks are typically computationally expensive to train, making it difficult to test large numbers of alternative representations. This paper introduces fast quality measures for neural network representations, allowing one to quickly and ac(cid:173) curately estimate which of a collection of possible representations for a problem is the best. We show that our measures for ranking representations are more accurate than a previously published mea(cid:173) sure, based on experiments with three difficult, real-world pattern recognition problems.

****

Generating Accurate and Diverse Members of a Neural-Network Ensemble

David Opitz, Jude Shavlik

Neural-network ensembles have been shown to be very accurate classification techniques. Previous work has shown that an effec(cid:173) tive ensemble should consist of networks that are not only highly correct, but ones that make their errors on different parts of the input space as well. Most existing techniques, however, only in(cid:173) directly address the problem of creating such a set of networks. In this paper we present a technique called ADDEMUP that uses genetic algorithms to directly search for an accurate and diverse set of trained networks. ADDEMUP works by first creating an ini(cid:173) tial population, then uses genetic operators to continually create new networks, keeping the set of networks that are as accurate as possible while disagreeing with each other as much as possible. Ex(cid:173) periments on three DNA problems show that ADDEMUP is able to generate a set of trained networks that is more accurate than sev(cid:173) eral existing approaches. Experiments also show that ADDEMUP is able to effectively incorporate prior knowledge, if available, to improve the quality of its ensemble.

****

A Model of Spatial Representations in Parietal Cortex Explains Hemineglect

Alexandre Pouget, Terrence J. Sejnowski

We have recently developed a theory of spatial representations in which the position of an object is not encoded in a particular frame of reference but, instead, involves neurons computing basis func(cid:173) tions of their sensory inputs. This type of representation is able to perform nonlinear sensorimotor transformations and is consis(cid:173) tent with the response properties of parietal neurons. We now ask whether the same theory could account for the behavior of human patients with parietal lesions. These lesions induce a deficit known as hemineglect that is characterized by a lack of reaction to stimuli located in the hemispace contralateral to the lesion. A simulated lesion in a basis function representation was found to replicate three of the most important aspects of hemineglect: i) The models failed to cross the leftmost lines in line cancellation experiments, ii) the deficit affected multiple frames of reference and, iii) it could be object centered. These results strongly support the basis function hypothesis for spatial representations and provide a computational theory of hemineglect at the single cell level.

****

Temporal Difference Learning in Continuous Time and Space

Kenji Doya

A continuous-time, continuous-state version of the temporal differ(cid:173) ence (TD) algorithm is derived in order to facilitate the application of reinforcement learning to real-world control tasks and neurobi(cid:173) ological modeling. An optimal nonlinear feedback control law was also derived using the derivatives of the value function. The per(cid:173) formance of the algorithms was tested in a task of swinging up a pendulum with limited torque. Both the "critic" that specifies the paths to the upright position and the "actor" that works as a non(cid:173) linear feedback controller were successfully implemented by radial basis function (RBF) networks.

****

Tempering Backpropagation Networks: Not All Weights are Created Equal
Nicol Schraudolph, Terrence J. Sejnowski
Terrence J. Sejnowski
****

Improving Policies without Measuring Merits
Peter Dayan, Satinder Singh
Performing policy iteration in dynamic programming should only require knowledge of relative rather than absolute measures of the utility of actions (Werbos, 1991) - what Baird (1993) calls the ad(cid:173)vantages of actions at states. Nevertheless, most existing methods in dynamic programming (including Baird's) compute some form of absolute utility function . For smooth problems, advantages satisfy two differential consistency conditions (including the requirement that they be free of curl), and we show that enforcing these can lead to appropriate policy improvement solely in terms of advantages.
****

Selective Attention for Handwritten Digit Recognition
Ethem Alpaydin
Completely parallel object recognition is NP-complete. Achieving a recognizer with feasible complexity requires a compromise be(cid:173)tween parallel and sequential processing where a system selectively focuses on parts of a given image, one after another. Successive fixations are generated to sample the image and these samples are processed and abstracted to generate a temporal context in which results are integrated over time. A computational model based on a partially recurrent feedforward network is proposed and made cred(cid:173)ible by testing on the real-world problem of recognition of hand(cid:173)written digits with encouraging results.
****

Learning Model Bias
Jonathan Baxter
In this paper the problem of learning appropriate domain-specific bias is addressed. It is shown that this can be achieved by learning many related tasks from the same domain, and a theorem is given bounding the number tasks that must be learnt. A corollary of the theorem is that if the tasks are known to possess a common inter(cid:173)nal representation or preprocessing then the number of examples required per task for good generalisation when learning n tasks si(cid:173)multaneously scales like O(a + ~), where O(a) is a bound on the minimum number of examples requred to learn a single task, and O(a + b) is a bound on the number of examples required to learn each task independently. An experiment providing strong qualita(cid:173)tive support for the theoretical results is reported.
****

Estimating the Bayes Risk from Sample Data
Robert Snapp, Tong Xu
A new nearest-neighbor method is described for estimating the Bayes risk of a multiclass pattern claSSification problem from sample data (e.g., a classified training set). Although it is assumed that the classification prob(cid:173)lem can be accurately described by sufficiently smooth class-conditional distributions, neither these distributions, nor the corresponding prior prob(cid:173)abilities of the classes are required. Thus this method can be applied to practical problems where the underlying probabilities are not known. This method is illustrated using two different pattern recognition problems.
****

A Model of Auditory Streaming
Susan McCabe, Michael Denham
An essential feature of intelligent sensory processing is the ability to focus on the part of the signal of interest against a background of distracting signals, and to be able to direct this focus at will. In this paper the problem of auditory scene segmentation is considered and a model of the early stages of the process is proposed. The behaviour of the model is shown to be in agreement wi

th a number of well known psychophysical results. The principal contribution of this model lies in demonstrating how streaming might result from interactions between the tonotopic patterns of activity of input signals and traces of previous activity which feedback and influence the way in which subsequent signals are processed.

****

Context-Dependent Classes in a Hybrid Recurrent Network-HMM Speech Recognition System

Dan Kershaw, Anthony Robinson, Mike Hochberg

A method for incorporating context-dependent phone classes in a connectionist-HMM hybrid speech recognition system is intro(cid:173) duced. A modular approach is adopted, where single-layer networks discriminate between different context classes given the phone class and the acoustic data. The context networks are combined with a context-independent (CI) network to generate context-dependent (CD) phone probability estimates. Experiments show an average reduction in word error rate of 16% and 13% from the CI system on ARPA 5,000 word and SQALE 20,000 word tasks respectively. Due to improved modelling, the decoding speed of the CD system is more than twice as fast as the CI system.

****

Exploiting Tractable Substructures in Intractable Networks

Lawrence Saul, Michael Jordan

We develop a refined mean field approximation for inference and learning in probabilistic neural networks. Our mean field theory, unlike most, does not assume that the units behave as independent degrees of freedom; instead, it exploits in a principled way the existence of large substructures that are computationally tractable. To illustrate the advantages of this framework, we show how to incorporate weak higher order interactions into a first-order hidden Markov model, treating the corrections (but not the first order structure) within mean field theory.

****

Statistical Theory of Overtraining - Is Cross-Validation Asymptotically Effective?

Shun-ichi Amari, Noboru Murata, Klaus-Robert Müller, Michael Finke, Howard Yang

A statistical theory for overtraining is proposed. The analysis treats realizable stochastic neural networks, trained with Kullback(cid:173) Leibler loss in the asymptotic case. It is shown that the asymptotic gain in the generalization error is small if we perform early stop(cid:173) ping, even if we have access to the optimal stopping time. Consider(cid:173) ing cross-validation stopping we answer the question: In what ratio the examples should be divided into training and testing sets in or(cid:173) der to obtain the optimum performance. In the non-asymptotic region cross-validated early stopping always decreases the general(cid:173) ization error. Our large scale simulations done on a CM5 are in nice agreement with our analytical findings.

****

Primitive Manipulation Learning with Connectionism

Yoky Matsuoka

Infants' manipulative exploratory behavior within the environment is a vehicle of cognitive stimulation[McCall 1974]. During this time, infants practice and perfect sensorimotor patterns that become be(cid:173) havioral modules which will be seriated and imbedded in more com(cid:173) plex actions. This paper explores the development of such primitive learning systems using an embodied light-weight hand which will be used for a humanoid being developed at the MIT Artificial In(cid:173) telligence Laboratory[Brooks and Stein 1993]. Primitive grasping procedures are learned from sensory inputs using a connectionist reinforcement algorithm while two submodules preprocess sensory data to recognize the hardness of objects and detect shear using competitive learning and back-propagation algorithm strategies, respectively. This system is not only consistent and quick dur(cid:173) ing the initial learning stage, but also adaptable to new situations after training is comp

leted.
****

Silicon Models for Auditory Scene Analysis
John Lazzaro, John Wawrzynek
We are developing special-purpose, low-power analog-to-digital converters for speech and music applications, that feature analog circuit models of biological audition to process the audio signal before conversion. This paper describes our most recent converter design, and a working system that uses several copies ofthe chip to compute multiple representations of sound from an analog input. This multi-representation system demonstrates the plausibility of inexpensively implementing an auditory scene analysis approach to sound processing.
****

Human Face Detection in Visual Scenes
Henry Rowley, Shumeet Baluja, Takeo Kanade
We present a neural network-based face detection system. A retinally connected neural network examines small windows of an image, and decides whether each window contains a face. The system arbitrates between multiple networks to improve performance over a single network. We use a bootstrap algorithm for training, which adds false detections into the training set as training progresses. This eliminates the difficult task of manually selecting non-face training examples, which must be chosen to span the entire space of non-face images. Comparisons with another state-of-the-art face detection system are presented; our system has better performance in terms of detection and false-positive rates.
****

SPERT-II: A Vector Microprocessor System and its Application to Large Problems in Backpropagation Training
John Wawrzynek, Krste Asanovic, Brian Kingsbury, James Beck, David Johnson, Nelson Morgan
We report on our development of a high-performance system for neural network and other signal processing applications. We have designed and implemented a vector microprocessor and pack(cid:173)aged it as an attached processor for a conventional workstation. We present performance comparisons with commercial worksta(cid:173)tions on neural network backpropagation training. The SPERT-II system demonstrates significant speedups over extensively hand(cid:173)optimization code running on the workstations.
****

Pruning with generalization based weight saliencies: λOBD, λOBS
Morten Pedersen, Lars Hansen, Jan Larsen
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.
****

Stochastic Hillclimbing as a Baseline Method for Evaluating Genetic Algorithms
Ari Juels, Martin Wattenberg
We investigate the effectiveness of stochastic hillclimbing as a baseline for evaluating the performance of genetic algorithms (GAs) as combinato(cid:173)rial function optimizers. In particular, we address two problems to which GAs have been applied in the literature: Koza's ll-multiplexer problem and the jobshop problem. We demonstrate that simple stochastic hill(cid:173)climbing methods are able to achieve results comparable or superior to those obtained by the GAs designed to address these two problems. We further illustrate, in the case of the jobshop problem, how insights ob(cid:173)tained in the formulation of a stochastic hillclimbing algorithm can lead to improvements in the encoding used by a GA.
****

Using the Future to "Sort Out" the Present: Rankprop and Multitask Learning for Medical Risk Evaluation

