PAC-Bayes Learning of Conjunctions and Classification of Gene-Expression Data
Mario Marchand, Mohak Shah
We propose a “soft greedy” learning algorithm for building small conjunctions of simple threshold functions, called rays, de■ned on single real-valued attributes. We also propose a PAC-Bayes risk bound which is minimized for classi■ers achieving a non-trivial tradeo■ between sparsity (the number of rays used) and the mag- nitude of the separating margin of each ray. Finally, we test the soft greedy algorithm on four DNA micro-array data sets.
****
A Second Order Cone programming Formulation for Classifying Missing Data
Chiranjib Bhattacharyya, Pannagadatta Shivaswamy, Alex Smola
We propose a convex optimization based strategy to deal with uncertainty in the observations of a classification problem. We assume that instead of a sample (xi, yi) a distribution over (xi, yi) is specified. In particular, we derive a robust formulation when the distribution is given by a normal distribution. It leads to Second Order Cone Programming formulation. Our method is applied to the problem of missing data, where it outperforms direct imputation.
****
Learning first-order Markov models for control
Pieter Abbeel, Andrew Ng
First-order Markov models have been successfully applied to many prob- lems, for example in modeling sequential data using Markov chains, and modeling control problems using the Markov decision processes (MDP) formalism. If a ■rst-order Markov model’s parameters are estimated from data, the standard maximum likelihood estimator considers only the ■rst-order (single-step) transitions. But for many problems, the ■rst- order conditional independence assumptions are not satis■ed, and as a re- sult the higher order transition probabilities may be poorly approximated. Motivated by the problem of learning an MDP’s parameters for control, we propose an algorithm for learning a ■rst-order Markov model that ex- plicitly takes into account higher order interactions during training. Our algorithm uses an optimization criterion different from maximum likeli- hood, and allows us to learn models that capture longer range effects, but without giving up the bene■ts of using ■rst-order Markov models. Our experimental results also show the new algorithm outperforming conven- tional maximum likelihood estimation in a number of control problems where the MDP’s parameters are estimated from data.
****
Who's In the Picture
Tamara Berg, Alexander Berg, Jaety Edwards, David Forsyth
The context in which a name appears in a caption provides powerful cues as to who is depicted in the associated image. We obtain 44,773 face images, using a face detector, from approximately half a million captioned news images and automatically link names, obtained using a named entity recognizer, with these faces. A simple clustering method can produce fair results. We improve these results significantly by combining the clustering process with a model of the probability that an individual is depicted given its context. Once the labeling procedure is over, we have an accurately labeled set of faces, an appearance model for each individual depicted, and a natural language model that can produce accurate results on captions in isolation.
****
Optimal Aggregation of Classifiers and Boosting Maps in Functional Magnetic Resonance Imaging
Vladimir Koltchinskii, Manel Martínez-ramón, Stefan Posse
We study a method of optimal data-driven aggregation of classifiers in a convex combination and establish tight upper bounds on its excess risk with respect to a convex loss function under the assumption that the solution of optimal aggregation problem is sparse. We use a boosting type algorithm of optimal aggregation to develop aggregate classifiers of activation patterns in fMRI based on locally trained SVM classifiers.

The aggregation coefficients are then used to design a "boosting map" of the brain needed to identify the regions with most significant impact on classification.

****

Binet-Cauchy Kernels

Alex Smola, S.v.n. Vishwanathan

We propose a family of kernels based on the Binet-Cauchy theorem and its extension to Fredholm operators. This includes as special cases all currently known kernels derived from the behavioral framework, diffusion processes, marginalized kernels, kernels on graphs, and the kernels on sets arising from the subspace angle approach. Many of these kernels can be seen as the extrema of a new continuum of kernel functions, which leads to numerous new special cases. As an application, we apply the new class of kernels to the problem of clustering of video sequences with encouraging results.

****

Synergistic Face Detection and Pose Estimation with Energy-Based Models

Margarita Osadchy, Matthew Miller, Yann Cun

We describe a novel method for real-time, simultaneous multi-view face detection and facial pose estimation. The method employs a convolutional network to map face images to points on a manifold, parametrized by pose, and non-face images to points far from that manifold. This network is trained by optimizing a loss function of three variables: image, pose, and face/non-face label. We test the resulting system, in a single configuration, on three standard data sets one for frontal pose, one for rotated faces, and one for profiles and find that its performance on each set is comparable to previous multi-view face detectors that can only handle one form of pose variation. We also show experimentally that the system's accuracy on both face detection and pose estimation is improved by training for the two tasks together.

****

The power of feature clustering: An application to object detection

Shai Avidan, Moshe Butman

We give a fast rejection scheme that is based on image segments and demonstrate it on the canonical example of face detection. However, instead of focusing on the detection step we focus on the rejection step and show that our method is simple and fast to be learned, thus making it an excellent pre-processing step to accelerate standard machine learning classifiers, such as neural-networks, Bayes classifiers or SVM. We decompose a collection of face images into regions of pixels with similar behavior over the image set. The relationships between the mean and variance of image segments are used to form a cascade of rejectors that can reject over 99.8% of image patches, thus only a small fraction of the image patches must be passed to a full-scale classifier. Moreover, the training time for our method is much less than an hour, on a standard PC. The shape of the features (i.e. image segments) we use is data-driven, they are very cheap to compute and they form a very low dimensional feature space in which exhaustive search for the best features is tractable.

****

Synergies between Intrinsic and Synaptic Plasticity in Individual Model Neurons

Jochen Triesch

This paper explores the computational consequences of simultaneous intrinsic and synaptic plasticity in individual model neurons. It proposes a new intrinsic plasticity mechanism for a continuous activation model neuron based on low order moments of the neuron's firing rate distribution. The goal of the intrinsic plasticity mechanism is to enforce a sparse distribution of the neuron's activity level. In conjunction with Hebbian learning at the neuron's synapses, the neuron is shown to discover sparse directions in the input.

****

A Probabilistic Model for Online Document Clustering with Application to Novelty Detection
Jian Zhang, Zoubin Ghahramani, Yiming Yang
In this paper we propose a probabilistic model for online document clustering. We use non-parametric Dirichlet process prior to model the growing number of clusters, and use a prior of general English language model as the base distribution to handle the generation of novel clusters. Furthermore, cluster uncertainty is modeled with a Bayesian Dirichlet-multinomial distribution. We use empirical Bayes method to estimate hyperparameters based on a historical dataset. Our probabilistic model is applied to the novelty detection task in Topic Detection and Tracking (TDT) and compared with existing approaches in the literature.
****
Non-Local Manifold Tangent Learning
Yoshua Bengio, Martin Monperrus
We claim and present arguments to the effect that a large class of manifold learning algorithms that are essentially local and can be framed as kernel learning algorithms will suffer from the curse of dimensionality, at the dimension of the true underlying manifold. This observation suggests to explore non-local manifold learning algorithms which attempt to discover shared structure in the tangent planes at different positions. A criterion for such an algorithm is proposed and experiments estimating a tangent plane prediction function are presented, showing its advantages with respect to local manifold learning algorithms: it is able to generalize very far from training data (on learning handwritten character image rotations), where a local non-parametric method fails.
****
Maximal Margin Labeling for Multi-Topic Text Categorization
Hideto Kazawa, Tomonori Izumitani, Hirotoshi Taira, Eisaku Maeda
In this paper, we address the problem of statistical learning for multi- topic text categorization (MTC), whose goal is to choose all relevant top- ics (a label) from a given set of topics. The proposed algorithm, Max- imal Margin Labeling (MML), treats all possible labels as independent classes and learns a multi-class classi■er on the induced multi-class cate- gorization problem. To cope with the data sparseness caused by the huge number of possible labels, MML combines some prior knowledge about label prototypes and a maximal margin criterion in a novel way. Experi- ments with multi-topic Web pages show that MML outperforms existing learning algorithms including Support Vector Machines.
****
Conditional Random Fields for Object Recognition
Ariadna Quattoni, Michael Collins, Trevor Darrell
We present a discriminative part-based approach for the recognition of object classes from unsegmented cluttered scenes. Objects are modeled as flexible constellations of parts conditioned on local observations found by an interest operator. For each object class the probability of a given assignment of parts to local features is modeled by a Conditional Random Field (CRF). We propose an extension of the CRF framework that incorporates hidden variables and combines class conditional CRFs into a unified framework for part-based object recognition. The parameters of the CRF are estimated in a maximum likelihood framework and recognition proceeds by finding the most likely class under our model. The main advantage of the proposed CRF framework is that it allows us to relax the assumption of conditional independence of the observed data (i.e. local features) often used in generative approaches, an assumption that might be too restrictive for a considerable number of object classes.
****
Inference, Attention, and Decision in a Bayesian Neural Architecture
Angela J. Yu, Peter Dayan
We study the synthesis of neural coding, selective attention and perceptual decision making. A hierarchical neural architecture is proposed,

