

English Alphabet Recognition with Telephone Speech

Mark Fanty, Ronald Cole, Krist Roginski

A recognition system is reported which recognizes names spelled over the telephone with brief pauses between letters. The system uses separate neural networks to locate segment boundaries and classify letters. The letter scores are then used to search a database of names to find the best scoring name. The speaker-independent classification rate for spoken letters is 89%. The system retrieves the correct name, spelled with pauses between letters, 91 % of the time from a database of 50,000 names.

Recognizing Overlapping Hand-Printed Characters by Centered-Object Integrated Segmentation and Recognition

Gale L. Martin, Mosfeq Rashid

This paper describes an approach, called centered object integrated segmentation and recognition (COISR), for integrating object segmentation and recognition within a single neural network. The application is hand-printed character recognition. Two versions of the system are described. One uses a backpropagation network that scans exhaustively over a field of characters and is trained to recognize whether it is centered over a single character or between characters. When it is centered over a character, the net classifies the character. The approach is tested on a dataset of hand-printed digits. Very low error rates are reported. The second version, COISR-SACCADE, avoids the need for exhaustive scans. The net is trained as before, but also is trained to compute ballistic 'eye' movements that enable the input window to jump from one character to the next.

Adaptive Soft Weight Tying using Gaussian Mixtures

Steven Nowlan, Geoffrey E. Hinton

One way of simplifying neural networks so they generalize better is to add an extra term to the error function that will penalize complexity. We propose a new penalty term in which the distribution of weight values is modelled as a mixture of multiple Gaussians. Under this model, a set of weights is simple if the weights can be clustered into subsets so that the weights in each cluster have similar values. We allow the parameters of the mixture model to adapt at the same time as the network learns. Simulations demonstrate that this complexity term is more effective than previous complexity terms.

Multi-State Time Delay Networks for Continuous Speech Recognition

Patrick Haffner, Alex Waibel

We present the "Multi-State Time Delay Neural Network" (MS-TDNN) as an extension of the TDNN to robust word recognition. Unlike most other hybrid methods, the MS-TDNN embeds an alignment search procedure into the connectionist architecture, and allows for word level supervision. The resulting system has the ability to manage the sequential order of subword units, while optimizing for the recognizer performance. In this paper we present extensive new evaluations of this approach over speaker-dependent and speaker-independent connected alphabet.

Neural Network - Gaussian Mixture Hybrid for Speech Recognition or Density Estimation

Yoshua Bengio, Renato De Mori, Giovanni Flammia, Ralf Kompe

The subject of this paper is the integration of multi-layered Artificial Neural Networks (ANN) with probability density functions such as Gaussian mixtures found in continuous density Hidden Markov Models (HMM). In the first part of this paper we present an ANN/HMM hybrid in which all the parameters of the system are simultaneously optimized with respect to a single criterion. In the second part of this paper, we study the relationship between the density of the inputs of the network and the density of the outputs of the networks. A few experiments are presented to explore how to perform density estimation with

ANNs.

Threshold Network Learning in the Presence of Equivalences

John Shawe-Taylor

This paper applies the theory of Probably Approximately Correct (PAC) learning to multiple output feedforward threshold networks in which the weights conform to certain equivalences. It is shown that the sample size for reliable learning can be bounded above by a formula similar to that required for single output networks with no equivalences. The best previously obtained bounds are improved for all cases.

Node Splitting: A Constructive Algorithm for Feed-Forward Neural Networks

Mike Wynne-Jones

A constructive algorithm is proposed for feed-forward neural networks, which uses node-splitting in the hidden layers to build large networks from smaller ones. The small network forms an approximate model of a set of training data, and the split creates a larger more powerful network which is initialised with the approximate solution already found. The insufficiency of the smaller network in modelling the system which generated the data leads to oscillation in those hidden nodes whose weight vectors cover regions in the input space where more detail is required in the model. These nodes are identified and split in two using principal component analysis, allowing the new nodes to cover the two main modes of each oscillating vector. Nodes are selected for splitting using principal component analysis on the oscillating weight vectors, or by examining the Hessian matrix of second derivatives of the network error with respect to the weights. The second derivative method can also be applied to the input layer, where it provides a useful indication of the relative importances of parameters for the classification task. Node splitting in a standard Multi Layer Perceptron is equivalent to introducing a hinge in the decision boundary to allow more detail to be learned. Initial results were promising, but further evaluation indicates that the long range effects of decision boundaries cause the new nodes to slip back to the old node position, and nothing is gained. This problem does not occur in networks of localised receptive fields such as radial basis functions or gaussian mixtures, where the technique appears to work well.

Structural Risk Minimization for Character Recognition

I. Guyon, V. Vapnik, B. Boser, L. Bottou, S. A. Solla

The method of Structural Risk Minimization refers to tuning the capacity of the classifier to the available amount of training data. This capacity is influenced by several factors, including: (1) properties of the input space, (2) nature and structure of the classifier, and (3) learning algorithm. Actions based on these three factors are combined here to control the capacity of linear classifiers and improve generalization on the problem of handwritten digit recognition.

Computer Recognition of Wave Location in Graphical Data by a Neural Network

Donald Freeman

Five experiments were performed using several neural network architectures to identify the location of a wave in the time ordered graphical results from a medical test. Baseline results from the first experiment found correct identification of the target wave in 85% of cases (n=20). Other experiments investigated the effect of different architectures and preprocessing the raw data on the results. The methods used seem most appropriate for time oriented graphical data which has a clear starting point such as electrophoresis or spectrometry rather than continuous tests such as ECGs and EEGs.

Improved Hidden Markov Model Speech Recognition Using Radial Basis Function Networks

Elliot Singer, Richard P. Lippmann

A high performance speaker-independent isolated-word hybrid speech recognizer was developed which combines Hidden Markov Models (HMMs) and Radial Basis Function (RBF) neural networks. In recognition experiments using a speaker-independent E-set database, the hybrid recognizer had an error rate of 11.5% compared to 15.7% for the robust unimodal Gaussian HMM recognizer upon which the hybrid system was based. These results and additional experiments demonstrate that RBF networks can be successfully incorporated in hybrid recognizers and suggest that they may be capable of good performance with fewer parameters than required by Gaussian mixture classifiers. A global parameter optimization method designed to minimize the overall word error rather than the frame recognition error failed to reduce the error rate.

Connectionist Optimisation of Tied Mixture Hidden Markov Models

Steve Renals, Nelson Morgan, Hervé Bourlard, Horacio Franco, Michael Cohen

Issues relating to the estimation of hidden Markov model (HMM) local probabilities are discussed. In particular we note the isomorphism of radial basis functions (RBF) networks to tied mixture density modelling; additionally we highlight the differences between these methods arising from the different training criteria employed. We present a method in which connectionist training can be modified to resolve these differences and discuss some preliminary experiments. Finally, we discuss some outstanding problems with discriminative training.

Principled Architecture Selection for Neural Networks: Application to Corporate Bond Rating Prediction

John Moody, Joachim Utans

The notion of generalization ability can be defined precisely as the prediction risk, the expected performance of an estimator in predicting new observations. In this paper, we propose the prediction risk as a measure of the generalization ability of multi-layer perceptron networks and use it to select an optimal network architecture from a set of possible architectures. We also propose a heuristic search strategy to explore the space of possible architectures. The prediction risk is estimated from the available data; here we estimate the prediction risk by v -fold cross-validation and by asymptotic approximations of generalized cross-validation or Akaike's final prediction error. We apply the technique to the problem of predicting corporate bond ratings. This problem is very attractive as a case study, since it is characterized by the limited availability of the data and by the lack of a complete a priori model which could be used to impose a structure to the network architecture.

Extracting and Learning an Unknown Grammar with Recurrent Neural Networks

C. L. Giles, C. B. Miller, D. Chen, G. Z. Sun, H. H. Chen, Y. C. Lee

Simple second-order recurrent networks are shown to readily learn simple regular grammars when trained with positive and negative strings examples. We show that similar methods are appropriate for learning unknown grammars from examples of their strings. The training algorithm is an incremental real-time, recurrent learning (RTRL) method that computes the complete gradient and updates the weights at the end of each string. After or during training, a dynamic clustering algorithm extracts the production rules that the neural network has learned. The methods are illustrated by extracting rules from unknown deterministic regular grammars. For many cases the extracted grammar outperforms the neural net from which it was extracted in correctly classifying unseen strings.

Network activity determines spatio-temporal integration in single cells

Öjvind Bernander, Christof Koch, Rodney Douglas

Single nerve cells with static properties have traditionally been viewed as the building blocks for networks that show emergent phenomena. In contrast to this approach, we study here how the overall network activity can control single cell parameters such as input resistance, as well as time and space constants, pa

parameters that are crucial for excitability and spatio(cid:173) temporal integration. Using detailed computer simulations of neocortical pyramidal cells, we show that the spontaneous background firing of the network provides a means for setting these parameters. The mechanism for this control is through the large conductance change of the membrane that is induced by both non-NMDA and NMDA excitatory and inhibitory synapses activated by the spontaneous background activity.

Image Segmentation with Networks of Variable Scales

Hans Graf, Craig Nohl, Jan Ben

We developed a neural net architecture for segmenting complex images, i.e., to localize two-dimensional geometrical shapes in a scene, without prior knowledge of the objects' positions and sizes. A scale variation is built into the network to deal with varying sizes. This algo(cid:173) rithm has been applied to video images of railroad cars, to find their identification numbers. Over 95% of the characters were located correctly in a data base of 300 images, despite a large variation in light(cid:173) ing conditions and often a poor quality of the characters. A part of the network is executed on a processor board containing an analog neural net chip (Graf et al. 1991). while the rest is implemented as a software model on a workstation or a digital signal processor.

Decoding of Neuronal Signals in Visual Pattern Recognition

Emad Eskandar, Barry Richmond, John Hertz, Lance Optican, Troels Kjær

We have investigated the properties of neurons in inferior temporal (IT) cortex in monkeys performing a pattern matching task. Simple back(cid:173) :173) propagation networks were trained to discriminate the various stimulus conditions on the basis of the measured neuronal signal. We also trained networks to predict the neuronal response waveforms from the spatial pat(cid:173) :173) terns of the stimuli. The results indicate that IT neurons convey tempo(cid:173) :173) rally encoded information about both current and remembered patterns, as well as about their behavioral context.

Analog LSI Implementation of an Auto-Adaptive Network for Real-Time Separation of Independent Signals

Marc Cohen, Philippe Pouliquen, Andreas Andreou

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

Learning in Feedforward Networks with Nonsmooth Functions

Nicholas Redding, T. Downs

This paper is concerned with the problem of learning in networks where some or all of the functions involved are not smooth. Examples of such networks are those whose neural transfer functions are piecewise-linear and those whose error function is defined in terms of the l_1 norm. Up to now, networks whose neural transfer functions are piecewise-linear have received very little consideration in the literature, but the possibility of using an error function defined in terms of the l_1 norm has received some attention. In this latter work, however, the problems that can occur when gradient methods are used for non smooth error functions have not been addressed. In this paper we draw upon some recent results from the field of nonsmooth optimization (NSO) to present an algorithm for the non smooth case. Our motivation for this work arose out of the fact that we have been able to show that, in backpropagation, an error function based upon the l_1 norm overcomes the difficulties which can occur when using the l_2 norm.