Rich Caruana, Shumeet Baluja, Tom Mitchell
A patient visits the doctor; the doctor reviews the patient's history, asks qu estions, makes basic measurements (blood pressure, .. .), and prescribes tes ts or treatment . The prescribed course of action is based on an asses sment of patient risk-patients at higher risk are given more and faster at tention. It is also sequential- it is too expensive to immediately order all tests which might later be of value. This paper presents two m ethods that together improve the accuracy of backprop nets on a pneum onia risk assessment problem by 10-50%. Rankprop improves on backpropagati on with sum of squares error in ranking patients by risk. Multitask learning takes advantage of future lab tests available in the training set, but not available in practice when predictions must be made. Both methods are b roadly applicable.
****

Exponentially many local minima for single neurons
Peter Auer, Mark Herbster, Manfred K. K. Warmuth
We show that for a single neuron with the logistic function as the transfer fun ction the number of local minima of the error function based on the square los s can grow exponentially in the dimension.
****

An Information-theoretic Learning Algorithm for Neural Network Classification
David J. Miller, Ajit Rao, Kenneth Rose, Allen Gersho
A new learning algorithm is developed for the design of statistical class ifiers minimizing the rate of misclassification. The method, which is b ased on ideas from information theory and analogies to statistical phys ics, assigns data to classes in probability. The dis(cid:173) tributions are chosen to minimize the expected classification error while simultane ously enforcing the classifier's structure and a level of "randomness" measur ed by Shannon's entropy. Achievement of the classifier structure is quantif ied by an associated cost. The con(cid:173) strained optimization problem is equivalent to the minimization of a Helmholtz free energy, and the resul ting optimization method is a basic extension of the deterministic anne aling algorithm that explicitly enforces structural constraints on assign ments while re(cid:173) ducing the entropy and expected cost with tempe rature. In the limit of low temperature, the error rate is minimized directly and a hard classifier with the requisite structure is obtained. This learn(c id:173) ing algorithm can be used to design a variety of classifier structures. The approach is compared with standard methods for radial basis function de sign and is demonstrated to substantially outperform other design metho ds on several benchmark examples, while of(cid:173) ten retaining design c omplexity comparable to, or only moderately greater than that of strict descen t-based methods.
****

Improving Elevator Performance Using Reinforcement Learning
Robert Crites, Andrew Barto
This paper describes the application of reinforcement learning (RL) to the diff icult real world problem of elevator dispatching. The el(cid:173) evator doma in poses a combination of challenges not seen in most RL research to date. Elevator systems operate in continuous state spaces and in continuous time as discrete event dynamic systems. Their states are not fully observable and they are nonstationary due to changing passenger arrival rates. In addi tion, we use a team of RL agents, each of which is responsible for contr olling one ele(cid:173) vator car. The team receives a global reinforcement signal which appears noisy to each agent due to the effects of the actions of the other agents, the random nature of the arrivals and the incomplete observ ation of the state. In spite of these complications, we show results that in simulation surpass the best of the heuristic elevator control algorithms o f which we are aware. These results demon(cid:173) strate the power of RL on a very large scale stochastic dynamic optimization problem of prac tical utility.

****

Optimal Asset Allocation using Adaptive Dynamic Programming
Ralph Neuneier

In recent years, the interest of investors has shifted to computer(cid:173) ized asset allocation (portfolio management) to exploit the growing dynamics of the capital markets. In this paper, asset allocation is formalized as a Markovian Decision Problem which can be opti(cid:173) mized by applying dynamic programming or reinforcement learning based algorithms. Using an artificial exchange rate, the asset allo(cid:173) cation strategy optimized with reinforcement learning (Q-Learning) is shown to be equivalent to a policy computed by dynamic pro(cid:173) gramming. The approach is then tested on the task to invest liquid capital in the German stock market. Here, neural networks are used as value function approximators. The resulting asset alloca(cid:173) tion strategy is superior to a heuristic benchmark policy. This is a further example which demonstrates the applicability of neural network based reinforcement learning to a problem setting with a high dimensional state space.

****

Recursive Estimation of Dynamic Modular RBF Networks
Visakan Kadirkamanathan, Maha Kadirkamanathan

In this paper, recursive estimation algorithms for dynamic modular networks are developed. The models are based on Gaussian RBF networks and the gating network is considered in two stages: At first, it is simply a time-varying scalar and in the second, it is based on the state, as in the mixture of local experts scheme. The resulting algorithm uses Kalman filter estimation for the model estimation and the gating probability estimation. Both, 'hard' and 'soft' competition based estimation schemes are developed where in the former, the most probable network is adapted and in the latter all networks are adapted by appropriate weighting of the data.

****

A Neural Network Classifier for the I100 OCR Chip
John Platt, Timothy Allen

This paper describes a neural network classifier for the 11000 chip, which optically reads the E13B font characters at the bottom of checks. The first layer of the neural network is a hardware linear classifier which recognizes the characters in this font . A second software neural layer is implemented on an inexpensive microprocessor to clean up the re(cid:173) sults of the first layer. The hardware linear classifier is mathematically specified using constraints and an optimization principle. The weights of the classifier are found using the active set method, similar to Vap(cid:173) nik's separating hyperplane algorithm. In 7.5 minutes of SPARC 2 time, the method solves for 1523 Lagrange mUltipliers, which is equivalent to training on a data set of approximately 128,000 examples. The result(cid:173) ing network performs quite well: when tested on a test set of 1500 real checks, it has a 99.995% character accuracy rate.

****

Modeling Saccadic Targeting in Visual Search
Rajesh Rao, Gregory Zelinsky, Mary Hayhoe, Dana Ballard

Visual cognition depends criticalIy on the ability to make rapid eye movements known as saccades that orient the fovea over targets of interest in a visual scene. Saccades are known to be ballistic: the pattern of muscle activation for foveating a prespecified target location is computed prior to the movement and visual feedback is precluded. Despite these distinctive properties, there has been no general model of the saccadic targeting strategy employed by the human visual system during visual search in natural scenes. This paper proposes a model for saccadic targeting that uses iconic scene representations derived from oriented spatial filters at multiple scales. Visual search proceeds in a coarse-to-fine fashion with the largest scale filter responses being compared first. The model was empirically tested by comparing its perfonnance with actual eye movement data from human subjects in a natural visual search task; preliminary resu

lts indicate substantial agreement between eye movements predicted by the model and those recorded from human subjects.

****

Plasticity of Center-Surround Opponent Receptive Fields in Real and Artificial Neural Systems of Vision

S. Yasui, T. Furukawa, M. Yamada, T. Saito

Despite the phylogenic and structural differences, the visual sys(cid:173) tems of different species, whether vertebrate or invertebrate, share certain functional properties. The center-surround opponent recep(cid:173) tive field (CSRF) mechanism represents one such example. Here, analogous CSRFs are shown to be formed in an artificial neural network which learns to localize contours (edges) of the luminance difference. Furthermore, when the input pattern is corrupted by a background noise, the CSRFs of the hidden units becomes shal(cid:173) lower and broader with decrease of the signal-to-noise ratio (SNR). The same kind of SNR-dependent plasticity is present in the CSRF of real visual neurons; in bipolar cells of the carp retina as is shown here experimentally, as well as in large monopolar cells of the fly compound eye as was described by others. Also, analogous SNR(cid:173) dependent plasticity is shown to be present in the biphasic flash responses (BPFR) of these artificial and biological visual systems . Thus, the spatial (CSRF) and temporal (BPFR) filtering proper(cid:173) ties with which a wide variety of creatures see the world appear to be optimized for detectability of changes in space and time.

****

Active Gesture Recognition using Learned Visual Attention

Trevor Darrell, Alex Pentland

We have developed a foveated gesture recognition system that runs in an uncons trained office environment with an active camera. Us(cid:173) ing vision routines previously implemented for an interactive envi(cid:173) ronment, we determine the spatial location of salient body parts of a user and guide an active camera to obtain images of gestures or expressions. A hidden-state reinforcement learning paradigm is used to implement visual attention. The attention module selects targets to foveate based on the goal of successful recognition, and uses a new multiple-model Q-Iearning formulation. Given a set of target and distractor gestures, our system can learn where to foveate to maximally discriminate a particular gesture.

****

On Neural Networks with Minimal Weights

Vasken Bohossian, Jehoshua Bruck

Linear threshold elements are the basic building blocks of artificial neural networks. A linear threshold element computes a function that is a sign of a weighted sum of the input variables. The weights are arbitrary integers; actually, they can be very big integers-(cid:173) exponential in the number of the input variables. However, in practice, it is difficult to implement big weights. In the present literature a distinction is made between the two extreme cases: linear threshold functions with polynomial-size weights as opposed to those with exponential-size weights. The main contribution of this paper is to fill up the gap by further refining that separation. Namely, we prove that the class of linear threshold functions with polynomial-size weights can be divided into subclasses according to the degree of the polynomial. In fact, we prove a more general that there exists a minimal weight linear threshold function result- for any arbitrary number of inputs and any weight size. To prove those results we have developed a novel technique for constructing linear threshold functions with minimal weights.

****

Neural Networks with Quadratic VC Dimension

Pascal Koiran, Eduardo Sontag

This paper shows that neural networks which use continuous acti(cid:173) vation functions have VC dimension at least as large as the square of the numbe

r of weights w. This result settles a long-standing open question, nam ely whether the well-known O(w log w) bound, known for hard-threshold nets, also held for more general sigmoidal nets. Implications for the number of s amples needed for valid gen(cid:173) eralization are discussed.

****

Factorial Hidden Markov Models

Zoubin Ghahramani, Michael Jordan

We present a framework for learning in hidden Markov models with distributed s tate representations. Within this framework , we de(cid:173) rive a learn ing algorithm based on the Expectation-Maximization (EM) procedure for max imum likelihood estimation. Analogous to the standard Baum-Welch update ru les, the M-step of our algo(cid:173) rithm is exact and can be solved analy tically. However, due to the combinatorial nature of the hidden state repre sentation, the exact E-step is intractable. A simple and tractable mean field approxima(cid:173) tion is derived. Empirical results on a set of problems su ggest that both the mean field approximation and Gibbs sampling are viable a lternatives to the computationally expensive exact algorithm.

****

Extracting Tree-Structured Representations of Trained Networks

Mark Craven, Jude Shavlik

A significant limitation of neural networks is that the represen(cid:173) 3) tations they learn are usually incomprehensible to humans. We presen t a novel algorithm , TREPAN, for extracting comprehensible, symbolic represe ntations from trained neural networks. Our algo(cid:173) rithm uses queries to induce a decision tree that approximates the concept represented by a gi ven network. Our experiments demon(cid:173) strate that TREPAN is able to pro duce decision trees that maintain a high level of fidelity to their respective networks while being com(cid:173) prehensible and accurate. Unlike previous work in this area, our algorithm is general in its applicability and scal es well to large net(cid:173) works and problems with high-dimensional input spaces.

****

Improving Committee Diagnosis with Resampling Techniques

Bambang Parmanto, Paul Munro, Howard Doyle

Central to the performance improvement of a committee relative to individual ne tworks is the error correlation between networks in the committee. We investi gated methods of achieving error indepen(cid:173) dence between the networks by training the networks with different resampling sets from the original t raining set. The methods were tested on the sinwave artificial task and the real-world problems of hepatoma (liver cancer) and breast cancer diagnoses.