which implements Bayesian integration of noisy sensory input and top-down attentional priors, leading to sound perceptual discrimination. The model offers an explicit explanation for the experimentally observed modulation that prior information in one stimulus feature (location) can have on an independent feature (orientation). The network's intermediate levels of representation instantiate known physiological properties of visual cortical neurons. The model also illustrates a possible reconciliation of cortical and neuromodulatory representations of uncertainty.

****

Exponential Family Harmoniums with an Application to Information Retrieval

Max Welling, Michal Rosen-zvi, Geoffrey E. Hinton

Directed graphical models with one layer of observed random variables and one or more layers of hidden random variables have been the dom- inant modelling paradigm in many research ■elds. Although this ap- proach has met with considerable success, the causal semantics of these models can make it dif■cult to infer the posterior distribution over the hidden variables. In this paper we propose an alternative two-layer model based on exponential family distributions and the semantics of undi- rected models. Inference in these “exponential family harmoniums” is fast while learning is performed by minimizing contrastive divergence. A member of this family is then studied as an alternative probabilistic model for latent semantic indexing. In experiments it is shown that they perform well on document retrieval tasks and provide an elegant solution to searching with keywords.

****

Distributed Occlusion Reasoning for Tracking with Nonparametric Belief Propagation

Erik Sudderth, Michael Mandel, William Freeman, Alan Willsky

We describe a threedimensional geometric hand model suitable for visual tracking applications. The kinematic constraints implied by the model's joints have a probabilistic structure which is well described by a graphical model. Inference in this model is complicated by the hand's many degrees of freedom, as well as multimodal likelihoods caused by ambiguous image measurements. We use nonparametric belief propagation (NBP) to develop a tracking algorithm which exploits the graph's structure to control complexity, while avoiding costly discretization. While kinematic constraints naturally have a local structure, self occlusions created by the imaging process lead to complex interpendencies in color and edgebased likelihood functions. However, we show that local structure may be recovered by introducing binary hidden variables describing the occlusion state of each pixel. We augment the NBP algorithm to infer these occlusion variables in a distributed fashion, and then analytically marginalize over them to produce hand position estimates which properly account for occlusion events. We provide simulations showing that NBP may be used to refine inaccurate model initializations, as well as track hand motion through extended image sequences.

****

An Investigation of Practical Approximate Nearest Neighbor Algorithms

Ting Liu, Andrew Moore, Ke Yang, Alexander Gray

This paper concerns approximate nearest neighbor searching algorithms, which have become increasingly important, especially in high dimensional perception areas such as computer vision, with dozens of publications in recent years. Much of this enthusiasm is due to a successful new approximate nearest neighbor approach called Locality Sensitive Hashing (LSH). In this paper we ask the question: can earlier spatial data structure approaches to exact nearest neighbor, such as metric trees, be altered to provide approximate answers to proximity queries and if so, how? We introduce a new kind of metric tree that allows overlap: certain datapoints may appear in both the children of a parent. We also introduce new approximate k-NN search algorithms on this structure. We show why these structures should be able to exploit the same random-projection-based approximations that LSH enjoys, but with a simpler al- g

orithm and perhaps with greater efficiency. We then provide a detailed empirical evaluation on five large, high dimensional datasets which show up to 31-fold accelerations over LSH. This result holds true throughout the spectrum of approximation levels.

****

Hierarchical Distributed Representations for Statistical Language Modeling
John Blitzer, Fernando Pereira, Kilian Q. Weinberger, Lawrence Saul

Statistical language models estimate the probability of a word occurring in a given context. The most common language models rely on a discrete enumeration of predictive contexts (e.g., n-grams) and consequently fail to capture and exploit statistical regularities across these contexts. In this paper, we show how to learn hierarchical, distributed representations of word contexts that maximize the predictive value of a statistical language model. The representations are initialized by unsupervised algorithms for linear and nonlinear dimensionality reduction [14], then fed as input into a hierarchical mixture of experts, where each expert is a multinomial distribution over predicted words [12]. While the distributed representations in our model are inspired by the neural probabilistic language model of Bengio et al. [2, 3], our particular architecture enables us to work with significantly larger vocabularies and training corpora. For example, on a large-scale bigram modeling task involving a sixty thousand word vocabulary and a training corpus of three million sentences, we demonstrate consistent improvement over class-based bigram models [10, 13]. We also discuss extensions of our approach to longer multiword contexts.

****

Discrete profile alignment via constrained information bottleneck
Sean O'rourke, Gal Chechik, Robin Friedman, Eleazar Eskin

Amino acid profiles, which capture position-specific mutation probabilities, are a richer encoding of biological sequences than the individual sequences themselves. However, profile comparisons are much more computationally expensive than discrete symbol comparisons, making profiles impractical for many large datasets. Furthermore, because they are such a rich representation, profiles can be difficult to visualize. To overcome these problems, we propose a discretization for profiles using an expanded alphabet representing not just individual amino acids, but common profiles. By using an extension of information bottleneck (IB) incorporating constraints and priors on the class distributions, we find an informationally optimal alphabet. This discretization yields a concise, informative textual representation for profile sequences. Also alignments between these sequences, while nearly as accurate as the full profile-profile alignments, can be computed almost as quickly as those between individual or consensus sequences. A full pairwise alignment of SwissProt would take years using profiles, but less than 3 days using a discrete IB encoding, illustrating how discrete encoding can expand the range of sequence problems to which profile information can be applied.