Constructing Proofs in Symmetric Networks

Gadi Pinkus

This paper considers the problem of expressing predicate calculus in con(cid:173) :173)

Connectionist networks that are based on energy minimization. Given a first-order logic knowledge base and a bound k , a symmetric network is constructed (like a Boltzman machine or a Hopfield network) that searches for a proof for a given query. If a resolution-based proof of length no longer than k exists, then the global minima of the energy function that is associated with the network represent such proofs. The network that is generated is of size cubic in the bound k and linear in the knowledge size. There are no restrictions on the type of logic formulas that can be represented. The network is inherently fault tolerant and can cope with inconsistency and nonmonotonicity.

A Weighted Probabilistic Neural Network

David Montana

The Probabilistic Neural Network (PNN) algorithm represents the likelihood function of a given class as the sum of identical, isotropic Gaussians. In practice, PNN is often an excellent pattern classifier, outperforming other classifiers including backpropagation. However, it is not robust with respect to affine transformations of feature space, and this can lead to poor performance on certain data. We have derived an extension of PNN called Weighted PNN (WPNN) which compensates for this flaw by allowing anisotropic Gaussians, i.e. Gaussians whose covariance is not a multiple of the identity matrix. The covariance is optimized using a genetic algorithm, some interesting features of which are its redundancy, logarithmic encoding and large population size. Experimental results validate our claims.

Shooting Craps in Search of an Optimal Strategy for Training Connectionist Pattern Classifiers

J. B. Hampshire II, B. Kumar

We compare two strategies for training connectionist (as well as nonconnectionist) models for statistical pattern recognition. The probabilistic strategy is based on the notion that Bayesian discrimination (i.e. optimal classification) is achieved when the classifier learns the a posteriori class distributions of the random feature vector. The differential strategy is based on the notion that the identity of the largest class a posteriori probability of the feature vector is all that is needed to achieve Bayesian discrimination. Each strategy is directly linked to a family of objective functions that can be used in the supervised training procedure. We prove that the probabilistic strategy - linked with error measure objective functions such as mean-squared-error and cross-entropy - typically used to train classifiers necessarily requires larger training sets and more complex classifier architectures than those needed to approximate the Bayesian discriminant function.

In contrast, we prove that the differential strategy - with classification figure-of-merit objective functions (CF-MmIo) [3] - requires the minimum classifier functional complexity and the fewest training examples necessary to approximate the Bayesian discriminant function with specified precision (measured in probability of error). We present our proofs in the context of a game of chance in which an unfair C -sided die is tossed repeatedly. We show that this rigged game of dice is a paradigm at the root of all statistical pattern recognition tasks, and demonstrate how a simple extension of the concept leads us to a general information-theoretic model of sample complexity for statistical pattern recognition.

Multimodular Architecture for Remote Sensing Operations.

Sylvie Thiria, Carlos Mejia, Fouad Badran, Michel Crépon

This paper deals with an application of Neural Networks to satellite remote sensing observations. Because of the complexity of the application and the large amount of data, the problem cannot be solved by using a single method. The solution we propose is to build multimodular NN architectures where several NN cooperate together. Such system suffer from generic problem for whom we propose solutions. They allow to reach accurate performances for multi-valued func

tion approximations and probability estimations. The results are compared with six other methods which have been used for this problem. We show that the methodology we have developed is general and can be used for a large variety of applications.

Learning in the Vestibular System: Simulations of Vestibular Compensation Using Recurrent Back-Propagation

Thomas Anastasio

Vestibular compensation is the process whereby normal functioning is regained following destruction of one member of the pair of peripheral vestibular receptors. Compensation was simulated by lesioning a dynamic neural network model of the vestibulo-ocular reflex (VOR) and retraining it using recurrent back-propagation. The model reproduced the pattern of VOR neuron activity experimentally observed in compensated animals, but only if connections heretofore considered uninvolved were allowed to be plastic. Because the model incorporated nonlinear units, it was able to reconcile previously conflicting, linear analyses of experimental results on the dynamic properties of VOR neurons in normal and compensated animals.

Repeat Until Bored: A Pattern Selection Strategy

Paul Munro

An alternative to the typical technique of selecting training examples independently from a fixed distribution is formulated and analyzed, in which the current example is presented repeatedly until the error for that item is reduced to some criterion value, ϵ ; then, another item is randomly selected.

The convergence time can be dramatically increased or decreased by this heuristic, depending on the task, and is very sensitive to the value of ϵ .

Refining PID Controllers using Neural Networks

Gary Scott, Jude Shavlik, W. Ray

The KBANN approach uses neural networks to refine knowledge that can be written in the form of simple propositional rules. We extend this idea further by presenting the MANNCON algorithm by which the mathematical equations governing a PID controller determine the topology and initial weights of a network, which is further trained using backpropagation. We apply this method to the task of controlling the outflow and temperature of a water tank, producing statistically-significant gains in accuracy over both a standard neural network approach and a non-learning PID controller. Furthermore, using the PID knowledge to initialize the weights of the network produces statistically less variation in test set accuracy when compared to networks initialized with small random numbers.

A Cortico-Cerebellar Model that Learns to Generate Distributed Motor Commands to Control a Kinematic Arm

N. Berthier, S. P. Singh, A. G. Barto, J. C. Houk

A neurophysiologically-based model is presented that controls a simulated kinematic arm during goal-directed reaches. The network generates a quasi-feedforward motor command that is learned using training signals generated by corrective movements. For each target, the network selects and sets the output of a subset of pattern generators. During the movement, feedback from proprioceptors turns off the pattern generators. The task facing individual pattern generators is to recognize when the arm reaches the target and to turn off. A distributed representation of the motor command that resembles population vectors seen in vivo was produced naturally by these simulations.

Linear Operator for Object Recognition

Ronen Basri, Shimon Ullman

Visual object recognition involves the identification of images of 3-D objects seen from arbitrary viewpoints. We suggest an approach to object recognition in which a view is represented as a collection of

points given by their location in the image. An object is modeled by a set of 2-D views together with the correspondence between the views. We show that any novel view of the object can be expressed as a linear combination of the stored views. Consequently, we build a linear operator that distinguishes between views of a specific object and views of other objects. This operator can be implemented using neural network architectures with relatively simple structures.

Learning Unambiguous Reduced Sequence Descriptions

Jürgen Schmidhuber

Do you want your neural net algorithm to learn sequences? Do not limit yourself to conventional gradient descent (or approximations thereof). Instead, use your sequence learning algorithm (any will do) to implement the following method for history compression. No matter what your final goals are, train a network to predict its next input from the previous ones. Since only unpredictable inputs convey new information, ignore all predictable inputs but let all unexpected inputs (plus information about the time step at which they occurred) become inputs to a higher-level network of the same kind (working on a slower, self-adjusting time scale). Go on building a hierarchy of such networks. This principle reduces the descriptions of event sequences without loss of information, thus easing supervised or reinforcement learning tasks. Alternatively, you may use two recurrent networks to collapse a multi-level predictor hierarchy into a single recurrent net. Experiments show that systems based on these principles can require less computation per time step and many fewer training sequences than conventional training algorithms for recurrent nets. Finally you can modify the above method such that predictability is not defined in a yes-or-no fashion but in a continuous fashion.

Dynamically-Adaptive Winner-Take-All Networks

Trent Lange

Winner-Take-All (WTA) networks, in which inhibitory interconnections are used to determine the most highly-activated of a pool of units, are an important part of many neural network models. Unfortunately, convergence of normal WTA networks is extremely sensitive to the magnitudes of their weights, which must be hand-tuned and which generally only provide the right amount of inhibition across a relatively small range of initial conditions. This paper presents Dynamically-Adaptive Winner-Take-All (DA WTA) networks.

Some Approximation Properties of Projection Pursuit Learning Networks

Ying Zhao, Christopher Atkeson

This paper will address an important question in machine learning: What kind of network architectures work better on what kind of problems? A projection pursuit learning network has a very similar structure to a one hidden layer sigmoidal neural network. A general method based on a continuous version of projection pursuit regression is developed to show that projection pursuit regression works better on angular smooth functions than on Laplacian smooth functions. There exists a ridge function approximation scheme to avoid the curse of dimensionality for approximating functions in $L_2(\mathbb{R}^d)$.

Unsupervised learning of distributions on binary vectors using two layer networks

Yoav Freund, David Haussler

We study a particular type of Boltzmann machine with a bipartite graph structure called a harmonic network. Our interest is in using such a machine to model a probability distribution on binary input vectors. We analyze the class of pro

bability distributions that can be modeled by such machines. showing that for each $n \sim 1$ this class includes arbitrarily good approximations to any distribution on the set of all n -vectors of binary inputs. We then present two learning algorithms for these machines. The first learning algorithm is the standard gradient ascent heuristic for computing maximum likelihood estimates for the parameters (i.e. weights and thresholds) of the model. Here we give a closed form for this gradient that is significantly easier to compute than the corresponding gradient for the general Boltzmann machine. The second learning algorithm is a greedy method that creates the hidden units and computes their weights one at a time. This method is a variant of the standard method for projection pursuit density estimation. We give experimental results for these learning methods on synthetic data and natural data from the domain of handwritten digits.

Best-First Model Merging for Dynamic Learning and Recognition

Stephen Omohundro

"Best-first model merging" is a general technique for dynamically choosing the structure of a neural or related architecture while avoiding overfitting. It is applicable to both learning and recognition tasks and often generalizes significantly better than fixed structures. We demonstrate the approach applied to the tasks of choosing radial basis functions for function learning, choosing local affine models for curve and constraint surface modelling, and choosing the structure of a balltree or bumptree to maximize efficiency of access.

Neural Computing with Small Weights

Kai-Yeung Siu, Jehoshua Bruck

An important issue in neural computation is the dynamic range of weights in the neural networks. Many experimental results on learning indicate that the weights in the networks can grow prohibitively large with the size of the inputs. Here we address this issue by studying the tradeoffs between the depth and the size of weights in polynomial-size networks of linear threshold elements (LTEs). We show that there is an efficient way of simulating a network of LTEs with large weights by a network of LTEs with small weights. In particular, we prove that every depth- d , polynomial-size network of LTEs with exponentially large integer weights can be simulated by a depth- $(2d + 1)$, polynomial-size network of LTEs with polynomially bounded integer weights. To prove these results, we use tools from harmonic analysis of Boolean functions. Our technique is quite general, it provides insights to some other problems. For example, we are able to improve the best known results on the depth of a network of linear threshold elements that computes the COMPARISON, SUM and PRODUCT of two n -bits numbers, and the MAXIMUM and the SORTING of n n -bit numbers.

A Segment-Based Automatic Language Identification System

Yeshwant Muthusamy, Ronald Cole

We have developed a four-language automatic language identification system for high-quality speech. The system uses a neural network-based segmentation algorithm to segment speech into seven broad phonetic categories. Phonetic and prosodic features computed on these categories are then input to a second network that performs the language classification. The system was trained and tested on separate sets of speakers of American English, Japanese, Mandarin Chinese and Tamil. It currently performs with an accuracy of 89.5% on the utterances of the test set.

A Topographic Product for the Optimization of Self-Organizing Feature Maps

Hans-Ulrich Bauer, Klaus Pawelzik, Theo Geisel

Optimizing the performance of self-organizing feature maps like the Kohonen map involves the choice of the output space topology. We present a topographic product which measures the preservation of neighborhood relations as a cr

iterion to optimize the output space topology of the map with regard to the global dimensionality DA as well as to the dimensions in the individual directions. We test the topographic product method not only on synthetic mapping examples, but also on speech data. In the latter application our method suggests an output space dimensionality of $DA = 3$, in coincidence with recent recognition results on the same data set.