****

The Capacity of a Bump

Gary Flake

Recently, several researchers have reported encouraging experimental re(cid:173) sults when using Gaussian or bump-like activation functions in multilayer perc eptrons. Networks of this type usually require fewer hidden layers and units and often learn much faster than typical sigmoidal networks. To explain th ese results we consider a hyper-ridge network, which is a simple perceptron w ith no hidden units and a rid¥e activation function. If we are interested in pa rtitioningp points in d dimensions into two classes then in the limit as d appr oaches infinity the capacity of a hyper-ridge and a perceptron is identical. H owever, we show that for p ~ d, which is the usual case in practice, the ratio of hyper-ridge to perceptron dichotomies approaches pl2(d + 1).

****

Boosting Decision Trees

Harris Drucker, Corinna Cortes

We introduce a constructive, incremental learning system for regression proble ms that models data by means of locally linear experts. In contrast to other approaches, the experts are trained independently and do not compete fo r data during learning. Only when a prediction for a query is required do

the experts cooperate by blending their individual predic(cid:173) tions. Each expert is trained by minimizing a penalized local cross vali(cid:173) dation error using second order methods. In this way, an expert is able to find a local distance metric by adjusting the size and shape of the recep(cid:173) tive field in which its predictions are valid, and also to detect relevant in(cid:173) put features by adjusting its bias on the importance of individual input dimensions. We derive asymptotic results for our method. In a variety of simulations the properties of the algorithm are demonstrated with respect to interference, learning speed, prediction accuracy, feature detection, and task oriented incremental learning.

****

Some results on convergent unlearning algorithm

Serguei Semenov, Irina Shuvalova

In this paper we consider probabilities of different asymptotics of convergent unlearning algorithm for the Hopfield-type neural net(cid:173) work (Plakhov & Semenov, 1994) treating the case of unbiased random patterns. We show also that failed unlearning results in total memory breakdown.

****

Forward-backward retraining of recurrent neural networks

Andrew W. Senior, Anthony Robinson

This paper describes the training of a recurrent neural network as the letter posterior probability estimator for a hidden Markov model, off-line handwriting recognition system. The network esti(cid:173) mates posterior distributions for each of a series of frames repre(cid:173) senting sections of a handwritten word. The supervised training algorithm, backpropagation through time, requires target outputs to be provided for each frame. Three methods for deriving these targets are presented. A novel method based upon the forward(cid:173) backward algorithm is found to result in the recognizer with the lowest error rate.

****

KODAK lMAGELINK™ OCR Alphanumeric Handprint Module

Alexander Shustorovich, Christopher W. Thrasher

This paper describes the Kodak Imageliok TM OCR alphanumeric handprint module. There are two neural network algorithms at its cme: the first network is trained to find individual characters in an alphamuneric field, while the second one perfmns the classification. Both networks were trained on Gabor projections of the ociginal pixel images, which resulted in higher recognition rates and greater noise immunity. Compared its purely numeric counterpart (Shusurovich and Thrasher, 1995), this version of the system has a significant applicatim specific postprocessing module. The system has been implemented in specialized parallel hardware, which allows it to run at 80 char/sec/board. It has been installed at the Driver and Vehicle Licensing Agency (DVLA) in the United Kingdom. and its overall success rate exceeds 96% (character level without rejects). which translates into 85% field rate. If approximately 20% of the fields are rejected. the system achieves 99.8% character and 99.5% field success rate.

****

Predictive Q-Routing: A Memory-based Reinforcement Learning Approach to Adaptive Traffic Control

Samuel Choi, Dit-Yan Yeung

In this paper, we propose a memory-based Q-Iearning algorithm called predictive Q-routing (PQ-routing) for adaptive traffic con(cid:173) trol. We attempt to address two problems encountered in Q-routing (Boyan & Littman, 1994), namely, the inability to fine-tune rout(cid:173) ing policies under low network load and the inability to learn new optimal policies under decreasing load conditions. Unlike other memory-based reinforcement learning algorithms in which mem(cid:173) ory is used to keep past experiences to increase learning speed, PQ-routing keeps the best experiences

learned and reuses them by predicting the traffic trend. The effectiv eness of PQ-routing has been verified under various network topologies and traffic con(cid:173) ditions. Simulation results show that PQ-routing is superior to Q-routing in terms of both learning speed and adaptability.

****

High-Performance Job-Shop Scheduling With A Time-Delay TD(λ) Network

Wei Zhang, Thomas Dietterich

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.

****

Family Discovery

Stephen Omohundro

"Family discovery" is the task of learning the dimension and struc(cid:173) tur e of a parameterized family of stochastic models. It is espe(cid:173) ci ally appropriate when the training examples are partitioned into "episodes" of samples drawn from a single parameter value. We present three family discovery algorithms based on surface learn(cid:173) ing and show that they significantly improve performance over two alternatives on a parameterized cl assification task.

****

Modeling Interactions of the Rat's Place and Head Direction Systems

A. Redish, David Touretzky

We have developed a computational theory of rodent navigation that incl udes analogs of the place cell system, the head direction system, and path int egration. In this paper we present simulation results showing how interactions between the place and head direction systems can account for recent observatio ns about hippocampal place cell responses to doubling and/or rotation of cue cards in a cylindrical arena (Sharp et at., 1990).

****

Empirical Entropy Manipulation for Real-World Problems

Paul Viola, Nicol Schraudolph, Terrence J. Sejnowski

No finite sample is sufficient to determine the density, and therefore the entr opy, of a signal directly. Some assumption about either the functional form of the density or about its smoothness is necessary. Both amount to a prior over the space of possible density functions. By far the most common approach is to assume that the density has a parametric form.

****

A Neural Network Model of 3-D Lightness Perception

Luiz Pessoa, William Ross

A neural network model of 3-D lightness perception is presented which builds upon the FACADE Theory Boundary Contour Sys(cid:173) tem/Feature C ontour System of Grossberg and colleagues. Early ratio encoding by retina l ganglion neurons as well as psychophysi(cid:173) cal results on constancy across different backgrounds (background constancy) are used to provide fun ctional constraints to the theory and suggest a contrast negation hypothes is which states that ratio measures between coplanar regions are given more weight in the determination of lightness of the respective regions. Simulations of the model address data on lightness perception, includin g the coplanar ratio hypothesis, the Benary cross, and White's illusion.

****

Adaptive Retina with Center-Surround Receptive Field

Shih-Chii Liu, Kwabena Boahen

Both vertebrate and invertebrate retinas are highly efficient in ex(cid:173) tr acting contrast independent of the background intensity over five or more dec ades. This efficiency has been rendered possible by the adaptation of th e DC operating point to the background inten(cid:173) sity while maintaining hig h gain transient responses. The center(cid:173) surround properties of the ret ina allows the system to extract in(cid:173) formation at the edges in the imag

e. This silicon retina models the adaptation properties of the receptors and t he antagonistic center(cid:173) surround properties of the laminar cells of the invertebrate retina and the outer-plexiform layer of the vertebrate retina. W e also illus(cid:173) trate the spatio-temporal responses of the silicon retina on moving bars. The chip has 59x64 pixels on a 6.9x6.8mm2 die and it is fabricated in 2 J-tm n-well technology.

****

Does the Wake-sleep Algorithm Produce Good Density Estimators?

Brendan J. Frey, Geoffrey E. Hinton, Peter Dayan

The wake-sleep algorithm (Hinton, Dayan, Frey and Neal 1995) is a rel(cid:173) a tively efficient method of fitting a multilayer stochastic generative model to high-dimensional data. In addition to the top-down connec(cid:173) tions in th e generative model, it makes use of bottom-up connections for approximating the probability distribution over the hidden units given the data, and it trains t hese bottom-up connections using a simple delta rule. We use a variety of synth etic and real data sets to compare the per(cid:173) formance of the wake-sleep a lgorithm with Monte Carlo and mean field methods for fitting the same generati ve model and also compare it with other models that are less powerful but easie r to fit.

****

EM Optimization of Latent-Variable Density Models

Christopher Bishop, Markus Svensén, Christopher Williams

There is currently considerable interest in developing general non(cid:173) l inear density models based on latent, or hidden, variables. Such model s have the ability to discover the presence of a relatively small number of und erlying 'causes' which, acting in combination, give rise to the apparent complexity of the observed data set. Unfortu(cid:173) nately, to train such m odels generally requires large computational effort. In this paper we introdu ce a novel latent variable algorithm which retains the general non-linear cap abilities of previous models but which uses a training procedure based on the EM algorithm. We demonstrate the performance of the model on a toy problem and on data from flow diagnostics for a multi-phase oil pipeli ne.

****

Modern Analytic Techniques to Solve the Dynamics of Recurrent Neural Networks

A.C.C. Coolen, S. Laughton, D. Sherrington

We describe the use of modern analytical techniques in solving the dynamics of symmetric and nonsymmetric recurrent neural net(cid:173) works near sa turation. These explicitly take into account the cor(cid:173) relations between the post-synaptic potentials, and thereby allow for a reliable predict ion of transients.

****

A Realizable Learning Task which Exhibits Overfitting

Siegfried Bös

In this paper we examine a perceptron learning task. The task is real izable since it is provided by another perceptron with identi(cid:173) cal architecture. Both perceptrons have nonlinear sigmoid output functio ns. The gain of the output function determines the level of nonlinearity o f the learning task. It is observed that a high level of nonlinearity l eads to overfitting. We give an explanation for this rather surprising obse rvation and develop a method to avoid the overfitting. This method ha s two possible interpretations, one is learning with noise, the other cro ss-validated early stopping.

****

Simulation of a Thalamocortical Circuit for Computing Directional Heading in the Rat

Hugh Blair

Several regions of the rat brain contain neurons known as head-direc(ci d:173) tion celis, which encode the animal's directional heading during spatial navigation. This paper presents a biophysical model of head-direction ce

ll acti vity, which suggests that a thalamocortical circuit might com(cid:173) d:173) pute the rat's head direction by integrating the angular velocity of the head over time. The model was implemented using the neural simulator NEURON, and makes testable predictions about the structure and func(cid:173) ti on of the rat head-direction circuit.

****

Clustering data through an analogy to the Potts model

Marcelo Blatt, Shai Wiseman, Eytan Domany

A new approach for clustering is proposed. This method is based on an a nalogy to a physical model; the ferromagnetic Potts model at thermal equili brium is used as an analog computer for this hard optimization problem . We do not assume any structure of the un(cid:173) derlying distribution of th e data. Phase space of the Potts model is divided into three regions; ferro magnetic, super-paramagnetic and paramagnetic phases. The region of interes t is that corresponding to the super-paramagnetic one, where domains of ali gned spins ap(cid:173) pear. The range of temperatures where these structures are stable is indicated by a non-vanishing magnetic susceptibility. We use a very efficient Monte Carlo algorithm to measure the susceptibil (cid:173) ity and the spin spin correlation function. The values of the spin spin correlation function, at the super-paramagnetic phase, serve to identify the partition of the data points into clusters.

****

Learning Fine Motion by Markov Mixtures of Experts

Marina Meila, Michael Jordan

Compliant control is a standard method for performing fine manip(cid:173) ulati on tasks, like grasping and assembly, but it requires estimation of the state o f contact (s.o.c.) between the robot arm and the ob(cid:173) jects involv ed. Here we present a method to learn a model of the movement from meas ured data. The method requires little or no prior knowledge and the r esulting model explicitly estimates the s.o.c. The current s.o.c. is v iewed as the hidden state variable of a discrete HMM. The control de pendent transition probabilities between states are modeled as parametri zed functions of the mea(cid:173) surement. We show that their parameters can be estimated from measurements at the same time as the parameters of th e movement in each s.o.c. The learning algorithm is a variant of the EM p roce(cid:173) dure. The E step is computed exactly ; solving the M step e xactly is not possible in general. Here, gradient ascent is used to prod uce an increase in likelihood .