****

Multiple Relational Embedding
Roland Memisevic, Geoffrey E. Hinton

We describe a way of using multiple different types of similarity relationship to learn a low-dimensional embedding of a dataset. Our method chooses different, possibly overlapping representations of similarity by individually reweighting the dimensions of a common underlying latent space. When applied to a single similarity relation that is based on Euclidean distances between the input data points, the method reduces to simple dimensionality reduction. If additional information is available about the dataset or about subsets of it, we can use this information to clean up or otherwise improve the embedding. We demonstrate the potential usefulness of this form of semi-supervised dimensionality reduc-

tion on some simple examples.
****
Conditional Models of Identity Uncertainty with Application to Noun Coreference
Andrew McCallum, Ben Wellner
Coreference analysis, also known as record linkage or identity uncer- t
ainty, is a difficult and important problem in natural language process-
ing, databases, citation matching and many other tasks. This paper intro-
duces several discriminative, conditional-probability models for coref-
erence analysis, all examples of undirected graphical models. Unlike
many historical approaches to coreference, the models presented here
are relational--they do not assume that pairwise coreference decisions
should be made independently from each other. Unlike other relational
models of coreference that are generative, the conditional model here can
incorporate a great variety of features of the input without having to be
concerned about their dependencies--paralleling the advantages of con-
ditional random fields over hidden Markov models. We present positive
results on noun phrase coreference in two standard text data sets.
****
Outlier Detection with One-class Kernel Fisher Discriminants
Volker Roth
The problem of detecting "atypical objects" or "outliers" is one of the
classical topics in (robust) statistics. Recently, it has been proposed to
address this problem by means of one-class SVM classifiers. The main
conceptual shortcoming of most one-class approaches, however, is that in
a strict sense they are unable to detect outliers, since the expected fract
ion of outliers has to be specified in advance. The method presented in
this paper overcomes this problem by relating kernelized one-class cla
ssifica- tion to Gaussian density estimation in the induced feature spa
ce. Having established this relation, it is possible to identify "atypi
cal objects" by quantifying their deviations from the Gaussian model. F
or RBF kernels it is shown that the Gaussian model is "rich enough" in
the sense that it asymptotically provides an unbiased estimator for the
true density. In or- der to overcome the inherent model selection prob
lem, a cross-validated likelihood criterion for selecting all free mode
l parameters is applied.
****
Kernel Projection Machine: a New Tool for Pattern Recognition
Laurent Zwald, Gilles Blanchard, Pascal Massart, Régis Vert
This paper investigates the effect of Kernel Principal Component Analy-
sis (KPCA) within the classification framework, essentially the regular-
ization properties of this dimensionality reduction method. KPCA has
been previously used as a pre-processing step before applying an SVM
but we point out that this method is somewhat redundant from a reg-
ularization point of view and we propose a new algorithm called Ker-
nel Projection Machine to avoid this redundancy, based on an analogy
with the statistical framework of regression for a Gaussian white n
oise model. Preliminary experimental results show that this algorith
m reaches the same performances as an SVM.
****
Seeing through water
Alexei Efros, Volkan Isler, Jianbo Shi, Mirkó Visontai
We consider the problem of recovering an underwater image distorted by
surface waves. A large amount of video data of the distorted image is
acquired. The problem is posed in terms of finding an undistorted im-
age patch at each spatial location. This challenging reconstruction task
can be formulated as a manifold learning problem, such that the center
of the manifold is the image of the undistorted patch. To compute the
center, we present a new technique to estimate global distances on th
e manifold. Our technique achieves robustness through convex flow com
- putations and solves the "leakage" problem inherent in recent manif

old embedding techniques.
****
Generative Affine Localisation and Tracking
John Winn, Andrew Blake
We present an extension to the Jojic and Frey (2001) layered sprite model which allows for layers to undergo affine transformations. This extension allows for affine object pose to be inferred whilst simultaneously learning the object shape and appearance. Learning is carried out by applying an augmented variational inference algorithm which includes a global search over a discretised transform space followed by a local optimisation. To aid correct convergence, we use bottom-up cues to restrict the space of possible affine transformations. We present results on a number of video sequences and show how the model can be extended to track an object whose appearance changes throughout the sequence.
****
Schema Learning: Experience-Based Construction of Predictive Action Models
Michael Holmes, Charles Jr.
Schema learning is a way to discover probabilistic, constructivist, pre- dictive action models (schemas) from experience. It includes meth- ods for ■nding and using hidden state to make predictions more accu- rate. We extend the original schema mechanism [1] to handle arbitrary discrete-valued sensors, improve the original learning criteria to handle POMDP domains, and better maintain hidden state by using schema pre- dictions. These extensions show large improvement over the original schema mechanism in several rewardless POMDPs, and achieve very low prediction error in a dif■cult speech modeling task. Further, we compare extended schema learning to the recently introduced predictive state rep- resentations [2], and ■nd their predictions of next-step action effects to be approximately equal in accuracy. This work lays the foundation for a schema-based system of integrated learning and planning.
****
Machine Learning Applied to Perception: Decision Images for Gender Classification
Felix A. Wichmann, Arnulf Graf, Heinrich Bülthoff, Eero Simoncelli, Bernhard Schölkopf
We study gender discrimination of human faces using a combination of psychophysical classi■cation and discrimination experiments together with methods from machine learning. We reduce the dimensionality of a set of face images using principal component analysis, and then train a set of linear classi■ers on this reduced representation (linear support vec- tor machines (SVMs), relevance vector machines (RVMs), Fisher linear discriminant (FLD), and prototype (prot) classi■ers) using human clas- si■cation data. Because we combine a linear preprocessor with linear classi■ers, the entire system acts as a linear classi■er, allowing us to visu- alise the decision-image corresponding to the normal vector of the separ- ating hyperplanes (SH) of each classi■er. We predict that the female-to- maleness transition along the normal vector for classi■ers closely mim- icking human classi■cation (SVM and RVM [1]) should be faster than the transition along any other direction. A psychophysical discrimina- tion experiment using the decision images as stimuli is consistent with this prediction.
****
An Information Maximization Model of Eye Movements
Laura Renninger, James Coughlan, Preeti Verghese, Jitendra Malik
We propose a sequential information maximization model as a general strategy for programming eye movements. The model reconstructs high-resolution visual information from a sequence of fixations, taking into account the fall-off in resolution from the fovea to the periphery. From this framework we get a simple rule for predicting fixation sequences: after each fixation, fixate next at the location that minimizes uncertainty (maximizes information) about the stimulus. By comparing our model performance to human eye movement data and to predictions from a saliency and random model, we demonstrate that our model is best at predicting fixation locations. Modeling additional biological constraints will imp

rove the prediction of fixation sequences. Our results suggest that informatio n maximization is a useful principle for programming eye movements.
****

Learning Hyper-Features for Visual Identification
Andras Ferencz, Erik Learned-miller, Jitendra Malik
We address the problem of identifying specific instances of a class (cars) from a set of images all belonging to that class. Although we cannot build a model for any particular instance (as we may be provided with only o ne "training" example of it), we can use information extracted from ob serv- ing other members of the class. We pose this task as a learning problem, in which the learner is given image pairs, labeled as matchin g or not, and must discover which image features are most consistent f or matching in- stances and discriminative for mismatches. We explore a patch based representation, where we model the distributions of simi larity measure- ments defined on the patches. Finally, we describe an algorithm that selects the most salient patches based on a mutual info rmation criterion. This algorithm performs identification well for our challenging dataset of car images, after matching only a few, well ch osen patches.
****

Parametric Embedding for Class Visualization
Tomoharu Iwata, Kazumi Saito, Naonori Ueda, Sean Stromsten, Thomas Griffiths, Jo shua Tenenbaum
In this paper, we propose a new method, Parametric Embedding (PE), for visualizi ng the posteriors estimated over a mixture model. PE simultane- ously embeds bot h objects and their classes in a low-dimensional space. PE takes as input a set of class posterior vectors for given data points, and tries to preserve the post erior structure in an embedding space by minimizing a sum of Kullback-Leibler di vergences, under the assump- tion that samples are generated by a Gaussian mixtu re with equal covari- ances in the embedding space. PE has many potential uses d epending on the source of the input data, providing insight into the classi■er’s be- havior in supervised, semi-supervised and unsupervised settings. The PE alg orithm has a computational advantage over conventional embedding methods based o n pairwise object relations since its complexity scales with the product of the number of objects and the number of classes. We demonstrate PE by visualizing su pervised categorization of web pages, semi-supervised categorization of digits, and the relations of words and latent topics found by an unsupervised algorithm, Latent Dirichlet Allo- cation.
****