Learning How to Teach or Selecting Minimal Surface Data

Davi Geiger, Ricardo Pereira

Learning a map from an input set to an output set is similar to the problem of reconstructing hypersurfaces from sparse data (Poggio and Girosi, 1990). In this framework, we discuss the problem of automatically selecting "minimal" surface data. The objective is to be able to approximately reconstruct the surface from the selected sparse data. We show that this problem is equivalent to the one of compressing information by data removal and the one of learning how to teach. Our key step is to introduce a process that statistically selects the data according to the model. During the process of data selection (learning how to teach) our system (teacher) is capable of predicting the new surface, the approximated one provided by the selected data. We concentrate on piecewise smooth surfaces, e.g. images, and use mean field techniques to obtain a deterministic network that is shown to compress image data.

Visual Grammars and Their Neural Nets

Eric Mjølness

I exhibit a systematic way to derive neural nets for vision problems. It involves formulating a vision problem as Bayesian inference or decision on a comprehensive model of the visual domain given by a probabilistic grammar.

Hierarchical Transformation of Space in the Visual System

Alexandre Pouget, Stephen Fisher, Terrence J. Sejnowski

Neurons encoding simple visual features in area VI such as orientation, direction of motion and color are organized in retinotopic maps. However, recent physiological experiments have shown that the responses of many neurons in VI and other cortical areas are modulated by the direction of gaze. We have developed a neural network model of the visual cortex to explore the hypothesis that visual features are encoded in head-centered coordinates at early stages of visual processing. New experiments are suggested for testing this hypothesis using electrical stimulations and psychophysical observations.

Illumination and View Position in 3D Visual Recognition

Amnon Shashua

It is shown that both changes in viewing position and illumination conditions can be compensated for, prior to recognition, using combinations of images taken from different viewing positions and different illumination conditions. It is also shown that, in agreement with psychophysical findings, the computation requires at least a sign-bit image as input - contours alone are not sufficient.

Constant-Time Loading of Shallow 1-Dimensional Networks

Stephen Judd

The complexity of learning in shallow 1-Dimensional neural networks has been shown elsewhere to be linear in the size of the network. However, when the network has a huge number of units (as cortex has) even linear time might be unacceptable. Furthermore, the algorithm that was given to achieve this time was based on a single serial processor and was biologically implausible. In this work we consider the more natural parallel model of processing and demonstrate an expected-time complexity that is constant (i.e. dependent of the size of the network).

k). This holds even when inter-node communication channels are short and local, thus adhering to more bio(cid:173) logical and VLSI constraints.

Combined Neural Network and Rule-Based Framework for Probabilistic Pattern Recognition and Discovery

Hayit K. Greenspan, Rodney Goodman, Rama Chellappa

A combined neural network and rule-based approach is suggested as a general framework for pattern recognition. This approach enables unsu(cid:173) pervised and supervised learning, respectively, while providing probability estimates for the output classes. The probability maps are utilized for higher level analysis such as a feedback for smoothing over the output la(cid:173) bel maps and the identification of unknown patterns (pattern "discovery"). The suggested approach is presented and demonstrated in the texture - analysis task. A correct classification rate in the 90 percentile is achieved for both unstructured and structured natural texture mosaics. The advan(cid:173) tages of the probabilistic approach to pattern analysis are demonstrated.

Experimental Evaluation of Learning in a Neural Microsystem

Joshua Alspector, Anthony Jayakumar, Stephan Luna

We report learning measurements from a system composed of a cascadable learning chip, data generators and analyzers for training pattern presentation, and an X-windows based software interface. The 32 neuron learning chip has 496 adaptive synapses and can perform Boltzmann and mean-field learning using separate noise and gain controls. We have used this system to do learning experiments on the parity and replication problem. The system settling time limits the learning speed to about 100,000 patterns per second roughly independent of system size.

Recurrent Eye Tracking Network Using a Distributed Representation of Image Motion

Paul Viola, Stephen Lisberger, Terrence J. Sejnowski

This paper briefly describes an artificial neural network for preattentive visual processing. The network is capable of determining image motion in a type of stimulus which defeats most popular methods of motion detection - a subset of second-order visual motion stimuli known as drift-balanced stimuli(DBS). The processing stages of the network described in this paper are integratable into a model capable of simultaneous motion extraction, edge detection, and the determination of occlusion.

Against Edges: Function Approximation with Multiple Support Maps

Trevor Darrell, Alex Pentland

Networks for reconstructing a sparse or noisy function often use an edge field to segment the function into homogeneous regions. This approach assumes that these regions do not overlap or have disjoint parts, which is often false. For example, images which contain regions split by an occluding object can't be properly reconstructed using this type of network. We have developed a network that overcomes these limitations, using support maps to represent the segmentation of a signal. In our approach, the support of each region in the signal is explicitly represented. Results from an initial implementation demonstrate that this method can reconstruct images and motion sequences which contain complicated occlusion.

Induction of Multiscale Temporal Structure

Michael C. Mozer

Learning structure in temporally-extended sequences is a difficult computational problem because only a fraction of the relevant information is available at any instant. Although variants of back propagation can in principle be used to find structure in sequences, in practice they are not sufficiently powerful to discover arbitrary contingencies, especially those spanning long temporal intervals or involving high order statistics. For example, in designing a connectionist network for mus

ic composition, we have encountered the problem that the net is able to learn musical structure that occurs locally in time-e.g., relations among notes within a musical phrase-but not structure that occurs over longer time periods--e.g., relations among phrases. To address this problem, we require a means of constructing a reduced description of the sequence that makes global aspects more explicit or more readily detectable. I propose to achieve this using hidden units that operate with different time constants. Simulation experiments indicate that slower time-scale hidden units are able to pick up global structure, structure that simply can not be learned by standard back propagation.

Locomotion in a Lower Vertebrate: Studies of the Cellular Basis of Rhythmogenesis and Oscillator Coupling

James Buchanan

To test whether the known connectivities of neurons in the lamprey spinal cord are sufficient to account for locomotor rhythmogenesis, a connectionist neural network simulation was done using identical cells connected according to experimentally established patterns. It was demonstrated that the network oscillates in a stable manner with the same phase relationships among the neurons as observed in the lamprey. The model was then used to explore coupling between identical

Single Neuron Model: Response to Weak Modulation in the Presence of Noise

A. R. Bulsara, W. Jacobs

We consider a bistable single neuron model driven by a periodic external modulation. The modulation introduces a correlated switching between states driven by the noise. The information flow through the system from the modulation to the output switching events, leads to a succession of strong peaks in the power spectrum. The signal-to-noise ratio (SNR) obtained from this power spectrum is a measure of the information content in the neuron response. With increasing noise intensity, the SNR passes through a maximum, an effect which has been called stochastic resonance. We treat the problem within the framework of a recently developed approximate theory, valid in the limits of weak noise intensity, weak periodic forcing and low forcing frequency. A comparison of the results of this theory with those obtained from a linear system FFT is also presented.

Fast Learning with Predictive Forward Models

Carlos Brody

A method for transforming performance evaluation signals distal both in space and time into proximal signals usable by supervised learning algorithms, presented in [Jordan & Jacobs 90], is examined. A simple observation concerning differentiation through models trained with redundant inputs (as one of their networks is) explains a weakness in the original architecture and suggests a modification: an internal world model that encodes action-space exploration and, crucially, cancels input redundancy to the forward model is added. Learning time on an example task, cart pole balancing, is thereby reduced about 50 to 100 times.

Nonlinear Pattern Separation in Single Hippocampal Neurons with Active Dendritic Membrane

Anthony Zador, Brenda Claiborne, Thomas Brown

The dendritic trees of cortical pyramidal neurons seem ideally suited to perform local processing on inputs. To explore some of the implications of this complexity for the computational power of neurons, we simulated a realistic biophysical model of a hippocampal pyramidal cell in which a "cold spot"-a high density patch of inhibitory Ca-dependent K channels and a colocalized patch of Ca channels-was present at a dendritic branch point. The cold spot induced a non monotonic relationship between the strength of the synaptic in

put and the probability of neuronal firing. This effect could also be interpreted as an analog stochastic XOR.

A Neural Network for Motion Detection of Drift-Balanced Stimuli

Hilary Tunley

This paper briefly describes an artificial neural network for preattentive visual processing. The network is capable of determining image motion in a type of stimulus which defeats most popular methods of motion detection - a subset of second-order visual motion stimuli known as drift-balanced stimuli (DBS). The processing stages of the network described in this paper are integratable into a model capable of simultaneous motion extraction, edge detection, and the determination of occlusion.

Generalization Performance in PARSEC - A Structured Connectionist Parsing Architecture

Ajay Jain

This paper presents PARSEC-a system for generating connectionist parsing networks from example parses. PARSEC is not based on formal grammar systems and is geared toward spoken language tasks. PARSEC networks exhibit three strengths important for application to speech processing: 1) they learn to parse, and generalize well compared to hand-coded grammars; 2) they tolerate several types of noise; 3) they can learn to use multi-modal input. Presented are the PARSEC architecture and performance analyses along several dimensions that demonstrate PARSEC's features. PARSEC's performance is compared to that of traditional grammar-based parsing systems.

Hierarchies of adaptive experts

Michael Jordan, Robert Jacobs

In this paper we present a neural network architecture that discovers a recursive decomposition of its input space. Based on a generalization of the modular architecture of Jacobs, Jordan, Nowlan, and Hinton (1991), the architecture uses competition among networks to recursively split the input space into nested regions and to learn separate associative mappings within each region. The learning algorithm is shown to perform gradient ascent in a log likelihood function that captures the architecture's hierarchical structure.

Optical Implementation of a Self-Organizing Feature Extractor

Dana Anderson, Claus Benkert, Verena Hebler, Ju-Seog Jang, Don Montgomery, Mark Saffman

We demonstrate a self-organizing system based on photorefractive ring oscillators. We employ the system in two ways that can both be thought of as feature extractors; one acts on a set of images exposed repeatedly to the system strictly as a linear feature extractor, and the other serves as a signal demultiplexer for fiber optic communications. Both systems implement unsupervised competitive learning embedded within the mode interaction dynamics between the modes of a set of ring oscillators. After a training period, the modes of the rings become associated with frequencies within the incoming data stream.

A Neurocomputer Board Based on the ANNA Neural Network Chip

Eduard Säckinger, Bernhard Boser, Lawrence Jackel

A board is described that contains the ANNA neural-network chip, and a DSP32C digital signal processor. The ANNA (Analog Neural Network Arithmetic unit) chip performs mixed analog/digital processing. The combination of ANNA with the DSP allows high-speed, end-to-end execution of numerous signal-processing applications, including the preprocessing, the neural-net calculations, and the postprocessing steps. The ANNA board evaluates neural networks 10 to 100 times faster than the DSP alone. The board is suitable for implementing large (million connections) networks with sparse weight matrices. Three applications

have been implemented on the board: a convolver network for slant detection of text blocks, a handwritten digit recognizer, and a neural network for recognition-based segmentation.

Adaptive Synchronization of Neural and Physical Oscillators

Kenji Doya, Shuji Yoshizawa

Animal locomotion patterns are controlled by recurrent neural networks called central pattern generators (CPGs). Although a CPG can oscillate autonomously, its rhythm and phase must be well coordinated with the state of the physical system using sensory inputs. In this paper we propose a learning algorithm for synchronizing neural and physical oscillators with specific phase relationships. Sensory input connections are modified by the correlation between cellular activities and input signals. Simulations show that the learning rule can be used for setting sensory feedback connections to a CPG as well as coupling connections between CPGs.