****

Stable LInear Approximations to Dynamic Programming for Stochastic Control Probl ems with Local Transitions

Benjamin Van Roy, John Tsitsiklis

We consider the solution to large stochastic control problems by means of methods that rely on compact representations and a vari(cid:173) ant of the value iteration algorithm to compute approximate cost(cid:173) to-go functions. While such methods are known to be unstable in general, we identify a new cl ass of problems for which convergence, as well as graceful error bounds, are guaranteed. This class in(cid:173) volves linear parameterizations of t he cost-to- go function together with an assumption that the dynamic progr amming operator is a contraction with respect to the Euclidean norm whe n applied to functions in the parameterized class. We provide a speci al case where this assumption is satisfied, which relies on the local ity of transitions in a state space. Other cases will be discussed in a fu ll length version of this paper.

****

Generalisation of A Class of Continuous Neural Networks

John Shawe-Taylor, Jieyu Zhao

We propose a way of using boolean circuits to perform real valued computat ion in a way that naturally extends their boolean func(cid:173) tionali ty. The functionality of multiple fan in threshold gates in this model

is shown to mimic that of a hardware implementation of continuous Neural Networks. A Vapnik-Chervonenkis dimension and sample size analysis for the systems is performed giving best known sample sizes for a real valued Neural Network. Experimen(cid:173) tal results confirm the conclusion that the sample sizes required for the networks are significantly smaller than for sigmoidal networks.

****

Visual gesture-based robot guidance with a modular neural system

Enno Littmann, Andrea Drees, Helge Ritter

We report on the development of the modular neural system "SEE(cid:173) EAGLE" for the visual guidance of robot pick-and-place actions. Several neural networks are integrated to a single system that vi(cid:173) sually recognizes human hand pointing gestures from stereo pairs of color video images. The output of the hand recognition stage is processed by a set of color-sensitive neural networks to determine the cartesian location of the target object that is referenced by the pointing gesture. Finally, this information is used to guide a robot to grab the target object and put it at another location that can be specified by a second pointing gesture. The accuracy of the cur(cid:173) rent system allows to identify the location of the referenced target object to an accuracy of 1 cm in a workspace area of 50x50 cm. In our current environment, this is sufficient to pick and place arbi(cid:173) trarily positioned target objects within the workspace. The system consists of neural networks that perform the tasks of image seg(cid:173) mentation, estimation of hand location, estimation of 3D-pointing direction, object recognition, and necessary coordinate transforms. Drawing heavily on the use of learning algorithms, the functions of all network modules were created from data examples only.

****

Symplectic Nonlinear Component Analysis

Lucas Parra

Statistically independent features can be extracted by finding a fac(cid:173) torial representation of a signal distribution. Principal Component Analysis (PCA) accomplishes this for linear correlated and Gaus(cid:173) sian distributed signals. Independent Component Analysis (ICA), formalized by Comon (1994), extracts features in the case of lin(cid:173) ear statistical dependent but not necessarily Gaussian distributed signals. Nonlinear Component Analysis finally should find a facto(cid:173) rial representation for nonlinear statistical dependent distributed signals. This paper proposes for this task a novel feed-forward, information conserving, nonlinear map - the explicit symplectic transformations. It also solves the problem of non-Gaussian output distributions by considering single coordinate higher order statis(cid:173) tics.

****

Independent Component Analysis of Electroencephalographic Data

Scott Makeig, Anthony Bell, Tzyy-Ping Jung, Terrence J. Sejnowski

Because of the distance between the skull and brain and their differ(cid:173) ent resistivities, electroencephalographic (EEG) data collected from any point on the human scalp includes activity generated within a large brain area. This spatial smearing of EEG data by volume conduction does not involve significant time delays, however, sug(cid:173) gesting that the Independent Component Analysis (ICA) algorithm of Bell and Sejnowski [1] is suitable for performing blind source sep(cid:173) aration on EEG data. The ICA algorithm separates the problem of source identification from that of source localization. First results of applying the ICA algorithm to EEG and event-related potential (ERP) data collected during a sustained auditory detection task show: (1) ICA training is insensitive to different random seeds. (2) ICA may be used to segregate obvious artifactual EEG components (line and muscle noise, eye movements) from other sources. (3) ICA is capable of isolating overlapping EEG phenomena, including al(cid:173) pha and theta bursts and spatially-separable ERP components, to separate ICA channels. (4) Nonstationarit

ies in EEG and behav(cid:173) ioral state can be tracked using ICA via cha nges in the amount of residual correlation between ICA-filtered output c hannels.
****

Competence Acquisition in an Autonomous Mobile Robot using Hardware Neural Techn iques
Geoffrey Jackson, Alan Murray
In this paper we examine the practical use of hardware neural network s in an autonomous mobile robot. We have developed a hardware neural system based around a custom VLSI chip, EP(cid:173) SILON III, designe d specifically for embedded hardware neural applications. We present he re a demonstration application of an autonomous mobile robot that highlight s the flexibility of this sys(cid:173) tem. This robot gains basic mobility com petence in very few train(cid:173) ing epochs using an "instinct-rule" trai ning methodology.
****

Gaussian Processes for Regression
Christopher Williams, Carl Rasmussen
The Bayesian analysis of neural networks is difficult because a sim(cid:173) ple prior over weights implies a complex prior distribution over function s . In this paper we investigate the use of Gaussian process priors over fun ctions, which permit the predictive Bayesian anal(cid:173) ysis for fixe d values of hyperparameters to be carried out exactly using matrix operat ions. Two methods, using optimization and av(cid:173) eraging (via Hybrid Monte Carlo) over hyperparameters have been tested on a number of challenging problems and have produced excellent results.
****

Reorganisation of Somatosensory Cortex after Tactile Training
Rasmus Petersen, John Taylor
Topographic maps in primary areas of mammalian cerebral cortex reor(cid:173) ga nise as a result of behavioural training. The nature of this reorgani(cid:1 73) sation seems consistent with the behaviour of competitive neural net (cid:173) works, as has been demonstrated in the past by computer simu lation. We model tactile training on the hand representation in primate somato(cid:173) sensory cortex, using the Neural Field Theory of Amari and his col(cid:173) leagues. Expressions for changes in both receptive field size and mag(cid:173) nification factor are derived, which are consistent with owl monkey ex(cid:173) periments and make a prediction which goes beyond them.
****

Prediction of Beta Sheets in Proteins
Anders Krogh, Soren Riis
Most current methods for prediction of protein secondary structure use a small window of the protein sequence to predict the structure of the central amino ac id. We describe a new method for prediction of the non-local structure call ed ,8-sheet, which consists of two or more ,8-strands that are connect ed by hydrogen bonds. Since,8- strands are often widely separated in the protein chain, a network with two windows is introduced. After training on a set of proteins the network predicts the sheets well, but there are many f alse pos(cid:173) itives. By using a global energy function the ,8-sheet prediction is combined with a local prediction of the three secondary structu res a-helix, ,8-strand and coil. The energy function is minimized using simul ated annealing to give a final prediction.
****

Active Learning in Multilayer Perceptrons
Kenji Fukumizu
We propose an active learning method with hidden-unit reduction. which is dev ised specially for multilayer perceptrons (MLP). First, we review our active learning method, and point out that many Fisher-information-based metho ds applied to MLP have a critical problem: the information matrix may be singular. To solve this problem, we derive the singularity condition o

f an information ma(cid:173) trix, and propose an active learning technique tha t is applicable to MLP. Its effectiveness is verified through experiments.
****

A model of transparent motion and non-transparent motion aftereffects
Alexander Grunewald
A model of human motion perception is presented. The model contains tw o stages of direction selective units. The first stage con(cid:173) tains broa dly tuned units, while the second stage contains units that are narrow ly tuned. The model accounts for the motion af(cid:173) tereffect throu gh adapting units at the first stage and inhibitory interactions at the second stage. The model explains how two pop(cid:173) ulations of dots moving in slightly different directions are perceived as a single population movin g in the direction of the vector sum, and how two populations moving in st rongly different directions are perceived as transparent motion. The model al so explains why the motion aftereffect in both cases appears as non-transparent motion.
****

Improved Gaussian Mixture Density Estimates Using Bayesian Penalty Terms and Net work Averaging
Dirk Ormoneit, Volker Tresp
Volker Tresp
****

A Practical Monte Carlo Implementation of Bayesian Learning
Carl Rasmussen
A practical method for Bayesian training of feed-forward neural networks using sophisticated Monte Carlo methods is presented and evaluated. In reasonably small amounts of computer time this approach outperforms other s tate-of-the-art methods on 5 data(cid:173) limited tasks from real world d omains.
****

A Smoothing Regularizer for Recurrent Neural Networks
Lizhong Wu, John Moody
We derive a smoothing regularizer for recurrent network models by requiring rob ustness in prediction performance to perturbations of the training data. The r egularizer can be viewed as a generaliza(cid:173) tion of the first order Tik honov stabilizer to dynamic models. The closed-form expression of the regul arizer covers both time-lagged and simultaneous recurrent nets, with feed forward nets and one(cid:173) layer linear nets as special cases. We have s uccessfully tested this regularizer in a number of case studies and found that it performs better than standard quadratic weight decay.
****

REMAP: Recursive Estimation and Maximization of A Posteriori Probabilities - App lication to Transition-Based Connectionist Speech Recognition
Yochai Konig, Hervé Bourlard, Nelson Morgan
In this paper, we introduce REMAP, an approach for the training and est imation of posterior probabilities using a recursive algorithm that is reminis cent of the EM-based Forward-Backward (Liporace 1982) algorithm for the e stimation of sequence likelihoods. Al(cid:173) though very general, the m ethod is developed in the context of a statistical model for transition -based speech recognition using Ar(cid:173) tificial Neural Networks (ANN) to generate probabilities for Hid(cid:173) den Markov Models (HMMs). In the new approach, we use local conditional posterior probabilities of tr ansitions to estimate global posterior probabilities of word sequences. Alth ough we still use ANNs to estimate posterior probabilities, the networ k is trained with targets that are themselves estimates of local posterior pr oba(cid:173) bilities. An initial experimental result shows a significant decr ease in error-rate in comparison to a baseline system.
****

Harmony Networks Do Not Work
René Gourley

Harmony networks have been proposed as a means by which con(cid:173) nectionist models can perform symbolic computation. Indeed, pro(cid:173) ponents claim that a harmony network can be built that constructs parse trees for strings in a context free language. This paper shows that harmony networks do not work in the following sense: they construct many outputs that are not valid parse trees.

****

Learning Sparse Perceptrons

Jeffrey Jackson, Mark Craven

We introduce a new algorithm designed to learn sparse percep(cid:173) trons over input representations which include high-order features. Our algorithm, which is based on a hypothesis-boosting method, is able to PAC-learn a relatively natural class of target concepts. Moreover, the algorithm appears to work well in practice: on a set of three problem domains, the algorithm produces classifiers that utilize small numbers of features yet exhibit good generalization performance. Perhaps most importantly, our algorithm generates concept descriptions that are easy for humans to understand.