A Topographic Support Vector Machine: Classification Using Local Label Configura tions
Johannes Mohr, Klaus Obermayer
The standard approach to the classification of objects is to consider the examples as independent and identically distributed (iid). In many real world settings, however, this assumption is not valid, because a topo- graphical relationship exists between the objects. In this contribution we consider the special case of image segmentation, where the objects are pixels and where the underlying topography is a 2D regular rectangular grid. We introduce a classification method which not only uses measured vectorial feature information but also the label configuration within a to- pographic neighborhood. Due to the resulting dependence between th e labels of neighboring pixels, a collective classification of a set of pixels becomes necessary. We propose a new method called 'Topographic Sup- port Vector Machine' (TSVM), which is based on a topographic kerne l and a self-consistent solution to the label assignment shown to be eq uiv- alent to a recurrent neural network. The performance of the algori thm is compared to a conventional SVM on a cell image segmentation task .
****

New Criteria and a New Algorithm for Learning in Multi-Agent Systems

Rob Powers, Yoav Shoham

We propose a new set of criteria for learning algorithms in multi-agent systems, one that is more stringent and (we argue) better justified than previous proposed criteria. Our criteria, which apply most straightforwardly in repeated games with average rewards, consist of three requirements: (a) against a specified class of opponents (this class is a parameter of the criterion) the algorithm yield a payoff that approaches the payoff of the best response, (b) against other opponents the algorithm's payoff at least approach (and possibly exceed) the security level payoff (or max-imin value), and (c) subject to these requirements, the algorithm achieve a close to optimal payoff in self-play. We furthermore require that these average payoffs be achieved quickly. We then present a novel algorithm, and show that it meets these new criteria for a particular parameter class, the class of stationary opponents. Finally, we show that the algorithm is effective not only in theory, but also empirically. Using a recently introduced comprehensive game theoretic test suite, we show that the algorithm almost universally outperforms previous learning algorithms.

****

The Cerebellum Chip: an Analog VLSI Implementation of a Cerebellar Model of Classical Conditioning

Constanze Hofstoetter, Manuel Gil, Kynan Eng, Giacomo Indiveri, Matti Mintz, Jörg Kramer, Paul Verschure

We present a biophysically constrained cerebellar model of classical conditioning, implemented using a neuromorphic analog VLSI (aVLSI) chip. Like its biological counterpart, our cerebellar model is able to control adaptive behavior by predicting the precise timing of events. Here we describe the functionality of the chip and present its learning performance, as evaluated in simulated conditioning experiments at the circuit level and in behavioral experiments using a mobile robot. We show that this aVLSI model supports the acquisition and extinction of adaptively timed conditioned responses under real-world conditions with ultra-low power consumption.

****

Pictorial Structures for Molecular Modeling: Interpreting Density Maps

Frank Dimaio, George Phillips, Jude Shavlik

X-ray crystallography is currently the most common way protein structures are elucidated. One of the most time-consuming steps in the crystallographic process is interpretation of the electron density map, a task that involves finding patterns in a three-dimensional picture of a protein. This paper describes DEFT (DEFormable Template), an algorithm using pictorial structures to build a flexible protein model from the protein's amino-acid sequence. Matching this pictorial structure into the density map is a way of automating density-map interpretation. Also described are several extensions to the pictorial structure matching algorithm necessary for this automated interpretation. DEFT is tested on a set of density maps ranging from 2 to 4Å resolution, producing root- mean-squared errors ranging from 1.38 to 1.84Å.

****

Support Vector Classification with Input Data Uncertainty

Jinbo Bi, Tong Zhang

This paper investigates a new learning model in which the input data is corrupted with noise. We present a general statistical framework to tackle this problem. Based on the statistical reasoning, we propose a novel formulation of support vector classi■cation, which allows uncer- tainty in input data. We derive an intuitive geometric interpretation of the proposed formulation, and develop algorithms to ef■ciently solve it. Empirical results are included to show that the newly formed method is superior to the standard SVM for problems with noisy input.

****

Planning for Markov Decision Processes with Sparse Stochasticity

Maxim Likhachev, Sebastian Thrun, Geoffrey J. Gordon

Planning algorithms designed for deterministic worlds, such as A* search, usually run much faster than algorithms designed for worlds with uncertain action outcomes, such as value iteration. Real-world planning problems often exhibit uncertainty, which forces us to use the slower algorithms to solve them. Many real-world planning problems exhibit sparse uncertainty: there are long sequences of deterministic actions which accomplish tasks like moving sensor platforms into place, interspersed with a small number of sensing actions which have uncertain outcomes. In this paper we describe a new planning algorithm, called MCP (short for MDP Compression Planning), which combines A* search with value iteration for solving Stochastic Shortest Path problem in MDPs with sparse stochasticity. We present experiments which show that MCP can run substantially faster than competing planners in domains with sparse uncertainty; these experiments are based on a simulation of a ground robot cooperating with a helicopter to fill in a partial map and move to a goal location.

****

Expectation Consistent Free Energies for Approximate Inference

Manfred Opper, Ole Winther

We propose a novel a framework for deriving approximations for intractable probabilistic models. This framework is based on a free energy (negative log marginal likelihood) and can be seen as a generalization of adaptive TAP [1, 2, 3] and expectation propagation (EP) [4, 5]. The free energy is constructed from two approximating distributions which encode different aspects of the intractable model such a single node constraints and couplings and are by construction consistent on a chosen set of moments. We test the framework on a difficult benchmark problem with binary variables on fully connected graphs and 2D grid graphs. We find good performance using sets of moments which either specify factorized nodes or a spanning tree on the nodes (structured approximation). Surprisingly, the Bethe approximation gives very inferior results even on grids.

****

Using Machine Learning to Break Visual Human Interaction Proofs (HIPs)

Kumar Chellapilla, Patrice Simard

Machine learning is often used to automatically solve human tasks. In this paper, we look for tasks where machine learning algorithms are not as good as humans with the hope of gaining insight into their current limitations. We studied various Human Interactive Proofs (HIPs) on the market, because they are systems designed to tell computers and humans apart by posing challenges presumably too hard for computers. We found that most HIPs are pure recognition tasks which can easily be broken using machine learning. The harder HIPs use a combination of segmentation and recognition tasks. From this observation, we found that building segmentation tasks is the most effective way to confuse machine learning algorithms. This has enabled us to build effective HIPs (which we deployed in MSN Passport), as well as design challenging segmentation tasks for machine learning algorithms.

****

Making Latin Manuscripts Searchable using gHMM's

Jaety Edwards, Yee Teh, Roger Bock, Michael Maire, Grace Vesom, David Forsyth

We describe a method that can make a scanned, handwritten mediaeval latin manuscript accessible to full text search. A generalized HMM is fitted, using transcribed latin to obtain a transition model and one example each of 22 letters to obtain an emission model. We show results for unigram, bigram and trigram models. Our method transcribes 25 pages of a manuscript of Terence with fair accuracy (75% of letters correctly transcribed). Search results are very strong; we use examples of vari- a

nt spellings to demonstrate that the search respects the ink of the doc-
ument. Furthermore, our model produces fair searches on a document fr
om which we obtained no training data.
****

Nonparametric Transforms of Graph Kernels for Semi-Supervised Learning
Jerry Zhu, Jaz Kandola, Zoubin Ghahramani, John Lafferty
We present an algorithm based on convex optimization for constructing kernels fo
r semi-supervised learning. The kernel matrices are derived from the spectral de
composition of graph Laplacians, and combine la- beled and unlabeled data in a s
ystematic fashion. Unlike previous work using diffusion kernels and Gaussian ran
dom ■eld kernels, a nonpara- metric kernel approach is presented that incorporat
es order constraints during optimization. This results in ■exible kernels and av
oids the need to choose among different parametric forms. Our approach relies on
a quadratically constrained quadratic program (QCQP), and is compu- tationally
feasible for large datasets. We evaluate the kernels on real datasets using supp
ort vector machines, with encouraging results.
****