Recurrent Networks and NARMA Modeling

Jerome Connor, Les Atlas, Douglas Martin

There exist large classes of time series, such as those with nonlinear moving average components, that are not well modeled by feedforward networks or linear models, but can be modeled by recurrent networks. We show that recurrent neural networks are a type of nonlinear autoregressive-moving average (NARMA) model. Practical ability will be shown in the results of a competition sponsored by the Puget Sound Power and Light Company, where the recurrent networks gave the best performance on electric load forecasting.

ANN Based Classification for Heart Defibrillators

M. Jabri, S. Pickard, P. Leong, Z. Chi, B. Flower, Y. Xie

Current Intra-Cardiac defibrillators make use of simple classification algorithms to determine patient conditions and subsequently to enable proper therapy. The simplicity is primarily due to the constraints on power dissipation and area available for implementation. Sub-threshold implementation of artificial neural networks offer potential classifiers with higher performance than commercially available defibrillators. In this paper we explore several classifier architectures and discuss micro-electronic implementation issues.

Adaptive Development of Connectionist Decoders for Complex Error-Correcting Codes

Sheri Gish, Mario Blaum

We present an approach for development of a decoder for any complex binary error-correcting code (ECC) via training from examples of decoded received words. Our decoder is a connectionist architecture. We describe two separate solutions: A system-level solution (the Cascaded Networks Decoder); and the ECC-Enhanced Decoder, a solution which simplifies the mapping problem which must be solved for decoding. Although both solutions meet our basic approach constraint for simplicity and compactness, only the ECC-Enhanced Decoder meets our second basic constraint of being a generic solution.

Self-organization in real neurons: Anti-Hebb in 'Channel Space'?

Anthony Bell

Ion channels are the dynamical systems of the nervous system. Their distribution within the membrane governs not only communication of information between neurons, but also how that information is integrated within the cell. Here, an argument is presented for an 'anti-Hebbian' rule for changing the distribution of voltage-dependent ion channels in order to flatten voltage curvatures in dendrites. Simulations show that this rule can account for the self-organization of dynamical receptive field properties such as resonance and direction selectivity. It also creates the conditions for the faithful conduction within the

e cell of signals to which the cell has been exposed. Various possible cellular implementations of such a learning rule are proposed, including activity-dependent migration of channel proteins in the plane of the membrane.

Tangent Prop - A formalism for specifying selected invariances in an adaptive network

Patrice Simard, Bernard Victorri, Yann LeCun, John Denker

In many machine learning applications, one has access, not only to training data, but also to some high-level a priori knowledge about the desired behavior of the system. For example, it is known in advance that the output of a character recognizer should be invariant with respect to small spatial distortions of the input images (translations, rotations, scale changes, etcetera). We have implemented a scheme that allows a network to learn the derivative of its outputs with respect to distortion operators of our choosing.

This not only reduces the learning time and the amount of training data, but also provides a powerful language for specifying what generalizations we wish the network to perform.

Human and Machine 'Quick Modeling'

Jakob Bernasconi, Karl Gustafson

We present here an interesting experiment in 'quick modeling' by humans, performed independently on small samples, in several languages and two continents, over the last three years. Comparisons to decision tree procedures and neural net processing are given. From these, we conjecture that human reasoning is better represented by the latter, but substantially different from both. Implications for the 'strong convergence hypothesis' between neural networks and machine learning are discussed, now expanded to include human reasoning comparisons.

Practical Issues in Temporal Difference Learning

Gerald Tesauro

This paper examines whether temporal difference methods for training connectionist networks, such as Sutton's TD(0) algorithm, can be successfully applied to complex real-world problems. A number of important practical issues are identified and discussed from a general theoretical perspective. These practical issues are then examined in the context of a case study in which TD(0) is applied to learning the game of backgammon from the outcome of self-play. This is apparently the first application of this algorithm to a complex nontrivial task. It is found that, with zero knowledge built in, the network is able to learn from scratch to play the entire game at a fairly strong intermediate level of performance, which is clearly better than conventional commercial programs, and which in fact surpasses comparable networks trained on a massive human expert data set. The hidden units in these network have apparently discovered useful features, a longstanding goal of computer games research. Furthermore, when a set of hand-crafted features is added to the input representation, the resulting networks reach a near-expert level of performance, and have achieved good results against world-class human play.

A Network of Localized Linear Discriminants

Martin Glassman

The localized linear discriminant network (LLDN) has been designed to address classification problems containing relatively closely spaced data from different classes (encounter zones [1], the accuracy problem [2]). Locally trained hyperplane segments are an effective way to define the decision boundaries for these regions [3]. The LLD uses a modified perceptron training algorithm for effective discovery of separating hyperplane/sigmoid units within narrow boundaries. The basic unit of the network is the discriminant receptive field (DRF) which combines the LLD function with Gaussians representing the dispersion of the local training data with respect to the hyperplane. The DRF implemen

ts a local distance mea(cid:173) sure [4], and obtains the benefits of networks of localized units [5]. A constructive algorithm for the two-class case is described which incorporates DRF's into the hidden layer to solve local discrimination problems. The output unit produces a smoothed, piecewise linear decision boundary. Preliminary results indicate the ability of the LLDN to efficiently achieve separation when boundaries are narrow and complex, in cases where both the "standard" multilayer perceptron (MLP) and k-nearest neighbor (KNN) yielded high error rates on training data.

Multi-Digit Recognition Using a Space Displacement Neural Network

Ofer Matan, Christopher J. C. Burges, Yann LeCun, John Denker

We present a feed-forward network architecture for recognizing an uncon(cid:173) strained handwritten multi-digit string. This is an extension of previous work on recognizing isolated digits. In this architecture a single digit recognizer is replicated over the input. The output layer of the network is coupled to a Viterbi alignment module that chooses the best interpretation of the input. Training errors are propagated through the Viterbi module. The novelty in this procedure is that segmentation is done on the feature maps developed in the Space Displacement Neural Network (SDNN) rather than the input (pixel) space.

JANUS: Speech-to-Speech Translation Using Connectionist and Non-Connectionist Techniques

Alex Waibel, Ajay Jain, Arthur McNair, Joe Tebelskis, Louise Osterholtz, Hiroaki Saito, Otto Schmidbauer, Tilo Sloboda, Monika Woszczyna

We present JANUS, a speech-to-speech translation system that utilizes diverse processing strategies, including connectionist learning, traditional AI knowledge representation approaches, dynamic programming, and stochastic techniques. JANUS translates continuously spoken English and German into German, English, and Japanese. JANUS currently achieves 87% translation fidelity from English speech and 97% from German speech. We present the JANUS system along with comparative evaluations of its interchangeable processing components, with special emphasis on the connectionist modules.

A Parallel Analog CCD/CMOS Signal Processor

Charles Neugebauer, Amnon Yariv

A CCO based signal processing IC that computes a fully parallel single quadrant vector-matrix multiplication has been designed and fabricated with a 2j..un CC O/CMOS process. The device incorporates an array of Charge Coupled Devices (CCO) which hold an analog matrix of charge encoding the matrix elements. Input vectors are digital with 1 - 8 bit accuracy.

Iterative Construction of Sparse Polynomial Approximations

Terence Sanger, Richard S. Sutton, Christopher Matheus

Christopher J. Matheus

A Computational Mechanism to Account for Averaged Modified Hand Trajectories

Ealan Henis, Tamar Flash

Using the double-step target displacement paradigm the mechanisms underlying arm trajectory modification were investigated. Using short (10- 110 msec) inter-stimulus intervals the resulting hand motions were initially directed in between the first and second target locations. The kinematic features of the modified motions were accounted for by the superposition scheme, which involves the vectorial addition of two independent point-to-point motion units: one for moving the hand toward an internally specified location and a second one for moving between that location and the final target location. The similarity between the inferred internally specified locations and previously reported measured end-points of the first saccades in double-step eye-movement studies may suggest similarities between perceived target locations in eye and hand motor control.

Learning Global Direct Inverse Kinematics

David DeMers, Kenneth Kreutz-Delgado

We introduce and demonstrate a bootstrap method for construction of an inverse function for the robot kinematic mapping using only sample configuration space/workspace data. Unsupervised learning (clustering) techniques are used on pre-image neighborhoods in order to learn to partition the configuration space into subsets over which the kinematic mapping is invertible. Supervised learning is then used separately on each of the partitions to approximate the inverse function. The ill-posed inverse kinematics function is thereby regularized, and a global inverse kinematics solution for the wristless Puma manipulator is developed.

VISIT: A Neural Model of Covert Visual Attention

Subutai Ahmad

Visual attention is the ability to dynamically restrict processing to a subset of the visual field. Researchers have long argued that such a mechanism is necessary to efficiently perform many intermediate level visual tasks. This paper describes VISIT, a novel neural network model of visual attention. The current system models the search for target objects in scenes containing multiple distractors. This is a natural task for people, it is studied extensively by psychologists, and it requires attention. The network's behavior closely matches the known psychophysical data on visual search and visual attention. VISIT also matches much of the physiological data on attention and provides a novel view of the functionality of a number of visual areas. This paper concentrates on the biological plausibility of the model and its relationship to the primary visual cortex, pulvinar, superior colliculus and posterior parietal areas.

Kernel Regression and Backpropagation Training With Noise

Petri Koistinen, Lasse Holmström

One method proposed for improving the generalization capability of a feedforward network trained with the backpropagation algorithm is to use artificial training vectors which are obtained by adding noise to the original training vectors. We discuss the connection of such backpropagation training with noise to kernel density and kernel regression estimation. We compare by simulated examples (1) backpropagation, (2) backpropagation with noise, and (3) kernel regression in mapping estimation and pattern classification contexts.

Recognition of Manipulated Objects by Motor Learning

Hiroaki Gomi, Mitsuo Kawato

We present two neural network controller learning schemes based on feedback error-learning and modular architecture for recognition and control of multiple manipulated objects. In the first scheme, a Gating Network is trained to acquire object-specific representations for recognition of a number of objects (or sets of objects). In the second scheme, an Estimation Network is trained to acquire function-specific, rather than object-specific, representations which directly estimate physical parameters. Both recognition networks are trained to identify manipulated objects using somatic and/or visual information. After learning, appropriate motor commands for manipulation of each object are issued by the control networks.

Temporal Adaptation in a Silicon Auditory Nerve

John Lazzaro

Many auditory theorists consider the temporal adaptation of the auditory nerve a key aspect of speech coding in the auditory periphery. Experiments with models of auditory localization and pitch perception also suggest temporal adaptation is an important element of practical auditory processing. I have designed, fabricated, and successfully tested

an analog integrated circuit that models many aspects of auditory nerve response, including temporal adaptation.

Oscillatory Model of Short Term Memory

David Horn, Marius Usher

We investigate a model in which excitatory neurons have dynamical thresholds which display both fatigue and potentiation. The fatigue property leads to oscillatory behavior. It is responsible for the ability of the model to perform segmentation, i.e., decompose a mixed input into staggered oscillations of the activities of the cell-assemblies (memories) affected by it. Potentiation is responsible for sustaining these staggered oscillations after the input is turned off, i.e. the system serves as a model for short term memory. It has a limited STM capacity, reminiscent of the magical number 7 ± 2 .