****

The Role of Activity in Synaptic Competition at the Neuromuscular Junction

Samuel Joseph, David Willshaw

An extended version of the dual constraint model of motor end(cid:173) plate morphogenesis is presented that includes activity dependent and independent competition. It is supported by a wide range of recent neurophysiological evidence that indicates a strong relation(cid:173) ship between synaptic efficacy and survival. The computational model is justified at the molecular level and its predictions match the developmental and regenerative behaviour of real synapses.

****

Absence of Cycles in Symmetric Neural Networks

Xin Wang, Arun Jagota, Fernanda Botelho, Max Garzon

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.

****

Explorations with the Dynamic Wave Model

Thomas Rebotier, Jeffrey Elman

Following Shrager and Johnson (1995) we study growth of logi(cid:173) cal function complexity in a network swept by two overlapping waves: one of pruning, and the other of Hebbian reinforcement of connections. Results indicate a significant spatial gradient in the appearance of both linearly separable and non linearly separable functions of the two inputs of the network; the n.l.s. cells are much sparser and their slope of appearance is sensitive to parameters in a highly non-linear way.

****

Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding

Richard S. Sutton

On large problems, reinforcement learning systems must use parame(cid:173) terized function approximators such as neural networks in order to gen(cid:173) eralize between similar situations and actions. In these cases there are no strong theoretical results on the accuracy of convergence, and com(cid:173) putational results have been mixed. In particular, Boyan and Moore reported at last year's meeting a series of negative results in attempting to apply dynamic programming together with function approximation to simple control problems with continuous state spaces. In this paper, we present positive results for all the control tasks they attempted, and for one that is significantly larger. The most important differences are that we used sparse-coarse-coded function approximators (CMACs) whereas they used mostly global function approximators, and that we learned online whereas they learned offline. Bo

yan and Moore and others have suggested that the problems they encoun tered could be solved by using actual outcomes ("rollouts"), as in cl assical Monte Carlo methods, and as in the TD().) algorithm when). = 1. H owever, in our experiments this always resulted in substantially poorer perfor (cid:173) mance. We conclude that reinforcement learning can work robust ly in conjunction with function approximators, and that there is littl e justification at present for avoiding the case of general)..

****

A Multiscale Attentional Framework for Relaxation Neural Networks

Dimitris Tsioutsias, Eric Mjolsness

We investigate the optimization of neural networks governed by general objective functions. Practical formulations of such objec(cid:173) tives are notoriously difficult to solve; a common problem is the poor local extr ema that result by any of the applied methods. In this paper, a novel fr amework is introduced for the solution oflarge(cid:173) scale optimization prob lems. It assumes little about the objective function and can be applied to g eneral nonlinear, non-convex func(cid:173) tions; objectives in thousand of v ariables are thus efficiently min(cid:173) imized by a combination of tech niques - deterministic annealing, multiscale optimization, attention mechanis ms and trust region op(cid:173) timization methods.

****

VLSI Model of Primate Visual Smooth Pursuit

Ralph Etienne-Cummings, Jan Van der Spiegel, Paul Mueller

A one dimensional model of primate smooth pursuit mechanism has been i mplemented in 2 11m CMOS VLSI. The model consolidates Robinson's negat ive feedback model with Wyatt and Pola's positive feedback scheme, to pr oduce a smooth pursuit system which zero's the velocity of a target on the retina. Furthermore, the system uses the current eye motion as a predictor for future target motion. Analysis, stability and biological cor respondence of the system are discussed. For implementation at the focal p lane, a local correlation based visual motion detection technique is u sed. Velocity measurements, ranging over 4 orders of magnitude with < 15% v ariation, provides the input to the smooth pursuit system. The system per formed successful velocity tracking for high contrast scenes. Circuit design and performance of the complete smooth pursuit system is presented.

****

Reinforcement Learning by Probability Matching

Philip N. Sabes, Michael Jordan

We present a new algorithm for associative reinforcement learn(cid:173) ing. The algorithm is based upon the idea of matching a network's output proba bility with a probability distribution derived from the environment's reward si gnal. This Probability Matching algorithm is shown to perform faster and be less susceptible to local minima than previously existing algorithms. W e use Probability Match(cid:173) ing to train mixture of experts networks, a n architecture for which other reinforcement learning rules fail to converge reliably on even simple problems. This architecture is particularly well suited for our algorithm as it can compute arbitrarily complex functions y et calculation of the output probability is simple.

****

Generalized Learning Vector Quantization

Atsushi Sato, Keiji Yamada

We propose a new learning method, "Generalized Learning Vec(cid:173) tor Quantization (GLVQ)," in which reference vectors are updated based on the ste epest descent method in order to minimize the cost function . The cost funct ion is determined so that the obtained learning rule satisfies the co nvergence condition. We prove that Kohonen's rule as used in LVQ does not satisfy the convergence condition and thus degrades recognition a bility. Experimental re(cid:173) sults for printed Chinese character recogn ition reveal that GLVQ is superior to LVQ in recognition ability.

****

Bayesian Methods for Mixtures of Experts
Steve Waterhouse, David MacKay, Anthony Robinson
We present a Bayesian framework for inferring the parameters of a mixture of experts model based on ensemble learning by varia(cid:173) tional free energy minimisation. The Bayesian approach avoids the over-fitting and noise level under-estimation problems of traditional maximum likelihood inference. We demonstrate these methods on artificial problems and sunspot time series prediction.
****

High-Speed Airborne Particle Monitoring Using Artificial Neural Networks
Alistair Ferguson, Theo Sabisch, Paul Kaye, Laurence Dixon, Hamid Bolouri
Current environmental monitoring systems assume particles to be spherical, and do not attempt to classify them. A laser-based sys(cid:173) tem developed at the University of Hertfordshire aims at classify(cid:173) ing airborne particles through the generation of two-dimensional scattering profiles. The pedormances of template matching, and two types of neural network (HyperNet and semi-linear units) are compared for image classification. The neural network approach is shown to be capable of comparable recognition pedormance, while offering a number of advantages over template matching.
****

Adaptive Mixture of Probabilistic Transducers
Yoram Singer
We introduce and analyze a mixture model for supervised learning of probabilistic transducers. We devise an online learning algorithm that efficiently infers the structure and estimates the parameters of each model in the mixture. Theoretical analysis and comparative simulations indicate that the learning algorithm tracks the best model from an arbitrarily large (possibly infinite) pool of models. We also present an application of the model for inducing a noun phrase recognizer.
****

SEEMORE: A View-Based Approach to 3-D Object Recognition Using Multiple Visual Cues
Bartlett Mel
A neurally-inspired visual object recognition system is described called SEEMORE, whose goal is to identify common objects from a large known set-independent of 3-D viewiag angle, distance, and non-rigid distortion. SEEMORE's database consists of 100 ob(cid:173) jects that are rigid (shovel), non-rigid (telephone cord), articu(cid:173) lated (book), statistical (shrubbery), and complex (photographs of scenes). Recognition results were obtained using a set of 102 color and shape feature channels within a simple feedforward network ar(cid:173) chitecture. In response to a test set of 600 novel test views (6 of each object) presented individually in color video images, SEEMORE identified the object correctly 97% of the time (chance is 1%) using a nearest neighbor classifier. Similar levels of performance were obtained for the subset of 15 non-rigid objects. Generalization be(cid:173) havior reveals emergence of striking natural category structure not explicit in the input feature dimensions.
****

Using Pairs of Data-Points to Define Splits for Decision Trees
Geoffrey E. Hinton, Michael Revow
Conventional binary classification trees such as CART either split the data using axis-aligned hyperplanes or they perform a compu(cid:173) tationally expensive search in the continuous space of hyperplanes with unrestricted orientations. We show that the limitations of the former can be overcome without resorting to the latter. For every pair of training data-points, there is one hyperplane that is orthog(cid:173) onal to the line joining the data-points and bisects this line. Such hyperplanes are plausible candidates for splits. In a comparison on a suite of 12 datasets we found that this method of generating candidate splits outperformed the standard methods, particularly when the training sets were small.

****
Dynamics of On-Line Gradient Descent Learning for Multilayer Neural Networks
David Saad, Sara Solla
We consider the problem of on-line gradient descent learning for general two-layer neural networks. An analytic solution is pre(cid:173) sented and used to investigate the role of the learning rate in con(cid:173) trolling the evolution and convergence of the learning process.
****
Geometry of Early Stopping in Linear Networks
Robert Dodier
A theory of early stopping as applied to linear models is presented. The backpropagation learning algorithm is modeled as gradient descent in continuous time. Given a training set and a validation set, all weight vectors found by early stopping must lie on a cer(cid:173) tain quadric surface, usually an ellipsoid. Given a training set and a candidate early stopping weight vector, all validation sets have least-squares weights lying on a certain plane. This latter fact can be exploited to estimate the probability of stopping at any given point along the trajectory from the initial weight vector to the least(cid:173) squares weights derived from the training set, and to estimate the probability that training goes on indefinitely. The prospects for extending this theory to nonlinear models are discussed.
****
A Unified Learning Scheme: Bayesian-Kullback Ying-Yang Machine
Lei Xu
A Bayesian-Kullback learning scheme, called Ying-Yang Machine, is proposed based on the two complement but equivalent Bayesian representations for joint density and their Kullback divergence. Not only the scheme unifies existing major supervised and unsu(cid:173) pervised learnings, including the classical maximum likelihood or least square learning, the maximum information preservation, the EM & em algorithm and information geometry, the recent popular Helmholtz machine, as well as other learning methods with new variants and new results; but also the scheme provides a number of new learning models.
****
Using Feedforward Neural Networks to Monitor Alertness from Changes in EEG Correlation and Coherence
Scott Makeig, Tzyy-Ping Jung, Terrence J. Sejnowski
We report here that changes in the normalized electroencephalo(cid:173) graphic (EEG) cross-spectrum can be used in conjunction with feedforward neural networks to monitor changes in alertness of op(cid:173) erators continuously and in near-real time. Previously, we have shown that EEG spectral amplitudes covary with changes in alert(cid:173) ness as indexed by changes in behavioral error rate on an auditory detection task [6,4]. Here, we report for the first time that increases in the frequency of detection errors in this task are also accompa(cid:173) nied by patterns of increased and decreased spectral coherence in several frequency bands and EEG channel pairs. Relationships between EEG coherence and performance vary between subjects, but within subjects, their topographic and spectral profiles appear stable from session to session. Changes in alertness also covary with changes in correlations among EEG waveforms recorded at different scalp sites, and neural networks can also estimate alert(cid:173) ness from correlation changes in spontaneous and unobtrusively(cid:173) recorded EEG signals.
****
A Dynamical Model of Context Dependencies for the Vestibulo-Ocular Reflex
Olivier Coenen, Terrence J. Sejnowski
The vestibulo-ocular reflex (VOR) stabilizes images on the retina during rapid head motions. The gain of the VOR (the ratio of eye to head rotation velocity) is typically around -1 when the eyes are focused on a distant target. However, to stabilize images accurately, the VOR gain must vary with context

(eye position, eye vergence and head translation). We first describe a ki nematic model of the VOR which relies solely on sensory information available from the semicircular canals (head rotation), the otoliths (head translation), and neural correlates of eye position and vergence angle. We then propose a dy namical model and compare it to the eye velocity responses measured in monkeys. The dynamical model repro(cid:173) duces the observed amplitude and time course of the modulation of the VOR and suggests one way to combine the required neur al signals within the cerebellum and the brain stem. It also makes predictions for the responses of neurons to multiple inputs (head rotation and translation , eye position, etc.) in the oculomotor system.