Sparse Coding of Natural Images Using an Overcomplete Set of Limited Capacity Un
its
Eizaburo Doi, Michael Lewicki
It has been suggested that the primary goal of the sensory system is to
represent input in such a way as to reduce the high degree of redun-
dancy. Given a noisy neural representation, however, solely reducing
redundancy is not desirable, since redundancy is the only clue to reduce
the effects of noise. Here we propose a model that best balances redun-
dancy reduction and redundant representation. Like previous models, ou
r model accounts for the localized and oriented structure of simple c
ells, but it also predicts a different organization for the populatio
n. With noisy, limited-capacity units, the optimal representation bec
omes an overcom- plete, multi-scale representation, which, compared t
o previous models, is in closer agreement with physiological data. Th
ese results offer a new perspective on the expansion of the number of
neurons from retina to V1 and provide a theoretical model of incorpo
rating useful redundancy into efficient neural representations.
****

Optimal Information Decoding from Neuronal Populations with Specific Stimulus Se
lectivity
Marcelo Montemurro, Stefano Panzeri
A typical neuron in visual cortex receives most inputs from other cortical
neurons with a roughly similar stimulus preference. Does this arrange-
ment of inputs allow efficient readout of sensory information by the tar
- get cortical neuron? We address this issue by using simple modellin
g of neuronal population activity and information theoretic tools. We
find that efficient synaptic information transmission requires that
the tuning curve of the afferent neurons is approximately as wide as
the spread of stim- ulus preferences of the afferent neurons reaching
the target neuron. By meta analysis of neurophysiological data we fo
und that this is the case for cortico-cortical inputs to neurons in v
isual cortex. We suggest that the organization of V1 cortico-cortical
synaptic inputs allows optimal in- formation transmission.
****

Implicit Wiener Series for Higher-Order Image Analysis
Matthias Franz, Bernhard Schölkopf
The computation of classical higher-order statistics such as higher-order
moments or spectra is difficult for images due to the huge number of
terms to be estimated and interpreted. We propose an alternative ap- p
roach in which multiplicative pixel interactions are described by a se-
ries of Wiener functionals. Since the functionals are estimated implicitly
via polynomial kernels, the combinatorial explosion associated with the
classical higher-order statistics is avoided. First results show that imag

e structures such as lines or corners can be predicted correctly, and t
hat pixel interactions up to the order of five play an important role i
n natural images.
****
Following Curved Regularized Optimization Solution Paths
Saharon Rosset
Regularization plays a central role in the analysis of modern data, where
non-regularized fitting is likely to lead to over-fitted models, useless
for both prediction and interpretation. We consider the design of i
ncremen- tal algorithms which follow paths of regularized solutions
, as the regu- larization varies. These approaches often result in
methods which are both efficient and highly flexible. We suggest a
general path-following algorithm based on second-order approximatio
ns, prove that under mild conditions it remains "very close" to the
path of optimal solutions and illustrate it with examples.
****
Learning, Regularization and Ill-Posed Inverse Problems
Lorenzo Rosasco, Andrea Caponnetto, Ernesto Vito, Francesca Odone, Umberto Giova
nnini
Many works have shown that strong connections relate learning from ex-
amples to regularization techniques for ill-posed inverse problems. Nev-
ertheless by now there was no formal evidence neither that learning from
examples could be seen as an inverse problem nor that theoretical results
in learning theory could be independently derived using tools from re
g- ularization theory. In this paper we provide a positive answer to b
oth questions. Indeed, considering the square loss, we translate the l
earning problem in the language of regularization theory and show that
consis- tency results and optimal regularization parameter choice can
be derived by the discretization of the corresponding inverse problem
.
****
Density Level Detection is Classification
Ingo Steinwart, Don Hush, Clint Scovel
We show that anomaly detection can be interpreted as a binary classifi-
cation problem. Using this interpretation we propose a support vector
machine (SVM) for anomaly detection. We then present some theoret- ical
results which include consistency and learning rates. Finally, we expe
rimentally compare our SVM with the standard one-class SVM.
****
Learning Syntactic Patterns for Automatic Hypernym Discovery
Rion Snow, Daniel Jurafsky, Andrew Ng
Semantic taxonomies such as WordNet provide a rich source of knowl-
edge for natural language processing applications, but are expensive to
build, maintain, and extend. Motivated by the problem of automatically
constructing and extending such taxonomies, in this paper we present a
new algorithm for automatically learning hypernym (is-a) relations from
text. Our method generalizes earlier work that had relied on using smal
l
numbers of hand-crafted regular expression patterns to identify hyper-
nym pairs. Using "dependency path" features extracted from parse trees,
we introduce a general-purpose formalization and generalization of thes
e
patterns. Given a training set of text containing known hypernym pairs,
our algorithm automatically extracts useful dependency paths and applie
s
them to new corpora to identify novel pairs. On our evaluation task (de
-
termining whether two nouns in a news article participate in a hypernym
relationship), our automatically extracted database of hypernyms attain
s

both higher precision and higher recall than WordNet.
****
Joint Tracking of Pose, Expression, and Texture using Conditionally Gaussian Filters
Tim Marks, J. Roddey, Javier Movellan, John Hershey
We present a generative model and stochastic filtering algorithm for simultaneous tracking of 3D position and orientation, non-rigid motion, object texture, and background texture using a single camera. We show that the solution to this problem is formally equivalent to stochastic filtering of conditionally Gaussian processes, a problem for which well known approaches exist [3, 8]. We propose an approach based on Monte Carlo sampling of the nonlinear component of the process (object motion) and exact filtering of the object and background textures given the sampled motion. The smoothness of image sequences in time and space is exploited by using Laplace's method to generate proposal distributions for importance sampling [7]. The resulting inference algorithm encompasses both optic flow and template-based tracking as special cases, and elucidates the conditions under which these methods are optimal. We demonstrate an application of the system to 3D non-rigid face tracking.
****
Multiple Alignment of Continuous Time Series
Jennifer Listgarten, Radford Neal, Sam Roweis, Andrew Emili
Multiple realizations of continuous-valued time series from a stochastic process often contain systematic variations in rate and amplitude. To leverage the information contained in such noisy replicate sets, we need to align them in an appropriate way (for example, to allow the data to be properly combined by adaptive averaging). We present the Continuous Profile Model (CPM), a generative model in which each observed time series is a non-uniformly subsampled version of a single latent trace, to which local rescaling and additive noise are applied. After unsupervised training, the learned trace represents a canonical, high resolution fusion of all the replicates. As well, an alignment in time and scale of each observation to this trace can be found by inference in the model. We apply CPM to successfully align speech signals from multiple speakers and sets of Liquid Chromatography-Mass Spectrometry proteomic data.
****
Hierarchical Clustering of a Mixture Model
Jacob Goldberger, Sam Roweis
In this paper we propose an efficient algorithm for reducing a large mixture of Gaussians into a smaller mixture while still preserving the component structure of the original model; this is achieved by clustering (grouping) the components. The method minimizes a new, easily computed distance measure between two Gaussian mixtures that can be motivated from a suitable stochastic model and the iterations of the algorithm use only the model parameters, avoiding the need for explicit resampling of datapoints. We demonstrate the method by performing hierarchical clustering of scenery images and handwritten digits.
****
Hierarchical Bayesian Inference in Networks of Spiking Neurons
Rajesh PN Rao
There is growing evidence from psychophysical and neurophysiological studies that the brain utilizes Bayesian principles for inference and decision making. An important open question is how Bayesian inference for arbitrary graphical models can be implemented in networks of spiking neurons. In this paper, we show that recurrent networks of noisy integrate-and-fire neurons can perform approximate Bayesian inference for dynamic and hierarchical graphical models. The membrane potential dynamics of neurons is used to implement belief propagation in the log domain

. The spiking probability of a neuron is shown to approximate the poste
rior probability of the preferred state encoded by the neuron, given pa
st inputs. We illustrate the model using two examples: (1) a motion de-
tection network in which the spiking probability of a direction-selective
neuron becomes proportional to the posterior probability of motion in
a preferred direction, and (2) a two-level hierarchical network that pro-
duces attentional effects similar to those observed in visual cortical are
as V2 and V4. The hierarchical model offers a new Bayesian interpretati
on of attentional modulation in V2 and V4.