Estimating Average-Case Learning Curves Using Bayesian, Statistical Physics and VC Dimension Methods

David Haussler, Michael Kearns, Manfred Oppen, Robert Schapire

In this paper we investigate an average-case model of concept learning, and give results that place the popular statistical physics and VC dimension theories of learning curve behavior in a common framework.

Operators and curried functions: Training and analysis of simple recurrent networks

Janet Wiles, Anthony Bloesch

We present a framework for programming the hidden unit representations of simple recurrent networks based on the use of hint units (additional targets at the output layer). We present two ways of analysing a network trained within this framework: Input patterns act as operators on the information encoded by the context units; symmetrically, patterns of activation over the context units act as curried functions of the input sequences. Simulations demonstrate that a network can learn to represent three different functions simultaneously and canonical discriminant analysis is used to investigate how operators and curried functions are represented in the space of hidden unit activations.

The Efficient Learning of Multiple Task Sequences

Satinder Singh

I present a modular network architecture and a learning algorithm based on incremental dynamic programming that allows a single learning agent to learn to solve multiple Markovian decision tasks (MDTs) with significant transfer of learning across the tasks. I consider a class of MDTs, called composite tasks, formed by temporally concatenating a number of simpler, elemental MDTs. The architecture is trained on a set of composite and elemental MDTs. The temporal structure of a composite task is assumed to be unknown and the architecture learns to produce a temporal decomposition. It is shown that under certain conditions the solution of a composite MDT can be constructed by computationally inexpensive modifications of the solutions of its constituent elemental MDTs.

Benchmarking Feed-Forward Neural Networks: Models and Measures

Leonard Hamey

Existing metrics for the learning performance of feed-forward neural networks do not provide a satisfactory basis for comparison because the choice of the training epoch limit can determine the results of the comparison. I propose new metrics which have the desirable property of being independent of the training epoch limit. The efficiency measures the yield of correct networks in proportion to the training effort expended. The optimal epoch limit provides the greatest efficiency. The learning performance is modelled statistically, and asymptotic performance is estimated. Implementation details may be found in (Hamey, 1992).

CCD Neural Network Processors for Pattern Recognition

Alice Chiang, Michael Chuang, Jeffrey LaFranchise

A CCD-based processor that we call the NNC2 is presented. The NNC2 implements a fully connected 192-input, 32-output two-layer network and can be cascaded to form multilayer networks or used in parallel for additional input or output nodes. The device computes 1.92×10^9 connections/sec when clocked at 10 MHz. Network weights can be specified to six bits of accuracy and are stored on-chip in programmable digital memories. A neural network pattern recognition system using NNC2 and CCD image feature extractor (IFE) devices is described. Additionally, we report a CCD output circuit that exploits inherent nonlinearities in the charge injection process to realize an adjustable-threshold sigmoid in a chip area of $40 \times 80 \mu\text{m}^2$.

A Simple Weight Decay Can Improve Generalization

Anders Krogh, John Hertz

It has been observed in numerical simulations that a weight decay can improve generalization in a feed-forward neural network. This paper explains why. It is proven that a weight decay has two effects in a linear network. First, it suppresses any irrelevant components of the weight vector by choosing the smallest vector that solves the learning problem. Second, if the size is chosen right, a weight decay can suppress some of the effects of static noise on the targets, which improves generalization quite a lot. It is then shown how to extend these results to networks with hidden layers and non-linear units. Finally the theory is confirmed by some numerical simulations using the data from NetTalk.

Neural Network Diagnosis of Avascular Necrosis from Magnetic Resonance Images

Armando Manduca, Paul Christy, Richard Ehman

Avascular necrosis (AVN) of the femoral head is a common yet potentially serious disorder which can be detected in its very early stages with magnetic resonance imaging. We have developed multi-layer perceptron networks, trained with conjugate gradient optimization, which diagnose AVN from single magnetic resonance images of the femoral head with 100% accuracy on training data and 97% accuracy on test data.

Obstacle Avoidance through Reinforcement Learning

Tony Prescott, John Mayhew

A method is described for generating plan-like, reflexive, obstacle avoidance behaviour in a mobile robot. The experiments reported here use a simulated vehicle with a primitive range sensor. Avoidance behaviour is encoded as a set of continuous functions of the perceptual input space. These functions are stored using CMACs and trained by a variant of Barto and Sutton's adaptive critic algorithm. As the vehicle explores its surroundings it adapts its responses to sensory stimuli so as to minimise the negative reinforcement arising from collisions. Strategies for local navigation are therefore acquired in an explicitly goal-driven fashion. The resulting trajectories form elegant collision-free paths through the environment.

The VC-Dimension versus the Statistical Capacity of Multilayer Networks

Chuanyi Ji, Demetri Psaltis

A general relationship is developed between the VC-dimension and the statistical lower epsilon-capacity which shows that the VC-dimension can be lower bounded (in order) by the statistical lower epsilon-capacity of a network trained with random samples. This relationship explains quantitatively how generalization takes place after memorization, and relates the concept of generalization (consistency) with the capacity of the optimal classifier over a class of classifiers with the same structure and the capacity.

ity of the Bayesian classifier. Furthermore, it provides a general methodology to evaluate a lower bound for the VC-dimension of feedforward multilayer neural networks. This general methodology is applied to two types of networks which are important for hardware implementations: two layer $(N - 2L - 1)$ net(cid:173) works with binary weights, integer thresholds for the hidden units and zero threshold for the output unit, and a single neuron $((N - 1)$ net(cid:173) works) with binary weights and a zero threshold. Specifically, we obtain $OC \sim L$::; $d2 \sim O(W)$, and $d1 \sim O(N)$. Here W is the total number of weights of the $(N - 2L - 1)$ networks. $d1$ and $d2$ represent the VC(cid:173) dimensions for the $(N - 1)$ and $(N - 2L - 1)$ networks respectively.

Improving the Performance of Radial Basis Function Networks by Learning Center Locations

Dietrich Wettschereck, Thomas Dietterich

Three methods for improving the performance of (gaussian) radial basis function (RBF) networks were tested on the NETtaik task. In RBF, a new example is classified by computing its Euclidean distance to a set of centers chosen by unsupervised methods. The application of supervised learning to learn a non-Euclidean distance metric was found to reduce the error rate of RBF networks, while supervised learning of each center's variance resulted in inferior performance. The best improvement in accuracy was achieved by networks called generalized radial basis function (GRBF) networks. In GRBF, the center locations are determined by supervised learning. After training on 1000 words, RBF classifies 56.5% of letters correct, while GRBF scores 73.4% letters correct (on a separate test set). From these and other experiments, we conclude that supervised learning of center locations can be very important for radial basis function learning.

Models Wanted: Must Fit Dimensions of Sleep and Dreaming

J. Hobson, Adam Mamelak, Jeffrey Sutton

During waking and sleep, the brain and mind undergo a tightly linked and precisely specified set of changes in state. At the level of neurons, this process has been modeled by variations of Volterra-Lotka equations for cyclic fluctuations of brainstem cell populations. However, neural network models based upon rapidly developing knowledge of the specific population connectivities and their differential responses to drugs have not yet been developed. Furthermore, only the most preliminary attempts have been made to model across states. Some of our own attempts to link rapid eye movement (REM) sleep neurophysiology and dream cognition using neural network approaches are summarized in this paper.

Merging Constrained Optimisation with Deterministic Annealing to "Solve" Combinatorially Hard Problems

Paul Stolorz

Several parallel analogue algorithms, based upon mean field theory (MFT) approximations to an underlying statistical mechanics formulation, and requiring an externally prescribed annealing schedule, now exist for finding approximate solutions to difficult combinatorial optimisation problems. They have been applied to the Travelling Salesman Problem (TSP), as well as to various issues in computational vision and cluster analysis. I show here that any given MFT algorithm can be combined in a natural way with notions from the areas of constrained optimisation and adaptive simulated annealing to yield a single homogeneous and efficient parallel relaxation technique, for which an externally prescribed annealing schedule is no longer required. The results of numerical simulations on 50-city and 100-city TSP problems are presented, which show that the ensuing algorithms are typically an order of magnitude faster than the MFT algorithms alone, and which also show, on occasion, superior solutions as well.

Neural Control for Rolling Mills: Incorporating Domain Theories to Overcome Data Deficiency

Martin Röscheisen, Reimar Hofmann, Volker Tresp

In a Bayesian framework, we give a principled account of how domain-specific prior knowledge such as imperfect analytic domain theories can be optimally incorporated into networks of locally-tuned units: by choosing a specific architecture and by applying a specific training regimen. Our method proved successful in overcoming the data deficiency problem in a large-scale application to devise a neural control for a hot line rolling mill. It achieves in this application significantly higher accuracy than optimally-tuned standard algorithms such as sigmoidal backpropagation, and outperforms the state-of-the-art solution.

Polynomial Uniform Convergence of Relative Frequencies to Probabilities

Alberto Bertoni, Paola Campadelli, Anna Morpurgo, Sandra Panizza

We define the concept of polynomial uniform convergence of relative frequencies to probabilities in the distribution-dependent context. Let $X_n = \{0, 1\}^n$, let P_n be a probability distribution on X_n and let $F_n \subset 2^{X_n}$ be a family of events. The family $\{(X_n, P_n, F_n)\}_{n \geq 1}$ has the property of polynomial uniform convergence if the probability that the maximum difference (over F_n) between the relative frequency and the probability of an event exceed a given positive ϵ be at most $O(n^{-1})$ ($0 < \epsilon < 1$), when the sample on which the frequency is evaluated has size polynomial in $n, 1/\epsilon, 1/b$. Given at-sample (X_1, \dots, X_t) , let $C_t(X_1, \dots, X_t)$ be the Vapnik-Chervonenkis dimension of the family $\{f|f \in F_n\}$ and $M(n, t)$ the expectation $E(C_t)$. We show that $\{(X_n, P_n, F_n)\}_{n \geq 1}$ has the property of polynomial uniform convergence iff there exists $\epsilon_0 > 0$ such that $M(n, t) = O(n/t^3)$. Applications to distribution-dependent PAC learning are discussed.

Forward Dynamics Modeling of Speech Motor Control Using Physiological Data

Makoto Hirayama, Eric Vatikiotis-Bateson, Mitsuo Kawato, Michael Jordan

We propose a paradigm for modeling speech production based on neural networks. We focus on characteristics of the musculoskeletal system. Using real physiological data - articulator movements and EMG from muscle activity - a neural network learns the forward dynamics relating motor commands to muscles and the ensuing articulator behavior. After learning, simulated perturbations, were used to assess properties of the acquired model, such as natural frequency, damping, and interarticulator couplings. Finally, a cascade neural network is used to generate continuous motor commands from a sequence of discrete articulatory targets.

Learning to Segment Images Using Dynamic Feature Binding

Michael C. Mozer, Richard Zemel, Marlene Behrmann

Despite the fact that complex visual scenes contain multiple, overlapping objects, people perform object recognition with ease and accuracy. One operation that facilitates recognition is an early segmentation process in which features of objects are grouped and labeled according to which object they belong. Current computational systems that perform this operation are based on predefined grouping heuristics. We describe a system called MAGIC that learns how to group features based on a set of presegmented examples. In many cases, MAGIC discovers grouping heuristics similar to those previously proposed, but it also has the capability of finding nonintuitive structural regularities in images. Grouping is performed by a relaxation network that attempts to dynamically bind related features. Features transmit a complex-valued signal (amplitude and phase) to one another; binding can thus be represented by phase locking related features. MAGIC'S training procedure is a generalization of recurrent back propagation to complex-valued units.