****

Constructive Algorithms for Hierarchical Mixtures of Experts

Steve Waterhouse, Anthony Robinson

We present two additions to the hierarchical mixture of experts (HME) architecture. By applying a likelihood splitting criteria to each expert in the HME we "grow" the tree adaptively during train(cid:173) ing. Secondly, by considering only the most probable path through the tree we may "prune" bran ches away, either temporarily, or per(cid:173) manently if they become redund ant. We demonstrate results for the growing and path pruning algorithm s which show significant speed ups and more efficient use of parameters over the standard fixed structure in discriminating between two interlo cking spirals and classifying 8-bit parity patterns.

****

Discovering Structure in Continuous Variables Using Bayesian Networks

Reimar Hofmann, Volker Tresp

We study Bayesian networks for continuous variables using non(cid:173) l inear conditional density estimators. We demonstrate that use(cid:173) fu l structures can be extracted from a data set in a self-organized way and we present sampling techniques for belief update based on Markov blanket conditional density models.

****

Temporal coding in the sub-millisecond range: Model of barn owl auditory pathway

Richard Kempter, Wulfram Gerstner, J. van Hemmen, Hermann Wagner

Binaural coincidence detection is essential for the localization of ext ernal sounds and requires auditory signal processing with high temporal precision. We present an integrate-and-fire model of spike processing in the a uditory pathway of the barn owl. It is shown that a temporal precision in the microsecond range can be achieved with neuronal time constants which are at least one magnitude longer. An important feature of our model is an u nsupervised Hebbian learning rule which leads to a temporal fine tuning of t he neuronal connections.

****

Stable Dynamic Parameter Adaption

Stefan Rüger

A stability criterion for dynamic parameter adaptation is given. In the case of the learning rate of backpropagation, a class of stable algorithms is pre sented and studied, including a convergence proof.

****

Optimizing Cortical Mappings

Geoffrey Goodhill, Steven Finch, Terrence J. Sejnowski

"Topographic" mappings occur frequently in the brain. A pop(cid:173) ular ap proach to understanding the structure of such mappings is to map points repres enting input features in a space of a few dimensions to points in a 2 dime nsional space using some self(cid:173) organizing algorithm. We argue that a more general approach may be useful, where similarities between features a re not con(cid:173) strained to be geometric distances, and the objective funct ion for topographic matching is chosen explicitly rather than being spec(cid:173) 17 3) ified implicitly by the self-organizing algorithm. We investigate analytic ally an example of this more general approach applied to the structure of inte rdigitated mappings, such as the pattern of ocular dominance columns in primar

y visual cortex.
****

Experiments with Neural Networks for Real Time Implementation of Control
Peter Campbell, Michael Dale, Herman Ferrá, Adam Kowalczyk
This paper describes a neural network based controller for allocating capacity in a telecommunications network. This system was proposed in order to overcome a "real time" response constraint. Two basic architectures are evaluated: 1) a feedforward network-heuristic and; 2) a feedforward network-recurrent network. These architectures are compared against a linear programming (LP) optimiser as a benchmark. This LP optimiser was also used as a teacher to label the data samples for the feedforward neural network training algorithm. It is found that the systems are able to provide a traffic throughput of 99% and 95%, respectively, of the throughput obtained by the linear programming solution. Once trained, the neural network based solutions are found in a fraction of the time required by the LP optimiser.
****

Cholinergic suppression of transmission may allow combined associative memory function and self-organization in the neocortex
Michael Hasselmo, Milos Cekic
Selective suppression of transmission at feedback synapses during learning is proposed as a mechanism for combining associative feed(cid:173)back with self-organization of feed forward synapses. Experimental data demonstrates cholinergic suppression of synaptic transmission in layer I (feedback synapses), and a lack of suppression in layer IV (feed(cid:173)forward synapses). A network with this feature uses local rules to learn mappings which are not linearly separable. During learning, sensory stimuli and desired response are simultaneously presented as input. Feedforward connections form self-organized representations of input, while suppressed feedback connections learn the transpose of feedfor(cid:173)ward connectivity. During recall, suppression is removed, sensory input activates the self-organized representation, and activity generates the learned response.
****

Dynamics of Attention as Near Saddle-Node Bifurcation Behavior
Hiroyuki Nakahara, Kenji Doya
In consideration of attention as a means for goal-directed behav(cid:173)ior in non-stationary environments, we argue that the dynamics of attention should satisfy two opposing demands: long-term main(cid:173)tenance and quick transition. These two characteristics are con(cid:173)tradictory within the linear domain. We propose the near saddle(cid:173)node bifurcation behavior of a sigmoidal unit with self-connection as a candidate of dynamical mechanism that satisfies both of these demands. We further show in simulations of the 'bug-eat-food' tasks that the near saddle-node bifurcation behavior of recurrent networks can emerge as a functional property for survival in non(cid:173)stationary environments.
****

Softassign versus Softmax: Benchmarks in Combinatorial Optimization
Steven Gold, Anand Rangarajan
A new technique, termed soft assign, is applied for the first time to two classic combinatorial optimization problems, the travel(cid:173)ing salesman problem and graph partitioning. Soft assign, which has emerged from the recurrent neural network/statistical physics framework, enforces two-way (assignment) constraints without the use of penalty terms in the energy functions. The soft assign can also be generalized from two-way winner-take-all constraints to multiple membership constraints which are required for graph par(cid:173)titioning. The soft assign technique is compared to the softmax (Potts glass). Within the statistical physics framework, softmax and a penalty term has been a widely used method for enforcing the two-way constraints common within many combinatorial optimiza(cid:173)tion problems. The benchmarks present evidence that soft assign has clear advantages in accuracy, speed, parallelizabilityand algo(cid:173)rithmic simplici

ty over softmax and a penalty term in optimization problems with two-way constr aints.

****

From Isolation to Cooperation: An Alternative View of a System of Experts

Stefan Schaal, Christopher Atkeson

We introduce a constructive, incremental learning system for regression proble ms that models data by means of locally linear experts. In contrast to other approaches, the experts are trained independently and do not compete fo r data during learning. Only when a prediction for a query is required do the experts cooperate by blending their individual predic(cid:173) tions. Each expert is trained by minimizing a penalized local cross vali(cid:173) d ation error using second order methods. In this way, an expert is able to find a local distance metric by adjusting the size and shape of the recep(cid:173) t ive field in which its predictions are valid, and also to detect relevant in(cid :173) put features by adjusting its bias on the importance of individual i nput dimensions. We derive asymptotic results for our method. In a variety of simulations the properties of the algorithm are demonstrated with respect to i nterference, learning speed, prediction accuracy, feature detection, and task oriented incremental learning.

****

Fast Learning by Bounding Likelihoods in Sigmoid Type Belief Networks

Tommi Jaakkola, Lawrence Saul, Michael Jordan

Sigmoid type belief networks, a class of probabilistic neural net(cid:173) works, provide a natural framework for compactly representing probabil istic information in a variety of unsupervised and super(cid:173) vised l earning problems. Often the parameters used in these net(cid:173) works need to be learned from examples. Unfortunately, estimat(cid:173) ing the parameters via exact probabilistic calculations (i.e, the EM-algorithm) is intractable even for networks with fairly small numbers of hidden units. We propose to avoid the infeasibility of the E step by bounding likelihoods instead of computing them ex(cid:173) actly. We introduce extended and complementary representations for these networks and show that the estimation of the network parameters can be made fast (reduced to qua dratic optimization) by performing the estimation in either of the alternativ e domains. The complementary networks can be used for continuous density estimation as well.

****

Neural Control for Nonlinear Dynamic Systems

Ssu-Hsin Yu, Anuradha Annaswamy

A neural network based approach is presented for controlling two distinct types of nonlinear systems. The first corresponds to nonlinear systems with para metric uncertainties where the parameters occur nonlinearly. The second co rresponds to systems for which stabilizing control struc(cid:173) tures can not be determined. The proposed neural controllers are shown to result in clos ed-loop system stability under certain conditions.

****

Sample Complexity for Learning Recurrent Perceptron Mappings

Bhaskar DasGupta, Eduardo Sontag

Recurrent perceptron classifiers generalize the classical perceptron model. They take into account those correlations and dependences among input coord inates which arise from linear digital filtering. This paper provides tig ht bounds on sample complexity associated to the fitting of such models to exp erimental data.

****

Is Learning The n-th Thing Any Easier Than Learning The First?

Sebastian Thrun

This paper investigates learning in a lifelong context. Lifelong learning add resses situations in which a learner faces a whole stream of learn(cid :173) ing tasks. Such scenarios provide the opportunity to transfer knowledge across multiple learning tasks, in order to generalize more accurately from les

s training data. In this paper, several different approaches to lifelong learning are described, and applied in an object recognition domain. It is shown that across the board, lifelong learning approaches generalize consistently more accurately from less training data, by their ability to transfer knowledge across learning tasks.
****

Parallel analog VLSI architectures for computation of heading direction and time-to-contact
Giacomo Indiveri, Jörg Kramer, Christof Koch
We describe two parallel analog VLSI architectures that integrate optical flow data obtained from arrays of elementary velocity sen(cid:173) sors to estimate heading direction and time-to-contact. For heading direction computation, we performed simulations to evaluate the most important qualitative properties of the optical flow field and determine the best functional operators for the implementation of the architecture. For time-to-contact we exploited the divergence theorem to integrate data from all velocity sensors present in the architecture and average out possible errors.
****

A Dynamical Systems Approach for a Learnable Autonomous Robot
Jun Tani, Naohiro Fukumura
This paper discusses how a robot can learn goal-directed naviga(cid:173) tion tasks using local sensory inputs. The emphasis is that such learning tasks could be formulated as an embedding problem of dynamical systems: desired trajectories in a task space should be embedded into an adequate sensory-based internal state space so that an unique mapping from the internal state space to the motor command could be established. The paper shows that a recurrent neural network suffices in self-organizing such an adequate internal state space from the temporal sensory input. In our experiments, using a real robot with a laser range sensor, the robot navigated robustly by achieving dynamical coherence with the environment. It was also shown that such coherence becomes structurally sta(cid:173) ble as the global attractor is self-organized in the coupling of the internal and the environmental dynamics.
****

Optimization Principles for the Neural Code
Michael DeWeese
Recent experiments show that the neural codes at work in a wide range of creatures share some common features. At first sight, these observations seem unrelated. However, we show that these features arise naturally in a linear filtered threshold crossing (LFTC) model when we set the threshold to maximize the transmitted information. This maximization process requires neural adaptation to not only the DC signal level, as in conventional light and dark adaptation, but also to the statistical structure of the signal and noise distribu(cid:173) tions. We also present a new approach for calculating the mutual information between a neuron's output spike train and any aspect of its input signal which does not require reconstruction of the in(cid:173) put signal. This formulation is valid provided the correlations in the spike train are small, and we provide a procedure for checking this assumption. This paper is based on joint work (DeWeese [1], 1995). Preliminary results from the LFTC model appeared in a previous proceedings (DeWeese [2], 1995), and the conclusions we reached at that time have been reaffirmed by further analysis of the model.
****

A Framework for Non-rigid Matching and Correspondence
Suguna Pappu, Steven Gold, Anand Rangarajan
Matching feature point sets lies at the core of many approaches to object recognition. We present a framework for non-rigid match(cid:173) ing that begins with a skeleton module, affine point matching, and then integrates multiple features to improve correspondence and develops an object representation based on spatial regions to model local transformations. The algorithm for feature matching

iteratively updates the transformation parameters and the corre(cid:173) sponde nce solution, each in turn. The affine mapping is solved in closed form, which permits its use for data of any dimension. The correspondence is set via a meth od for two-way constraint satisfac(cid:173) tion, called softassign, which has r ecently emerged from the neural network/statistical physics realm. The complexi ty of the non-rigid matching algorithm with multiple features is the same as th at of the affine point matching algorithm. Results for synthetic and real worl d data are provided for point sets in 2D and 3D, and for 2D data with multiple types of features and parts.