****

Efficient Kernel Discriminant Analysis via QR Decomposition

Tao Xiong, Jieping Ye, Qi Li, Ravi Janardan, Vladimir Cherkassky

Linear Discriminant Analysis (LDA) is a well-known method for fea- tur
e extraction and dimension reduction. It has been used widely in many
applications such as face recognition. Recently, a novel LDA algo- rit
hm based on QR Decomposition, namely LDA/QR, has been proposed, which
is competitive in terms of classification accuracy with other LDA algo
rithms, but it has much lower costs in time and space. However, LDA/QR
is based on linear projection, which may not be suitable for data wit
h nonlinear structure. This paper first proposes an algorithm called K
DA/QR, which extends the LDA/QR algorithm to deal with nonlin- ear dat
a by using the kernel operator. Then an efficient approximation of KDA
/QR called AKDA/QR is proposed. Experiments on face image data show th
at the classification accuracy of both KDA/QR and AKDA/QR are competit
ive with Generalized Discriminant Analysis (GDA), a gen- eral kernel d
iscriminant analysis algorithm, while AKDA/QR has much lower time and
space costs.

****

Semi-parametric Exponential Family PCA

Sajama Sajama, Alon Orlitsky

We present a semi-parametric latent variable model based technique for
density modelling, dimensionality reduction and visualization. Unlike
previous methods, we estimate the latent distribution non-parametrically
which enables us to model data generated by an underlying low dimen-
sional, multimodal distribution. In addition, we allow the components
of latent variable models to be drawn from the exponential family which
makes the method suitable for special data types, for example binary
or count data. Simulations on real valued, binary and count data show
fa- vorable comparison to other related schemes both in terms of sep
arating different populations and generalization to unseen samples.

****

Self-Tuning Spectral Clustering

Lihi Zelnik-manor, Pietro Perona

We study a number of open issues in spectral clustering: (i) Selecting the appro
priate scale of analysis, (ii) Handling multi-scale data, (iii) Cluster- ing wit
h irregular background clutter, and, (iv) Finding automatically the number of gr
oups. We ■rst propose that a ‘local’ scale should be used to compute the af■nity
between each pair of points. This local scaling leads to better clustering espe
cially when the data includes multiple scales and when the clusters are placed w
ithin a cluttered background. We further suggest exploiting the structure of the
eigenvectors to infer automatically the number of groups. This leads to a new a
lgorithm in which the ■nal randomly initialized k-means stage is eliminated.

****

The Entire Regularization Path for the Support Vector Machine

Saharon Rosset, Robert Tibshirani, Ji Zhu, Trevor Hastie

In this paper we argue that the choice of the SVM cost parameter can be
critical. We then derive an algorithm that can fit the entire path of SVM
solutions for every value of the cost parameter, with essentially the same
computational cost as fitting one SVM model.

****

The Laplacian PDF Distance: A Cost Function for Clustering in a Kernel Feature S
pace
Robert Jenssen, Deniz Erdogmus, Jose Principe, Torbjørn Eltoft
A new distance measure between probability density functions (pdfs) is
introduced, which we refer to as the Laplacian pdf dis- tance. The La
placian pdf distance exhibits a remarkable connec- tion to Mercer kern
el based learning theory via the Parzen window technique for density e
stimation. In a kernel feature space defined by the eigenspectrum of t
he Laplacian data matrix, this pdf dis- tance is shown to measure the
cosine of the angle between cluster mean vectors. The Laplacian data m
atrix, and hence its eigenspec- trum, can be obtained automatically ba
sed on the data at hand, by optimal Parzen window selection. We show t
hat the Laplacian pdf distance has an interesting interpretation as a
risk function connected to the probability of error.
****
Experts in a Markov Decision Process
Eyal Even-dar, Sham M. Kakade, Yishay Mansour
We consider an MDP setting in which the reward function is allowed to
change during each time step of play (possibly in an adversarial manner),
yet the dynamics remain fixed. Similar to the experts setting, we addres
s the question of how well can an agent do when compared to the rewa
rd achieved under the best stationary policy over time. We provide e
fficient algorithms, which have regret bounds with no dependence on
the size of state space. Instead, these bounds depend only on a cert
ain horizon time of the process and logarithmically on the number of
actions. We also show that in the case that the dynamics change ove
r time, the problem becomes computationally hard.
****
Neighbourhood Components Analysis
Jacob Goldberger, Geoffrey E. Hinton, Sam Roweis, Russ R. Salakhutdinov
In this paper we propose a novel method for learning a Mahalanobis
distance measure to be used in the KNN classification algorithm. The
algorithm directly maximizes a stochastic variant of the leave-one-out
KNN score on the training set. It can also learn a low-dimensional lin-
ear embedding of labeled data that can be used for data visualization
and fast classification. Unlike other methods, our classification
model is non-parametric, making no assumptions about the shape of t
he class distributions or the boundaries between them. The performa
nce of the method is demonstrated on several data sets, both for me
tric learning and linear dimensionality reduction.
****
Learning Gaussian Process Kernels via Hierarchical Bayes
Anton Schwaighofer, Volker Tresp, Kai Yu
We present a novel method for learning with Gaussian process regres-
sion in a hierarchical Bayesian framework. In a first step, kernel matri-
ces on a fixed set of input points are learned from data using a simple
and efficient EM algorithm. This step is nonparametric, in that it does
not require a parametric form of covariance function. In a second ste
p, kernel functions are fitted to approximate the learned covariance
matrix using a generalized Nystrom method, which results in a complex
, data driven kernel. We evaluate our approach as a recommendation en
gine for art images, where the proposed hierarchical Bayesian method
leads to excellent prediction performance.
****
Nonlinear Blind Source Separation by Integrating Independent Component Analysis
and Slow Feature Analysis
Tobias Blaschke, Laurenz Wiskott
In contrast to the equivalence of linear blind source separation and linear inde
pendent component analysis it is not possible to recover the origi- nal source s
ignal from some unknown nonlinear transformations of the sources using only the

independence assumption. Integrating the ob- jectives of statistical independenc
e and temporal slowness removes this indeterminacy leading to a new method for n
onlinear blind source sepa- ration. The principle of temporal slowness is adopte
d from slow feature analysis, an unsupervised method to extract slowly varying f
eatures from a given observed vectorial signal. The performance of the algorithm
 is demonstrated on nonlinearly mixed speech data.
****

Solitaire: Man Versus Machine
Xiang Yan, Persi Diaconis, Paat Rusmevichientong, Benjamin Roy
In this paper, we use the rollout method for policy improvement to an-
 alyze a version of Klondike solitaire. This version, sometimes called
 thoughtful solitaire, has all cards revealed to the player, but then follows
        the usual Klondike rules. A strategy that we establish, using iterated r
oll-   outs, wins about twice as many games on average as an expert huma
n   player does.
****

A Machine Learning Approach to Conjoint Analysis
Olivier Chapelle, Zaïd Harchaoui
Choice-based conjoint analysis builds models of consumer preferences
over products with answers gathered in questionnaires. Our main goal is
   to bring tools from the machine learning community to solve this prob-
     lem more efficiently. Thus, we propose two algorithms to quickly and
     accurately estimate consumer preferences.
****