Simulation of Optimal Movements Using the Minimum-Muscle-Tension-Change Model

Menashe Dornay, Yoji Uno, Mitsuo Kawato, Ryoji Suzuki

This work discusses various optimization techniques which were proposed in models for controlling arm movements. In particular, the minimum-muscle-tension-change model is investigated. A dynamic simulator of the monkey's arm, including seventeen single and double joint muscles, is utilized to generate horizontal hand movements. The hand trajectories produced by this algorithm are discussed.

Neural Network Analysis of Event Related Potentials and Electroencephalogram Predicts Vigilance

Rita Venturini, William Lytton, Terrence J. Sejnowski

Automated monitoring of vigilance in attention intensive tasks such as air traffic control or sonar operation is highly desirable. As the operator monitors the instrument, the instrument would monitor the operator, insuring against lapses. We have taken a first step toward this goal by using feedforward neural networks trained with backpropagation to interpret event related potentials (ERPs) and electroencephalogram (EEG) associated with periods of high and low vigilance. The accuracy of our system on an ERP data set averaged over 28 minutes was 96%, better than the 83% accuracy obtained using linear discriminant analysis. Practical vigilance monitoring will require prediction over shorter time periods. We were able to average the ERP over as little as 2 minutes and still get 90% correct prediction of a vigilance measure. Additionally, we achieved similarly good performance using segments of EEG power spectrum as short as 56 sec.

HARMONET: A Neural Net for Harmonizing Chorales in the Style of J. S. Bach

Hermann Hild, Johannes Feulner, Wolfram Menzel

HARMONET, a system employing connectionist networks for music processing, is presented. After being trained on some dozen Bach chorales using error backpropagation, the system is capable of producing four-part chorales in the style of J.S.Bach, given a one-part melody. Our system solves a musical real-world problem on a performance level appropriate for musical practice. HARMONET's power is based on (a) a new coding scheme capturing musically relevant information and (b) the integration of backpropagation and symbolic algorithms in a hierarchical system, combining the advantages of both.

A Self-Organizing Integrated Segmentation and Recognition Neural Net

Jim Keeler, David Rumelhart

We present a neural network algorithm that simultaneously performs segmentation and recognition of input patterns that self-organizes to detect input pattern locations and pattern boundaries. We demonstrate this neural network architecture on character recognition using the NIST database and report on results herein. The resulting system simultaneously segments and recognizes touching or overlapping characters, broken characters, and noisy images with high accuracy.

A Comparison of Projection Pursuit and Neural Network Regression Modeling

Jenq-Neng Huang, Hang Li, Martin Maechler, R. Martin, Jim Schimert

Two projection based feedforward network learning methods for model-free regression problems are studied and compared in this paper: one is the popular back-propagation learning (BPL); the other is the projection pursuit learning (PPL). Unlike the totally parametric BPL method, the PPL non-parametrically estimates unknown nonlinear functions sequentially (neuron-by-neuron and layer-by-layer) at each iteration while jointly estimating the interconnection weights. In terms of learning efficiency, both methods have comparable training speed when based on a Gauss-Newton optimization algorithm while the PPL is more parsimonious. In terms of learning robustness toward noise outliers, the BPL

is more sensitive to the outliers.

Unsupervised Classifiers, Mutual Information and 'Phantom Targets

John Bridle, Anthony Heading, David MacKay

David J.e. MacKay

Fast, Robust Adaptive Control by Learning only Forward Models

Andrew Moore

A large class of motor control tasks requires that on each cycle the controller is told its current state and must choose an action to achieve a specified, state-dependent, goal behaviour. This paper argues that the optimization of learning rate, the number of experimental control decisions before adequate performance is obtained, and robustness is of prime importance-if necessary at the expense of computation per controller and memory requirement. This is motivated by the observation that a robot which requires two thousand learning steps to achieve adequate performance, or a robot which occasionally gets stuck while learning, will always be undesirable, whereas moderate computational expense can be accommodated by increasingly powerful computer hardware. It is not unreasonable to assume the existence of inexpensive 100 Mflop controllers within a few years and so even processes with control cycles in the low tens of milliseconds will have millions of machine instructions in which to make their decisions. This paper outlines a learning control scheme which aims to make effective use of such computational power.

Induction of Finite-State Automata Using Second-Order Recurrent Networks

Raymond Watrous, Gary Kuhn

Second-order recurrent networks that recognize simple finite state languages over $\{0,1\}^*$ are induced from positive and negative examples.

Using the complete gradient of the recurrent network and sufficient training examples to constrain the definition of the language to be induced, solutions are obtained that correctly recognize strings of arbitrary length. A method for extracting a finite state automaton corresponding to an optimized network is demonstrated.

Network generalization for production: Learning and producing styled letterforms
Igor Grebert, David Stork, Ron Keesing, Steve Mims

We designed and trained a connectionist network to generate letterforms in a new font given just a few exemplars from that font. During learning, our network constructed a distributed internal representation of fonts as well as letters.

Despite the fact that each training instance exemplified both a font and a letter. It was necessary to have separate but interconnected hidden units for "letter" and "font" representations - several alternative architectures were not successful.

Green's Function Method for Fast On-Line Learning Algorithm of Recurrent Neural Networks

Guo-Zheng Sun, Hsing-Hen Chen, Yee-Chun Lee

The two well known learning algorithms of recurrent neural networks are the back-propagation (Rumelhart & el al., Werbos) and the forward propagation (Williams and Zipser). The main drawback of back-propagation is its off-line backward path in time for error cumulation. This violates the on-line requirement in many practical applications. Although the forward propagation algorithm can be used in an on-line manner, the annoying drawback is the heavy computation load required to update the high dimensional sensitivity matrix ($O(n^2)$ operations for each time step). Therefore, to develop a fast forward algorithm is a challenging task. In this paper we proposed a forward learning algorithm which is one order faster ($O(n)$ operations for each time step) than the sensitivity matrix algorithm. The basic idea is that instead of

by integrating the high dimensional sensitivity dynamic equation we solve forward in time for its Green's function to avoid the redundant computations, and then update the weights whenever the error is to be corrected.

Constrained Optimization Applied to the Parameter Setting Problem for Analog Circuits

David Kirk, Kurt Fleischer, Lloyd Watts, Alan Barr

We use constrained optimization to select operating parameters for two circuits: a simple 3-transistor square root circuit, and an analog VLSI artificial cochlea. This automated method uses computer controlled measurement and test equipment to choose chip parameters which minimize the difference between the actual circuit's behavior and a specified goal behavior. Choosing the proper circuit parameters is important to compensate for manufacturing variations or adjust circuit performance within a certain range. As biologically-motivated analog VLSI circuits become increasingly complex, implying more parameters, setting these parameters by hand will become more cumbersome. Thus an automated parameter setting method can be of great value [Fleischer 90]. Automated parameter setting is an integral part of a goal-based engineering design methodology in which circuits are constructed with parameters enabling a wide range of behaviors, and are then "tuned" to the desired behaviors automatically.

Fault Diagnosis of Antenna Pointing Systems using Hybrid Neural Network and Signal Processing Models

Padhraic Smyth, Jeff Mellstrom

We describe in this paper a novel application of neural networks to system health monitoring of a large antenna for deep space communications. The paper outlines our approach to building a monitoring system using hybrid signal processing and neural network techniques, including autoregressive modelling, pattern recognition, and Hidden Markov models. We discuss several problems which are somewhat generic in applications of this kind - in particular we address the problem of detecting classes which were not present in the training data. Experimental results indicate that the proposed system is sufficiently reliable for practical implementation.

Network Model of State-Dependent Sequencing

Jeffrey Sutton, Adam Mamelak, J. Hobson

A network model with temporal sequencing and state-dependent modulatory features is described. The model is motivated by neurocognitive data characterizing different states of waking and sleeping. Computer studies demonstrate how unique states of sequencing can exist within the same network under different aminergic and cholinergic modulatory influences. Relationships between state-dependent modulation, memory, sequencing and learning are discussed.

Splines, Rational Functions and Neural Networks

Robert C. Williamson, Peter Bartlett

Connections between spline approximation, approximation with rational functions, and feedforward neural networks are studied. The potential improvement in the degree of approximation in going from single to two hidden layer networks is examined. Some results of Birman and Solomjak regarding the degree of approximation achievable when knot positions are chosen on the basis of the probability distribution of examples rather than the function values are extended.

Learning to Make Coherent Predictions in Domains with Discontinuities

Suzanna Becker, Geoffrey E. Hinton

We have previously described an unsupervised learning procedure that discovers spatially coherent properties of the world by maximizing the information that parameters extracted from different parts of the sensory input convey about some common underlying cause. When given random dot stereograms of curved surfaces, this procedure learns to extra

ct sur(cid:173) face depth because that is the property that is coherent across space. It also learns how to interpolate the depth at one location from the depths at nearby locations (Becker and Hinton, 1992). In this paper, we propose two new models which handle surfaces with discontinuities. The first model attempts to detect cases of discontinuities and reject them. The second model develops a mixture of expert interpolators. It learns to detect the locations of discontinuities and to invoke specialized, asymmetric interpolators that do not cross the discontinuities.

Bayesian Model Comparison and Backprop Nets

David MacKay

The Bayesian model comparison framework is reviewed, and the Bayesian Occam's razor is explained. This framework can be applied to feedforward networks, making possible (1) objective comparisons between solutions using alternative network architectures; (2) objective choice of magnitude and type of weight decay terms; (3) quantified estimates of the error bars on network parameters and on network output. The framework also generates a measure of the effective number of parameters determined by the data. The relationship of Bayesian model comparison to recent work on prediction of generalisation ability (Guyon et al., 1992, Moody, 1992) is discussed.

Networks for the Separation of Sources that are Superimposed and Delayed

John Platt, Federico Faggin

We have created new networks to unmix signals which have been mixed either with time delays or via filtering. We first show that a subset of the Herault-Jutten learning rules fulfills a principle of minimum output power. We then apply this principle to extensions of the Herault-Jutten network which have delays in the feedback path. Our networks perform well on real speech and music signals that have been mixed using time delays or filtering.

Burst Synchronization without Frequency Locking in a Completely Solvable Neural Network Model

Heinz Schuster, Christof Koch

The dynamic behavior of a network model consisting of all-to-all excitatory coupled binary neurons with global inhibition is studied analytically and numerically. We prove that for random input signals, the output of the network consists of synchronized bursts with apparently random interspersions of noisy activity. Our results suggest that synchronous bursts can be generated by a simple neuronal architecture which amplifies incoming coincident signals. This synchronization process is accompanied by dampened oscillations which, by themselves, however, do not play any constructive role in this and can therefore be considered to be an epiphenomenon.

Incrementally Learning Time-varying Half-planes

Anthony Kuh, Thomas Petsche, Ronald Rivest

We present a distribution-free model for incremental learning when concepts vary with time. Concepts are caused to change by an adversary while an incremental learning algorithm attempts to track the changing concepts by minimizing the error between the current target concept and the hypothesis. For a single half-plane and the intersection of two half-planes, we show that the average mistake rate depends on the maximum rate at which an adversary can modify the concept. These theoretical predictions are verified with simulations of several learning algorithms including back propagation.