****

Hierarchical Recurrent Neural Networks for Long-Term Dependencies

Salah Hihi, Yoshua Bengio

We have already shown that extracting long-term dependencies from se(cid:173) quential data is difficult, both for determimstic dynamical systems such as recurrent networks, and probabilistic models such as hidden Markov models (HMMs) or input/output hidden Markov models (IOHMMs). In practice , to avoid this problem, researchers have used domain specific a-prior i knowledge to give meaning to the hidden or state variables rep(cid:173) resenting past context. In this paper, we propose to use a more general type of a-priori knowledge, namely that the temporal dependencIes are structured hierarchically. This implies that long-term dependencies are r epresented by variables with a long time scale. This principle is applied to a recurrent network which includes delays and multiple time scales. Ex(cid:173) periments confirm the advantages of such structures. A similar approach i s proposed for HMMs and IOHMMs.

****

Finite State Automata that Recurrent Cascade-Correlation Cannot Represent

Stefan Kremer

This paper relates the computational power of Fahlman' s Recurrent Casca de Correlation (RCC) architecture to that of fInite state automata (FSA). Whi le some recurrent networks are FSA equivalent, RCC is not. The paper presents a theoretical analysis of the RCC architecture in the form of a proof describi ng a large class of FSA which cannot be realized by RCC.

****

Memory-based Stochastic Optimization

Andrew Moore, Jeff Schneider

In this paper we introduce new algorithms for optimizing noisy plants i n which each experiment is very expensive. The algorithms build a global non- linear model of the expected output at the same time as using Bayesian linear r egression analysis of locally weighted polynomial models. The local model ans wers queries about confi(cid:173) dence, noise, gradient and Hessians, a nd use them to make auto(cid:173) mated decisions similar to those made by a practitioner of Response Surface Methodology. The global and local models are combined naturally as a locally weighted regression. We examine th e ques(cid:173) tion of whether the global model can really help optimization, and we extend it to the case of time-varying functions. We compare th e new algorithms with a highly tuned higher-order stochastic op(cid:173) timiza tion algorithm on randomly-generated functions and a sim(cid:173) ulated ma nufacturing task. We note significant improvements in total regret , time to converge, and final solution quality.

****

Handwritten Word Recognition using Contextual Hybrid Radial Basis Function Netwo rk/Hidden Markov Models

Bernard Lemarié, Michel Gilloux, Manuel Leroux

A hybrid and contextual radial basis function networklhidden Markov model off- line handwritten word recognition system is presented. The task assigne d to the radial basis function networks is the estimation of emission p robabilities associated to Markov states. The model is contex(cid:173) tual beca use the estimation of emission probabilities takes into account the left conte xt of the current image segment as represented by its pred(cid:173) ecessor in t

he sequence. The new system does not outperform the previ(cid:173) ous system wi thout context but acts differently.
****

Universal Approximation and Learning of Trajectories Using Oscillators
Pierre Baldi, Kurt Hornik
Natural and artificial neural circuits must be capable of travers(cid:173) ing specific state space trajectories. A natural approach to this p roblem is to learn the relevant trajectories from examples. Un(cid:173) fortunately, gradient descent learning of complex trajectories in amorph ous networks is unsuccessful. We suggest a possible ap(cid:173) proach where trajectories are realized by combining simple oscil(cid:173) lators, in various modular ways. We contrast two regimes of fast and slow oscil lations. In all cases, we show that banks of oscillators with bounded frequen cies have universal approximation properties. Open questions are also discuss ed briefly.
****

Correlated Neuronal Response: Time Scales and Mechanisms
Wyeth Bair, Ehud Zohary, Christof Koch
We have analyzed the relationship between correlated spike count and the pe ak in the cross-correlation of spike trains for pairs of si(cid:173) multaneous ly recorded neurons from a previous study of area MT in the macaque monk ey (Zohary et al., 1994). We conclude that common input, responsible for creating peaks on the order of ten milliseconds wide in the spike t rain cross-correlograms (CCGs), is also responsible for creating the co rrelation in spike count ob(cid:173) served at the two second time sca le of the trial. We argue that both common excitation and inhibition ma y play significant roles in establishing this correlation.
****

Worst-case Loss Bounds for Single Neurons
David Helmbold, Jyrki Kivinen, Manfred K. K. Warmuth
We analyze and compare the well-known Gradient Descent algo(cid:173) rit hm and a new algorithm, called the Exponentiated Gradient algorithm, fo r training a single neuron with an arbitrary transfer function . Both algorithms are easily generalized to larger neural networks, and the gen eralization of Gradient Descent is the stan(cid:173) dard back-propagation algorithm. In this paper we prove worst(cid:173) case loss bounds for both algorithms in the single neuron case. Since local minima make it difficult to prove worst-case bounds for gradient-based algorithms, we must use a loss function that prevents the formation of spurious loca l minima. We define such a matching loss function for any strictly i ncreasing differentiable transfer function and prove worst-case loss bou nd for any such transfer function and its corresponding matching loss. For exam(cid:173) ple, the matching loss for the identity function is t he square loss and the matching loss for the logistic sigmoid is the entr opic loss. The different structure of the bounds for the two algorithms in di(cid:173) cates that the new algorithm out-performs Gradient Descent when t he inputs contain a large number of irrelevant components.
****

Stock Selection via Nonlinear Multi-Factor Models
Asriel Levin
This paper discusses the use of multilayer feed forward neural net(cid:173) works for predicting a stock's excess return based on its exposure to various technical and fundamental factors. To demonstrate the effectiveness of the approach a hedged portfolio which consists of equally capitalized long and short positions is constructed and its historical returns are benchmarked against T-bill returns and the S&P500 index.
****

The Geometry of Eye Rotations and Listing's Law
Amir Handzel, Tamar Flash
We analyse the geometry of eye rotations, and in particular saccades,

using basic Lie group theory and differential geome(cid:173) try. Vario us parameterizations of rotations are related through a unifying mathem atical treatment, and transformations between co-ordinate systems are com puted using the Campbell-Baker(cid:173) Hausdorff formula. Next, we descr ibe Listing's law by means of the Lie algebra so(3). This enables us to demonstrate a direct connection to Donders' law, by showing that e ye orientations are restricted to the quotient space 80(3)/80(2). The latt er is equiv(cid:173) alent to the sphere S2, which is exactly the space of gaze directions. Our analysis provides a mathematical framework for studying th e oculomotor system and could also be extended to investigate the geom etry of mUlti-joint arm movements.

****

When is an Integrate-and-fire Neuron like a Poisson Neuron?

Charles Stevens, Anthony Zador

In the Poisson neuron model, the output is a rate-modulated Pois(cid:173) so n process (Snyder and Miller, 1991); the time varying rate pa(cid:173) rameter ret) is an instantaneous function G[.] of the stimulus, ret) = G[s(t)]. In a Poisson neuron, then, ret) gives the instan(cid:173) taneous firing rate-the instantaneous probability of firing at any instant t -and the output is a stochastic function of the input. In part because of i ts great simplicity, this model is widely used (usu(cid:173) ally with the ad dition of a refractory period), especially in in vivo single unit electrop hysiological studies, where set) is usually taken to be the value of some s ensory stimulus. In the integrate-and-fire neuron model, by contrast, the o utput is a filtered and thresholded function of the input: the input is pass ed through a low-pass filter (determined by the membrane time constant T) an d integrated un(cid:173) til the membrane potential vet) reaches threshold 8, at which point vet) is reset to its initial value. By contrast with the Poi sson model, in the integrate-and-fire model the ouput is a deterministic functi on of the input. Although the integrate-and-fire model is a caricature of re al neural dynamics, it captures many of the qualitative fea(cid:173) ture s, and is often used as a starting point for conceptualizing the biophy sical behavior of single neurons. Here we show how a slightly modified Poisso n model can be derived from the integrate-and-fire model with noisy inputs yet) = set) + net). In the modified model, the transfer function G[.] i s a sigmoid (erf) whose shape is deter(cid:173) mined by the noise va riance /T~. Understanding the equivalence between the dominant in vivo and in vitro simple neuron models may help forge links between the two levels.

****

Examples of learning curves from a modified VC-formalism

Adam Kowalczyk, Jacek Szymanski, Peter Bartlett, Robert C. Williamson

We examine the issue of evaluation of model specific parameters in a mo dified VC-formalism. Two examples are analyzed: the 2-dimensional homogeneous perceptron and the I-dimensional higher order neuron. Both models are s olved theoretically, and their learning curves are com(cid:173) pared against t rue learning curves. It is shown that the formalism has the potential t o generate a variety of learning curves, including ones displaying ''pha se transitions."

****

Using Unlabeled Data for Supervised Learning

Geoffrey Towell

Many classification problems have the property that the only costly part of obt aining examples is the class label. This paper suggests a simple method for using distribution information contained in unlabeled examples to a ugment labeled examples in a supervised training framework. Empirical te sts show that the technique de(cid:173) scribed in this paper can sign ificantly improve the accuracy of a supervised learner when the learne r is well below its asymptotic accuracy level.