Intrinsically Motivated Reinforcement Learning
Nuttapong Chentanez, Andrew Barto, Satinder Singh
Psychologists call behavior intrinsically motivated when it is engaged in
      for its own sake rather than as a step toward solving a specific problem
           of clear practical value. But what we learn during intrinsically moti
vated   behavior is essential for our development as competent autonom
ous en-   tities able to efficiently solve a wide range of practical p
roblems as they   arise. In this paper we present initial results from
 a computational study   of intrinsically motivated reinforcement lear
ning aimed at allowing arti-   ficial agents to construct and extend h
ierarchies of reusable skills that are   needed for competent autonomy
.
****

Joint Probabilistic Curve Clustering and Alignment
Scott Gaffney, Padhraic Smyth
Clustering and prediction of sets of curves is an important problem in
many areas of science and engineering. It is often the case that curves
 tend to be misaligned from each other in a continuous manner, either in
  space (across the measurements) or in time. We develop a probabilistic
  framework that allows for joint clustering and continuous alignment of
  sets of curves in curve space (as opposed to a fixed-dimensional feature-
     vector space). The proposed methodology integrates new probabilistic
   alignment models with model-based curve clustering algorithms. The   p
robabilistic approach allows for the derivation of consistent EM learn-
 ing algorithms for the joint clustering-alignment problem. Experimental
  results are shown for alignment of human growth data, and joint cluster-
    ing and alignment of gene expression time-course data.
****

Triangle Fixing Algorithms for the Metric Nearness Problem
Suvrit Sra, Joel Tropp, Inderjit Dhillon
Various problems in machine learning, databases, and statistics involve
 pairwise distances among a set of objects. It is often desirable for these
     distances to satisfy the properties of a metric, especially the triangle in
-          equality. Applications where metric data is useful include clustering
,   classification, metric-based indexing, and approximation algorithms f
or   various graph problems. This paper presents the Metric Nearness Prob

- lem: Given a dissimilarity matrix, find the "nearest" matrix of distances that satisfy the triangle inequalities. For p nearness measures, this paper develops efficient triangle fixing algorithms that compute globally optimal solutions by exploiting the inherent structure of the problem. Empirically, the algorithms have time and storage costs that are linear in the number of triangle constraints. The methods can also be easily parallelized for additional speed.

****

Economic Properties of Social Networks

Sham M. Kakade, Michael Kearns, Luis E. Ortiz, Robin Pemantle, Siddharth Suri

We examine the marriage of recent probabilistic generative models for social networks with classical frameworks from mathematical economics. We are particularly interested in how the statistical structure of such networks influences global economic quantities such as price variation. Our findings are a mixture of formal analysis, simulation, and experiments on an international trade data set from the United Nations.

****

Beat Tracking the Graphical Model Way

Dustin Lang, Nando Freitas

We present a graphical model for beat tracking in recorded music. Using a probabilistic graphical model allows us to incorporate local information and global smoothness constraints in a principled manner. We evaluate our model on a set of varied and difficult examples, and achieve impressive results. By using a fast dual-tree algorithm for graphical model inference, our system runs in less time than the duration of the music being processed.

****

Large-Scale Prediction of Disulphide Bond Connectivity

Jianlin Cheng, Alessandro Vullo, Pierre Baldi

The formation of disulphide bridges among cysteines is an important feature of protein structures. Here we develop new methods for the prediction of disulphide bond connectivity. We first build a large curated data set of proteins containing disulphide bridges and then use 2-Dimensional Recursive Neural Networks to predict bonding probabilities between cysteine pairs. These probabilities in turn lead to a weighted graph matching problem that can be addressed efficiently. We show how the method consistently achieves better results than previous approaches on the same validation data. In addition, the method can easily cope with chains with arbitrary numbers of bonded cysteines. Therefore, it overcomes one of the major limitations of previous approaches restricting predictions to chains containing no more than 10 oxidized cysteines. The method can be applied both to situations where the bonded state of each cysteine is known or unknown, in which case bonded state can be predicted with 85% precision and 90% recall. The method also yields an estimate for the total number of disulphide bridges in each chain.

****

An Application of Boosting to Graph Classification

Taku Kudo, Eisaku Maeda, Yuji Matsumoto

This paper presents an application of Boosting for classifying labeled graphs, general structures for modeling a number of real-world data, such as chemical compounds, natural language texts, and bio sequences. The proposal consists of i) decision stumps that use subgraph as features, and ii) a Boosting algorithm in which subgraph-based decision stumps are used as weak learners. We also discuss the relation between our algorithm and SVMs with convolution kernels. Two experiments using natural language data and chemical compounds show that our method achieves comparable or even better performance than SVMs with convolution kernels as well as improves the testing efficiency.

****

Multi-agent Cooperation in Diverse Population Games

K. Wong, S. Lim, Z. Gao

We consider multi-agent systems whose agents compete for resources by striving to be in the minority group. The agents adapt to the environment by reinforcement learning of the preferences of the policies they hold. Diversity of preferences of policies is introduced by adding random bi- ases to the initial cumulative payoffs of their policies. We explain and provide evidence that agent cooperation becomes increasingly important when diversity increases. Analyses of these mechanisms yield excellent agreement with simulations over nine decades of data.

****

Modeling Nonlinear Dependencies in Natural Images using Mixture of Laplacian Distribution

Hyun Park, Te Lee

Capturing dependencies in images in an unsupervised manner is important for many image processing applications. We propose a new method for capturing nonlinear dependencies in images of natural scenes. This method is an extension of the linear Independent Component Analysis (ICA) method by building a hierarchical model based on ICA and mixture of Laplacian distribution. The model parameters are learned via an EM algorithm and it can accurately capture variance correlation and other high order structures in a simple manner. We visualize the learned variance structure and demonstrate applications to image segmentation and denoising.

****

Dependent Gaussian Processes

Phillip Boyle, Marcus Frean

Gaussian processes are usually parameterised in terms of their covariance functions. However, this makes it difficult to deal with multiple outputs, because ensuring that the covariance matrix is positive definite is problematic. An alternative formulation is to treat Gaussian processes as white noise sources convolved with smoothing kernels, and to parameterise the kernel instead. Using this, we extend Gaussian processes to handle multiple, coupled outputs.

****

Synchronization of neural networks by mutual learning and its application to cryptography

Einat Klein, Rachel Mislovaty, Ido Kanter, Andreas Ruttor, Wolfgang Kinzel

Two neural networks that are trained on their mutual output synchronize to an identical time dependant weight vector. This novel phenomenon can be used for creation of a secure cryptographic secret-key using a public channel. Several models for this cryptographic system have been suggested, and have been tested for their security under different sophisticated attack strategies. The most promising models are networks that involve chaos synchronization. The synchronization process of mutual learning is described analytically using statistical physics methods.

****

Semigroup Kernels on Finite Sets

Marco Cuturi, Jean-philippe Vert

Complex objects can often be conveniently represented by finite sets of simpler components, such as images by sets of patches or texts by bags of words. We study the class of positive definite (p.d.) kernels for two such objects that can be expressed as a function of the merger of their respective sets of components. We prove a general integral representation of such kernels and present two particular examples. One of them leads to a kernel for sets of points living in a space endowed itself with a positive definite kernel. We provide experimental results on a benchmark experiment of handwritten digits image classification which illustrate the validity of the approach.