Oscillatory Neural Fields for Globally Optimal Path Planning

Michael Lemmon

A neural network solution is proposed for solving path planning problems faced by mobile robots. The proposed network is a two-dimensional sheet of neurons for

forming a distributed representation of the robot's workspace. Lateral interconnections between neurons are "cooperative", so that the network exhibits oscillatory behaviour. These oscillations are used to generate solutions of Bellman's dynamic programming equation in the context of path planning. Simulation experiments imply that these networks locate global optimal paths even in the presence of substantial levels of circuit noise.

Retinogeniculate Development: The Role of Competition and Correlated Retinal Activity

Ron Keesing, David Stork, Carla Shatz

During visual development, projections from retinal ganglion cells (RGCs) to the lateral geniculate nucleus (LGN) in cat are refined to produce ocular dominance layering and precise topographic mapping. Normal development depends upon activity in RGCs, suggesting a key role for activity-dependent synaptic plasticity. Recent experiments on prenatal retina show that during early development, "waves" of activity pass across RGCs (Meister, et al., 1991). We provide the first simulations to demonstrate that such retinal waves, in conjunction with Hebbian synaptic competition and early arrival of contralateral axons, can account for observed patterns of retinogeniculate projections in normal and experimentally-treated animals.

A Neural Net Model for Adaptive Control of Saccadic Accuracy by Primate Cerebellum and Brainstem

Paul Dean, John Mayhew, Pat Langdon

Accurate saccades require interaction between brainstem circuitry and the cerebellum. A model of this interaction is described, based on Kawato's principle of feedback-error-learning. In the model a part of the brainstem (the superior colliculus) acts as a simple feedback controller with no knowledge of initial eye position, and provides an error signal for the cerebellum to correct for eye-muscle nonlinearities. This teaches the cerebellum, modelled as a CMAC, to adjust appropriately the gain on the brainstem burst-generator's internal feedback loop and so alter the size of burst sent to the motoneurons. With direction-only errors the system rapidly learns to make accurate horizontal eye movements from any starting position, and adapts realistically to subsequent simulated eye-muscle weakening or displacement of the saccadic target.

Segmentation Circuits Using Constrained Optimization

John Harris

A novel segmentation algorithm has been developed utilizing an absolute (cid:173) value smoothness penalty instead of the more common quadratic regularizer. This functional imposes a piece-wise constant constraint on the segmented data. Since the minimized energy is guaranteed to be convex, there are no problems with local minima and no complex continuation methods are necessary to find the unique global minimum. By interpreting (cid:173) the minimized energy as the generalized power of a nonlinear resistive network, a continuous-time analog segmentation circuit was constructed.

Rule Induction through Integrated Symbolic and Subsymbolic Processing

Clayton McMillan, Michael C. Mozer, Paul Smolensky

We describe a neural network, called RuleNet, that learns explicit, symbolic condition-action rules in a formal string manipulation domain. RuleNet discovers functional categories over elements of the domain, and, at various points during learning, extracts rules that operate on these categories. The rules are then injected back into RuleNet and training continues, in a process called iterative projection. By incorporating rules in this way, RuleNet exhibits enhanced learning and generalization performance over alternative neural net approaches. By integrating symbolic rule learning and subsymbolic category learning, RuleNet has capabilities that go beyond a purely symbolic system. We show how

this architecture can be applied to the problem of case-role assignment in natural language processing, yielding a novel rule-based solution.

A comparison between a neural network model for the formation of brain maps and experimental data

K. Obermayer, K. Schulten, G. G. Blasdel

Recently, high resolution images of the simultaneous representation of orientation preference, orientation selectivity and ocular dominance have been obtained for large areas in monkey striate cortex by optical imaging [1-3]. These data allow for the first time a "local" as well as "global" description of the spatial patterns and provide strong evidence for correlations between orientation selectivity and ocular dominance. A quantitative analysis reveals that these correlations arise when a five-dimensional feature space (two dimensions for retinotopic space, one each for orientation preference, orientation specificity, and ocular dominance) is mapped into the two available dimensions of cortex while locally preserving topology. These results provide strong evidence for the concept of topology preserving maps which have been suggested as a basic design principle of striate cortex [4-7].

A Connectionist Learning Approach to Analyzing Linguistic Stress

Prahlad Gupta, David Touretzky

We use connectionist modeling to develop an analysis of stress systems in terms of ease of learnability. In traditional linguistic analyses, learnability arguments determine default parameter settings based on the feasibility of logically deducing correct settings from an initial state. Our approach provides an empirical alternative to such arguments. Based on perceptron learning experiments using data from nineteen human languages, we develop a novel characterization of stress patterns in terms of six parameters. These provide both a partial description of the stress pattern itself and a prediction of its learnability, without invoking abstract theoretical constructs such as metrical feet. This work demonstrates that machine learning methods can provide a fresh approach to understanding linguistic phenomena.

Stationarity of Synaptic Coupling Strength Between Neurons with Nonstationary Discharge Properties

Mark R. Sydorenko, Eric Young

Based on a general non-stationary point process model, we computed estimates of the synaptic coupling strength (efficacy) as a function of time after stimulus onset between an inhibitory interneuron and its target postsynaptic cell in the feline dorsal cochlear nucleus. The data consist of spike trains from pairs of neurons responding to brief tone bursts recorded in vivo. Our results suggest that the synaptic efficacy is non-stationary. Further, synaptic efficacy is shown to be inversely and approximately linearly related to average presynaptic spike rate. A second-order analysis suggests that the latter result is not due to non-linear interactions. Synaptic efficacy is less strongly correlated with postsynaptic rate and the correlation is not consistent across neural pairs.

Time-Warping Network: A Hybrid Framework for Speech Recognition

Esther Levin, Roberto Pieraccini, Enrico Bocchieri

Interest has been generated regarding speech recognition systems based on Hidden Markov Models (HMMs) and neural network (NN) hybrids. Such systems attempt to combine the best features of both models: the temporal structure of HMMs and the discriminative power of neural networks. In this work we define a time-warping (TW) neuron that extends the operation of the formal neuron of a back-propagation network by warping the input pattern to match it optimally to its weights. We show that a single-layer network of TW neurons is equivalent to a Gaussian density HMM-based discriminative power of this system by using back-propagation discriminative training, and/or by generalizing the structure of the recognizer to a multi-layered net. The performance

rmance of the proposed network was evaluated on a highly confusable, isolated word. multi speaker recognition task. The results indicate that not only does the recognition performance improve. but the separation between classes to set up a rejection criterion to is enhanced also, allowing us improve the confidence of the system.

The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems

John Moody

We present an analysis of how the generalization performance (expected test set error) relates to the expected training set error for nonlinear learning systems, such as multilayer perceptrons and radial basis functions. The principal result is the following relationship (computed to second order) between the expected test set and training set errors:

Dual Inhibitory Mechanisms for Definition of Receptive Field Characteristics in a Cat Striate Cortex

A. Bonds

In single cells of the cat striate cortex, lateral inhibition across orientation and/or spatial frequency is found to enhance pre-existing biases. A contrast-dependent but spatially non-selective inhibitory component is also found. Stimulation with ascending and descending contrasts reveals the latter as a response hysteresis that is sensitive, powerful and rapid, suggesting that it is active in day-to-day vision. Both forms of inhibition are not recurrent but are rather network properties. These findings suggest two fundamental inhibitory mechanisms: a global mechanism that limits dynamic range and creates spatial selectivity through thresholding and a local mechanism that specifically refines spatial filter properties. Analysis of burst patterns in spike trains demonstrates that these two mechanisms have unique physiological origins.

Statistical Reliability of a Blowfly Movement-Sensitive Neuron

Rob de Ruyter van Steveninck, William Bialek

We develop a model-independent method for characterizing the reliability of neural responses to brief stimuli. This approach allows us to measure the discriminability of similar stimuli, based on the real-time response of a single neuron. Neurophysiological data were obtained from a movement-sensitive neuron (HI) in the visual system of the blowfly *Calliphora erythrocephala*. Furthermore, recordings were made from blowfly photoreceptor cells to quantify the signal to noise ratios in the peripheral visual system. As photoreceptors form the input to the visual system, the reliability of their signals ultimately determines the reliability of any visual discrimination task. For the case of movement detection, this limit can be computed, and compared to the HI neuron's reliability. Under favorable conditions, the performance of the HI neuron closely approaches the theoretical limit, which means that under these conditions the nervous system adds little noise in the process of computing movement from the correlations of signals in the photoreceptor array.

Networks with Learned Unit Response Functions

John Moody, Norman Yarvin

Feedforward networks composed of units which compute a sigmoidal function of a weighted sum of their inputs have been much investigated. We tested the approximation and estimation capabilities of networks using functions more complex than sigmoids. Three classes of functions were tested: polynomials, rational functions, and flexible Fourier series. Unlike sigmoids, these classes can fit non-monotonic functions. They were compared on three problems: prediction of Boston housing prices, the sunspot count, and robot arm inverse dynamics. The complex units attained clearly superior performance on the robot arm problem,

which is a highly non-monotonic, pure approximation problem. On the noisy and only mildly nonlinear Boston housing and sunspot problems, differences among the complex units were revealed; polynomials did poorly, whereas rationals and flexible Fourier series were comparable to sigmoids.

Adaptive Elastic Models for Hand-Printed Character Recognition

Geoffrey E. Hinton, Christopher Williams, Michael D. Revow

Hand-printed digits can be modeled as splines that are governed by about 8 control points. For each known digit, the control points have preferred locations, and deformations of the digit are generated by moving the control points away from their home locations. Images of digits can be produced by placing Gaussian ink generators uniformly along the spline. Real images can be recognized by finding the digit model most likely to have generated the data. For each digit model we use an elastic matching algorithm to minimize an energy function that includes both the deformation energy of the digit model and the log probability that the model would generate the inked pixels in the image. The model with the lowest total energy wins. If a uniform noise process is included in the model of image generation, some of the inked pixels can be rejected as noise as a digit model is fitting a poorly segmented image. The digit models learn by modifying the home locations of the control points.

Modeling Applications with the Focused Gamma Net

José Príncipe, Bert de Vries, Jyh-Ming Kuo, Pedro de Oliveira

The focused gamma network is proposed as one of the possible implementations of the gamma neural model. The focused gamma network is compared with the focused backpropagation network and TDNN for a time series prediction problem, and with ADALINE in a system identification problem.

Towards Faster Stochastic Gradient Search

Christian Darken, John Moody

Stochastic gradient descent is a general algorithm which includes LMS, on-line backpropagation, and adaptive k-means clustering as special cases. The standard choices of the learning rate η (both adaptive and fixed functions of time) often perform quite poorly. In contrast, our recently proposed class of "search then converge" learning rate schedules (Darken and Moody, 1990) display the theoretically optimal asymptotic convergence rate and a superior ability to escape from poor local minima. However, the user is responsible for setting a key parameter. We propose here a new methodology for creating the first completely automatic adaptive learning rates which achieve the optimal rate of convergence.

Gradient Descent: Second Order Momentum and Saturating Error

Barak Pearlmutter

Batch gradient descent, $\tilde{w}(t) = -\eta \nabla E / \nabla w(t)$, converges to a minimum of quadratic form with a time constant no better than $4\lambda_{\max} / \lambda_{\min}$ where λ_{\min} and λ_{\max} are the minimum and maximum eigenvalues of the Hessian matrix of E with respect to w . It was recently shown that adding a momentum term $\tilde{w}(t) = -\eta \nabla E / \nabla w(t) + Q[\tilde{w}(t-1)]$ improves this to $\sim \sqrt{\lambda_{\max} / \lambda_{\min}}$, although only in the batch case. Here we show that second order momentum, $\tilde{w}(t) = -\eta \nabla E / \nabla w(t) + Q[\tilde{w}(t-1) + (3-\alpha)\tilde{w}(t-2)]$, can lower this no further. We then regard gradient descent with momentum as a dynamic system and explore a non quadratic error surface, showing that saturation of the error accounts for a variety of effects observed in simulations and justifies some popular heuristics.