****

Implementation Issues in the Fourier Transform Algorithm
Yishay Mansour, Sigal Sahar
The Fourier transform of boolean functions has come to play an importan t role in proving many important learnability results. We aim to demonstrate t hat the Fourier transform techniques are also a useful and practical algo rithm in addition to being a powerful theoretical tool. We describe th e more prominent changes we have introduced to the algorithm, ones that were crucial and without which the performance of the algorithm would severely deterio(cid:173) rate. One of the benefits we present is the con fidence level for each prediction which measures the likelihood the predic tion is correct.
****

A Bound on the Error of Cross Validation Using the Approximation and Estimation Rates, with Consequences for the Training-Test Split
Michael Kearns
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Classifying Facial Action
Marian Bartlett, Paul Viola, Terrence J. Sejnowski, Beatrice Golomb, Jan Larsen, Joseph Hager, Paul Ekman
The Facial Action Coding System, (FACS), devised by Ekman and Friesen (1978), provides an objective meanS for measuring the facial muscle contractions invo lved in a facial expression. In this paper, we approach automated facial expression analysis by detecting and classifying facial actions. We gener ated a database of over 1100 image sequences of 24 subjects performing o ver 150 distinct facial actions or action combinations. We compare three different ap(cid:173) proaches to classifying the facial actions in these images: Holistic spatial analysis based on principal components of graylevel images; explicit measurement of local image features such as wrinkles; and template matching with motion flow fields. On a dataset contain(cid:173) i ng six individual actions and 20 subjects, these methods had 89%, 57%, and 85% performances respectively for generalization to novel subjects. When combined, performance improved to 92%.
****

Strong Unimodality and Exact Learning of Constant Depth μ-Perceptron Networks
Mario Marchand, Saeed Hadjifaradji
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Gradient and Hamiltonian Dynamics Applied to Learning in Neural Networks
James Howse, Chaouki Abdallah, Gregory Heileman
The process of machine learning can be considered in two stages: model selection and parameter estimation. In this paper a technique is present ed for constructing dynamical systems with desired qualitative properties. The approach is based on the fact that an n-dimensional nonlinear dynamical system can be decomposed into one gradient and (n - 1) Hamiltonian sys(cid:173) tems. Thus, the model selection stage consists of choosing the gradient an d Hamiltonian portions appropriately so that a certain behavior is obtainable. To estimate the parameters, a stably convergent learning rule is presented. This algorithm has been proven to converge to the desired system trajectory for all initial conditions and system inputs. This technique can be used to d esign neural network models which are guaranteed to solve the trajectory lea rning problem.
****

A New Learning Algorithm for Blind Signal Separation

Shun-ichi Amari, Andrzej Cichocki, Howard Yang
A new on-line learning algorithm which minimizes a statistical de(cid:173) pend ency among outputs is derived for blind separation of mixed signals. The d ependency is measured by the average mutual in(cid:173) formation (MI) of the outputs. The source signals and the mixing matrix are unknown except for the number of the sources. The Gram-Charlier expansion instead o f the Edgeworth expansion is used in evaluating the MI. The natural gr adient approach is used to minimize the MI. A novel activation function is pr oposed for the on-line learning algorithm which has an equivariant property and is easily implemented on a neural network like model. The validity of th e new learning algorithm are verified by computer simulations.
****

Adaptive Back-Propagation in On-Line Learning of Multilayer Networks
Ansgar West, David Saad
An adaptive back-propagation algorithm is studied and compared with gradien t descent (standard back-propagation) for on-line learning in two-layer neural networks with an arbitrary number of hidden units. Within a st atistical mechanics framework , both numerical studies and a rigorous a nalysis show that the adaptive back-propagation method results in faster tr aining by breaking the symmetry between hidden units more efficiently and b y providing faster convergence to optimal generalization than gradient descen t.
****

Neuron-MOS Temporal Winner Search Hardware for Fully-Parallel Data Processing
Tadashi Shibata, Tsutomu Nakai, Tatsuo Morimoto, Ryu Kaihara, Takeo Yamashita, T adahiro Ohmi
A unique architecture of winner search hardware has been de(cid:173) ve loped using a novel neuron-like high functionality device called Neuron MOS transistor (or vMOS in short) [1,2] as a key circuit element. The circuits developed in this work can find the location of the maximum (or minimum) signal among a number of input data on the continuous-time basis, thus enabling real-time winner tracking as well as fully-parallel so rting of multiple input data. We have developed two circuit schemes. One is an ensemble of self(cid:173) loop-selecting v M OS ring oscillators findi ng the winner as an oscil(cid:173) lating node. The other is an ensemble of vM OS variable threshold inverters receiving a common ramp-voltage for competitiv e excita(cid:173) tion where data sorting is conducted through consecutive winner search actions. Test circuits were fabricated by a double-polysilicon CMOS process and their operation has been experimentally veri(cid:173) fied.
****

Human Reading and the Curse of Dimensionality
Gale Martin
Whereas optical character recognition (OCR) systems learn to clas(cid:173) sify single characters; people learn to classify long character strings in paralle l, within a single fixation . This difference is surprising because hi gh dimensionality is associated with poor classification learning. This paper suggests that the human reading system avoids these problems bec ause the number of to-be-classified im(cid:173) ages is reduced by consi stent and optimal eye fixation positions, and by character sequence regu larities.
****

Unsupervised Pixel-prediction
William Softky
When a sensory system constructs a model of the environment from its input, it might need to verify the model's accuracy. One method of verification is multiv ariate time-series prediction: a good model could predict the near-future activ ity of its inputs, much as a good scientific theory predicts future data. Such a predict(cid:173) ing model would require copious top-down connections to compa re the predictions with the input. That feedback could improve the model's per

formance in two ways: by biasing internal activity to(cid:173) ward expected pat terns, and by generating specific error signals if the predictions fail. A proo f-of-concept model-an event-driven, computationally efficient layered network, incorporating "cortical" features like all-excitatory synapses and local inhibi tion- was con(cid:173) structed to make near-future predictions of a simple, mov ing stim(cid:173) ulus. After unsupervised learning, the network contained units not only tuned to obvious features of the stimulus like contour orienta(cid:17 3) tion and motion, but also to contour discontinuity ("end-stopping") and illu sory contours.

****

Control of Selective Visual Attention: Modeling the "Where" Pathway

Ernst Niebur, Christof Koch

Intermediate and higher vision processes require selection of a sub(cid:173) se t of the available sensory information before further processing. Usually , this selection is implemented in the form of a spatially circumscrib ed region of the visual field, the so-called "focus of at(cid:173) tention " which scans the visual scene dependent on the input and on the attent ional state of the subject. We here present a model for the control of the f ocus of attention in primates, based on a saliency map. This mechanism is not only expected to model the functional(cid:173) ity of biological vision but als o to be essential for the understanding of complex scenes in machine vision.

****

Quadratic-Type Lyapunov Functions for Competitive Neural Networks with Different Time-Scales

Anke Meyer-Bäse

The dynamics of complex neural networks modelling the self(cid:173) organization process in cortical maps must include the aspects of long and short-term memor y. The behaviour of the network is such characterized by an equation of neural activity as a fast phenom(cid:173) enon and an equation of synaptic modification as a slow part of the neural system. We present a quadratic-type Lyapunov func tion for the flow of a competitive neural system with fast and slow dynamic va riables. We also show the consequences of the stability analysis on the neural net parameters.

****

Learning long-term dependencies is not as difficult with NARX networks

Tsungnan Lin, Bill Horne, Peter Tiño, C. Giles

It has recently been shown that gradient descent learning algo(cid:173) rithms for recurrent neural networks can perform poorly on tasks that i nvolve long-term dependencies. In this paper we explore this problem f or a class of architectures called NARX networks, which have powerful representational capabilities. Previous work reported that gradient descent l earning is more effective in NARX networks than in recurrent networks wit h "hidden states". We show that although NARX networks do not circumve nt the prob(cid:173) lem of long-term dependencies, they can greatly impro ve perfor(cid:173) mance on such problems. We present some experimental 'results that show that NARX networks can often retain information for two to three times as long as conventional recurrent networks.

****

Learning the Structure of Similarity

Joshua Tenenbaum

The additive clustering (ADCL US) model (Shepard & Arabie, 1979) treats t he similarity of two stimuli as a weighted additive measure of their co mmon features. Inspired by recent work in unsupervised learning with mul tiple cause models, we propose anew, statistically well-motivated algorith m for discovering the structure of natural stimulus classes using the AD CLUS model, which promises substan(cid:173) tial gains in conceptual simplici ty, practical efficiency, and solution quality over earlier efforts. We a lso present preliminary results with artificial data and two classic similari ty data sets.

****

Investment Learning with Hierarchical PSOMs
Jörg Walter, Helge Ritter
We propose a hierarchical scheme for rapid learning of context dependent "skills" that is based on the recently introduced "Parameterized Self(cid:173) Organizing Map" ("PSOM"). The underlying idea is to first invest some learning effort to specialize the system into a rapid learner for a more restricted range of contexts.
****

How Perception Guides Production in Birdsong Learning
Christopher Fry
A c.:omputational model of song learning in the song sparrow (M elospiza melodia) learns to categorize the different syllables of a song sparrow song and uses this categorization to train itself to reproduce song. The model fills a crucial gap in the computational explanation of birdsong learning by exploring the organization of perception in songbirds. It shows how competitive learning may lead to the organization of a specific nucleus in the bird brain, replicates the song production results of a previous model (Doya and Sejnowski, 1995), and demonstrates how perceptual learning can guide production through reinforcement learning.
****

Stable Fitted Reinforcement Learning
Geoffrey J. Gordon
We describe the reinforcement learning problem, motivate algo(cid:173) rithms which seek an approximation to the Q function, and present new convergence results for two such algorithms.
****

Information through a Spiking Neuron
Charles Stevens, Anthony Zador
While it is generally agreed that neurons transmit information about their synaptic inputs through spike trains, the code by which this information is transmitted is not well understood. An upper bound on the information encoded is obtained by hypothesizing that the precise timing of each spike conveys information. Here we develop a general approach to quantifying the information carried by spike trains under this hypothesis, and apply it to the leaky integrate-and-fire (IF) model of neuronal dynamics. We formu(cid:173) late the problem in terms of the probability distribution peT) of interspike intervals (ISIs), assuming that spikes are detected with arbitrary but finite temporal resolution. In the absence of added noise, all the variability in the ISIs could encode information, and the information rate is simply the entropy of the lSI distribution, H (T) = (-p(T) log2 p(T)}, times the spike rate. H (T) thus pro(cid:173) vides an exact expression for the information rate. The methods developed here can be used to determine experimentally the infor(cid:173) mation carried by spike trains, even when the lower bound of the information rate provided by the stimulus reconstruction method is not tight. In a preliminary series of experiments, we have used these methods to estimate information rates of hippocampal neu(cid:173) rons in slice in response to somatic current injection. These pilot experiments suggest information rates as high as 6.3 bits/spike.
****

Discriminant Adaptive Nearest Neighbor Classification and Regression
Trevor Hastie, Robert Tibshirani
Nearest neighbor classification expects the class conditional prob(cid:173) abilities to be locally constant, and suffers from bias in high di(cid:173) mensions We propose a locally adaptive form of nearest neighbor classification to try to finesse this curse of dimensionality. We use a local linear discriminant analysis to estimate an effective met(cid:173) ric for computing neighborhoods. We determine the local decision boundaries from centroid information, and then shrink neighbor(cid:173) hoods in directions orthogonal to these local decision boundaries, and elongate

them parallel to the boundaries. Thereafter, any neighborhood-based classifier can be employed, using the modified neighborhoods. We also propose a method for global dimension reduction, that combines local dimension information. We indicate how these techniques can be extended to the regression problem.
****

A Predictive Switching Model of Cerebellar Movement Control
Andrew Barto, James Houk
We present a hypothesis about how the cerebellum could partici(cid:173) pate in regulating movement in the presence of significant feedback delays without resorting to a forward model of the motor plant. We show how a simplified cerebellar model can learn to control end(cid:173) point positioning of a nonlinear spring-mass system with realistic delays in both afferent and efferent pathways. The model's opera(cid:173) tion involves prediction, but instead of predicting sensory input, it directly regulates movement by reacting in an anticipatory fashion to input patterns that include delayed sensory feedback.
****

Recurrent Neural Networks for Missing or Asynchronous Data
Yoshua Bengio, Francois Gingras
In this paper we propose recurrent neural networks with feedback into the input units for handling two types of data analysis problems. On the one hand, this scheme can be used for static data when some of the input variables are missing. On the other hand, it can also be used for sequential data, when some of the input variables are missing or are available at different frequencies. Unlike in the case of probabilistic models (e.g. Gaussian) of the missing variables, the network does not attempt to model the distribution of the missmg variables given the observed variables. Instead it is a more "discriminant" approach that fills in the missing variables for the sole purpose of minimizing a learning criterion (e.g., to minimize an output error).
****