Active Exploration in Dynamic Environments

Sebastian B. Thrun, Knut Möller

Whenever an agent learns to control an unknown environment, two opposing principles have to be combined, namely: exploration (long-term optimization) and exploitation (short-term optimization). Many real-valued connectionist approaches to learning control realize exploration by random actions.

s in action selection. This might be disadvantageous when costs are assigned to "negative experiences". The basic idea presented in this paper is to make an agent explore unknown regions in a more directed manner. This is achieved by a so-called competence map, which is trained to predict the controller's accuracy, and is used for guiding exploration. Based on this, a bistable system enables smoothly switching attention between two behaviors - exploration and exploitation - depending on expected costs and knowledge gain. The appropriateness of this method is demonstrated by a simple robot navigation task.

Data Analysis using G/SPLINES

David Rogers

G/SPLINES is an algorithm for building functional models of data. It uses genetic search to discover combinations of basis functions which are then used to build a least-squares regression model. Because it produces a population of models which evolve over time rather than a single model, it allows analysis not possible with other regression-based approaches.

A Contrast Sensitive Silicon Retina with Reciprocal Synapses

Kwabena A. Boahen, Andreas Andreou

The goal of perception is to extract invariant properties of the underlying world. By computing contrast at edges, the retina reduces incident light intensities spanning twelve decades to a twentyfold variation. In one stroke, it solves the dynamic range problem and extracts relative reflectivity, bringing us a step closer to the goal. We have built a contrast sensitive silicon retina that models all major synaptic interactions in the outer plexiform layer of the vertebrate retina using current-mode CMOS circuits: namely, reciprocal synapses between cones and horizontal cells, which produce the antagonistic center/surround receptive field, and cone and horizontal cell gap junctions, which determine its size. The chip has 90 x 92 pixels on a 6.8 x 6.9mm die in 2/1m n-well technology and is fully functional.

Markov Random Fields Can Bridge Levels of Abstraction

Paul Cooper, Peter Prokopowicz

Network vision systems must make inferences from evidential information across levels of representational abstraction, from low level invariants, through intermediate scene segments, to high level behaviorally relevant object descriptions. This paper shows that such networks can be realized as Markov Random Fields (MRFs). We show first how to construct an MRF functionally equivalent to a Hough transform parameter network, thus establishing a principled probabilistic basis for visual networks. Second, we show that these MRF parameter networks are more capable and flexible than traditional methods. In particular, they have a well-defined probabilistic interpretation, intrinsically incorporate feedback, and offer richer representations and decision capabilities.

Information Measure Based Skeletonisation

Sowmya Ramachandran, Lorien Pratt

Automatic determination of proper neural network topology by trimming oversized networks is an important area of study, which has previously been addressed using a variety of techniques. In this paper, we present Information Measure Based Skeletonisation (IMBS), a new approach to this problem where superfluous hidden units are removed based on their information measure (IM). This measure, borrowed from decision tree induction techniques, reflects the degree to which the hyperplane formed by a hidden unit discriminates between training data classes. We show the results of applying IMBS to three classification tasks and demonstrate that it removes a substantial number of hidden units without significantly affecting network performance.

Interpretation of Artificial Neural Networks: Mapping Knowledge-Based Neural Net

works into Rules

Geoffrey Towell, Jude Shavlik

We propose and empirically evaluate a method for the extraction of expert(cid:173) comprehensible rules from trained neural networks. Our method operates in the context of a three-step process for learning that uses rule-based domain knowledge in combination with neural networks. Empirical tests using real(cid:173) worlds problems from molecular biology show that the rules our method extracts from trained neural networks: closely reproduce the accuracy of the network from which they came, are superior to the rules derived by a learning system that directly refines symbolic rules, and are expert-comprehensible.

Competitive Anti-Hebbian Learning of Invariants

Nicol Schraudolph, Terrence J. Sejnowski

Although the detection of invariant structure in a given set of input patterns is vital to many recognition tasks, connectionist learning rules tend to focus on directions of high variance (principal components). The prediction paradigm is often used to reconcile this dichotomy; here we suggest a more direct approach to invariant learning based on an anti-Hebbian learning rule. An unsupervised two-layer network implementing this method in a competitive setting learns to extract coherent depth information from random-dot stereograms.

3D Object Recognition Using Unsupervised Feature Extraction

Nathan Intrator, Joshua Gold, Heinrich Bülthoff, Shimon Edelman

Intrator (1990) proposed a feature extraction method that is related to recent statistical theory (Huber, 1985; Friedman, 1987), and is based on a biologically motivated model of neuronal plasticity (Bienenstock et al., 1982). This method has been recently applied to feature extraction in the context of recognizing 3D objects from single 2D views (Intrator and Gold, 1991). Here we describe experiments designed to analyze the nature of the extracted features, and their relevance to the theory and psychophysics of object recognition.

Information Processing to Create Eye Movements

David Robinson

Because eye muscles never cocontract and do not deal with external loads, one can write an equation that relates motoneuron firing rate to eye position and velocity - a very uncommon situation in the CNS. The semicircular canals transduce head velocity in a linear manner by using a high background discharge rate, imparting linearity to the premotor circuits that generate eye movements. This has allowed deducing some of the signal processing involved, including a neural network that integrates. These ideas are often summarized by block diagrams. Unfortunately, they are of little value in describing the behavior of single neurons - a finding supported by neural network models.

Propagation Filters in PDS Networks for Sequencing and Ambiguity Resolution

Ronald Sumida, Michael Dyer

We present a Parallel Distributed Semantic (PDS) Network architecture that addresses the problems of sequencing and ambiguity resolution in natural language understanding. A PDS Network stores phrases and their meanings using multiple PDP networks, structured in the form of a semantic net. A mechanism called Propagation Filters is employed: (1) to control communication between networks, (2) to properly sequence the components of a phrase, and (3) to resolve ambiguities. Simulation results indicate that PDS Networks and Propagation Filters can successfully represent high-level knowledge, can be trained relatively quickly, and provide for parallel inferencing at the knowledge level.

Neural Network Routing for Random Multistage Interconnection Networks

Mark Goudreau, C. Giles

A routing scheme that uses a neural network has been developed that can aid in establishing point-to-point communication routes through multi(cid:173)stage interconnection networks (MINs). The neural network is a network of the type that was examined by Hopfield (Hopfield, 1984 and 1985). In this work, the problem of establishing routes through random MINs (RMINs) in a shared-memory, distributed computing system is addressed. The performance of the neural network routing scheme is compared to two more traditional approaches - exhaustive search routing and greedy routing. The results suggest that a neural network router may be competitive for certain RMINs.

Direction Selective Silicon Retina that uses Null Inhibition

Ronald Benson, Tobi Delbrück

Biological retinas extract spatial and temporal features in an attempt to reduce the complexity of performing visual tasks. We have built and tested a silicon retina which encodes several useful temporal features found in vertebrate retinas. The cells in our silicon retina are selective to direction, highly sensitive to positive contrast changes around an ambient light level, and tuned to a particular velocity. Inhibitory connections in the null direction perform the direction selectivity we desire. This silicon retina is on a 4.6 x 6.8mm die and consists of a 47 x 41 array of photoreceptors.

Software for ANN training on a Ring Array Processor

Phil Kohn, Jeff Bilmes, Nelson Morgan, James Beck

Experimental research on Artificial Neural Network (ANN) algorithms requires either writing variations on the same program or making one monolithic program with many parameters and options. By using an object-oriented library, the size of these experimental programs is reduced while making them easier to read, write and modify. An efficient and flexible realization of this idea is Connection(cid:173)ist Layered Object-oriented Network Simulator (CLONES). CLONES runs on UNIX workstations and on the 100-1000 MFLOP Ring Array Processor (RAP) that we built with ANN algorithms in mind. In this report we describe CLONES and show how it is implemented on the RAP.

The Clusteron: Toward a Simple Abstraction for a Complex Neuron

Bartlett Mel

Are single neocortical neurons as powerful as multi-layered networks? A recent compartmental modeling study has shown that voltage-dependent membrane nonlinearities present in a complex dendritic tree can provide a virtual layer of local nonlinear processing elements between synaptic inputs and the final output at the cell body, analogous to a hidden layer in a multi-layer network. In this paper, an abstract model neuron is introduced, called a clusteron, which incorporates aspects of the dendritic "cluster-sensitivity" phenomenon seen in these detailed biophysical modeling studies. It is shown, using a clusteron, that a Hebb-type learning rule can be used to extract higher-order statistics from a set of training patterns, by manipulating the spatial ordering of synaptic connections onto the dendritic tree. The potential neurobiological relevance of these higher-order statistics for nonlinear pattern discrimination is then studied within a full compartmental model of a neocortical pyramidal cell, using a training set of 1000 high-dimensional sparse random patterns.

Application of Neural Network Methodology to the Modelling of the Yield Strength in a Steel Rolling Plate Mill

Ah Tsoi

In this paper, a tree based neural network viz. MARS (Friedman, 1991) for the modelling of the yield strength of a steel rolling plate mill is described. The inputs to the time series model are temperature, strain, strain rate,

and interpass time and the output is the corresponding yield stress. It is found that the MARS-based model reveals which variable's functional dependence is nonlinear, and significant. The results are compared with those obtained by using a Kalman filter based online tuning method and other classification methods, e.g. CART, C4.5, Bayesian classification. It is found that the MARS-based method consistently outperforms the other methods.

Reverse TDNN: An Architecture For Trajectory Generation

Patrice Simard, Yann Le Cun

The backpropagation algorithm can be used for both recognition and generation of time trajectories. When used as a recognizer, it has been shown that the performance of a network can be greatly improved by adding structure to the architecture. The same is true in trajectory generation. In particular a new architecture corresponding to a "reversed" TDNN is proposed. Results show dramatic improvement of performance in the generation of hand-written characters. A combination of TDNN and reversed TDNN for compact encoding is also suggested.

Principles of Risk Minimization for Learning Theory

V. Vapnik

Learning is posed as a problem of function estimation, for which two principles of solution are considered: empirical risk minimization and structural risk minimization. These two principles are applied to two different statements of the function estimation problem: global and local. Systematic improvements in prediction power are illustrated in application to zip-code recognition.

Perturbing Hebbian Rules

Peter Dayan, Geoffrey Goodhill

Feedforward networks composed of units which compute a sigmoidal function of a weighted sum of their inputs have been much investigated. We tested the approximation and estimation capabilities of networks using functions more complex than sigmoids. Three classes of functions were tested: polynomials, rational functions, and flexible Fourier series. Unlike sigmoids, these classes can fit non-monotonic functions. They were compared on three problems: prediction of Boston housing prices, the sunspot count, and robot arm inverse dynamics. The complex units attained clearly superior performance on the robot arm problem, which is a highly non-monotonic, pure approximation problem. On the noisy and only mildly nonlinear Boston housing and sunspot problems, differences among the complex units were revealed; polynomials did poorly, whereas rationals and flexible Fourier series were comparable to sigmoids.