****

A Hidden Markov Model for de Novo Peptide Sequencing

Bernd Fischer, Volker Roth, Jonas Grossmann, Sacha Baginsky, Wilhelm Gruissem, Franz Roos, Peter Widmayer, Joachim Buhmann

De novo Sequencing of peptides is a challenging task in proteome re- se
arch. While there exist reliable DNA-sequencing methods, the high- thro
ughput de novo sequencing of proteins by mass spectrometry is still an
open problem. Current approaches suffer from a lack in precision to det
ect mass peaks in the spectrograms. In this paper we present a novel me
thod for de novo peptide sequencing based on a hidden Markov model. Exp
eriments effectively demonstrate that this new method signif- icantly o
utperforms standard approaches in matching quality.
****

Instance-Based Relevance Feedback for Image Retrieval
Giorgio Gia\-cin\-to, Fabio Roli
High retrieval precision in content-based image retrieval can be
attained by adopting relevance feedback mechanisms. These mec
hanisms require that the user judges the quality of the results of the
query by marking all the retrieved images as being either r
elevant or not. Then, the search engine exploits this information to a
dapt the search to better meet user's needs. At present, the vast
majority of proposed relevance feedback mechanisms are fo
rmulated in terms of search model that has to be optimized. Such an o
ptimization involves the modification of some search parameters
so that the nearest neighbor of the query vector contains the largest
number of relevant images. In this paper, a different approa
ch to relevance feedback is proposed. After the user provides
the first feedback, following retrievals are not based on k- nn search
, but on the computation of a relevance score for each image
of the database. This score is computed as a function of two distance
s, namely the distance from the nearest non-relevant image and the d
istance from the nearest relevant one. Images are then ranked
according to this score and the top k images are displayed.
Reported results on three image data sets show that the proposed
mechanism outperforms other state-of-the-art relevance feedback
mechanisms.
****

Learning Preferences for Multiclass Problems
Fabio Aiolli, Alessandro Sperduti
Many interesting multiclass problems can be cast in the general frame-
work of label ranking defined on a given set of classes. The evaluation
for such a ranking is generally given in terms of the number of violated
order constraints between classes. In this paper, we propose the Prefer-
ence Learning Model as a unifying framework to model and solve a large
class of multiclass problems in a large margin perspective. In addition,
an original kernel-based method is proposed and evaluated on a ranking
dataset with state-of-the-art results.
****

Mass Meta-analysis in Talairach Space
Finn Nielsen
We provide a method for mass meta-analysis in a neuroinformatics datab
ase containing stereotaxic Talairach coordinates from neu- roimaging e
xperiments. Database labels are used to group the in- dividual experim
ents, e.g., according to cognitive function, and the consistent patter
n of the experiments within the groups are de- termined. The method vo
xelizes each group of experiments via a kernel density estimation, for
ming probability density volumes. The values in the probability densit
y volumes are compared to null-hypothesis distributions generated by r
esamplings from the entire unlabeled set of experiments, and the dista
nces to the null- hypotheses are used to sort the voxels across groups
of experi- ments. This allows for mass meta-analysis, with the constr
uction of a list with the most prominent associations between brain ar
- eas and group labels. Furthermore, the method can be used for
functional labeling of voxels.

****
Result Analysis of the NIPS 2003 Feature Selection Challenge
Isabelle Guyon, Steve Gunn, Asa Ben-Hur, Gideon Dror
The NIPS 2003 workshops included a feature selection competi- tion org anized by the authors. We provided participants with five datasets fro m different application domains and called for classifica- tion result s using a minimal number of features. The competition took place over a period of 13 weeks and attracted 78 research groups. Participants we re asked to make on-line submissions on the validation and test sets, with performance on the validation set being presented immediately to the participant and performance on the test set presented to the parti cipants at the workshop. In total 1863 entries were made on the valida tion sets during the development period and 135 entries on all test se ts for the final competition. The winners used a combination of Bayesi an neu- ral networks with ARD priors and Dirichlet diffusion trees. Ot her top entries used a variety of methods for feature selection, which combined filters and/or wrapper or embedded methods using Ran- dom Forests, kernel methods, or neural networks as a classification engine. The results of the benchmark (including the predictions made by the participants and the features they selected) and the scoring s oftware are publicly available. The benchmark is available at www.nips fsc.ecs.soton.ac.uk for post-challenge submissions to stimulate furthe r research.
****
Sub-Microwatt Analog VLSI Support Vector Machine for Pattern Classification and Sequence Estimation
Shantanu Chakrabartty, Gert Cauwenberghs
An analog system-on-chip for kernel-based pattern classification and se- quence estimation is presented. State transition probabilities conditioned on input data are generated by an integrated support vector machin e. Dot product based kernels and support vector coefficients are im plemented in analog programmable floating gate translinear circuits , and probabil- ities are propagated and normalized using sub-thres hold current-mode circuits. A 14-input, 24-state, and 720-support v ector forward decod- ing kernel machine is integrated on a 3mm3mm c hip in 0.5m CMOS technology. Experiments with the processor trained for speaker verifica- tion and phoneme sequence estimation demonst rate real-time recognition accuracy at par with floating-point soft ware, at sub-microwatt power.
****
Maximum Margin Clustering
Linli Xu, James Neufeld, Bryce Larson, Dale Schuurmans
We propose a new method for clustering based on finding maximum mar- gi n hyperplanes through data. By reformulating the problem in terms of th e implied equivalence relation matrix, we can pose the problem as a con vex integer program. Although this still yields a difficult com- putati onal problem, the hard-clustering constraints can be relaxed to a soft- clustering formulation which can be feasibly solved with a semidef- ini te program. Since our clustering technique only depends on the data thr ough the kernel matrix, we can easily achieve nonlinear clusterings in the same manner as spectral clustering. Experimental results show that our maximum margin clustering technique often obtains more accurate res ults than conventional clustering methods. The real benefit of our ap- proach, however, is that it leads naturally to a semi-supervised training method for support vector machines. By maximizing the margin simul- taneously on labeled and unlabeled training data, we achieve state of the art performance by using a single, integrated learning principle.
****
On the Adaptive Properties of Decision Trees
Clayton Scott, Robert Nowak

Decision trees are surprisingly adaptive in three important respects: They automatically (1) adapt to favorable conditions near the Bayes decision boundary; (2) focus on data distributed on lower dimensional manifolds; (3) reject irrelevant features. In this paper we examine a decision tree based on dyadic splits that adapts to each of these conditions to achieve minimax optimal rates of convergence. The proposed classifier is the first known to achieve these optimal rates while being practical and implementable.

****

Kernel Methods for Implicit Surface Modeling

Joachim Giesen, Simon Spalinger, Bernhard Schölkopf

We describe methods for computing an implicit model of a hypersurface that is given only by a finite sampling. The methods work by mapping the sample points into a reproducing kernel Hilbert space and then determining regions in terms of hyperplanes.

****

Neural Network Computation by In Vitro Transcriptional Circuits

Jongmin Kim, John Hopfield, Erik Winfree

The structural similarity of neural networks and genetic regulatory networks to digital circuits, and hence to each other, was noted from the very beginning of their study [1, 2]. In this work, we propose a simple biochemical system whose architecture mimics that of genetic regulation and whose components allow for in vitro implementation of arbitrary circuits. We use only two enzymes in addition to DNA and RNA molecules: RNA polymerase (RNAP) and ribonuclease (RNase). We develop a rate equation for in vitro transcriptional networks, and derive a correspondence with general neural network rate equations [3]. As proof-of-principle demonstrations, an associative memory task and a feedforward network computation are shown by simulation. A difference between the neural network and biochemical models is also highlighted: global coupling of rate equations through enzyme saturation can lead to global feedback regulation, thus allowing a simple network without explicit mutual inhibition to perform the winner-take-all computation. Thus, the full complexity of the cell is not necessary for biochemical computation: a wide range of functional behaviors can be achieved with a small set of biochemical components.

****

Similarity and Discrimination in Classical Conditioning: A Latent Variable Account

Aaron C. Courville, Nathaniel Daw, David Touretzky

We propose a probabilistic, generative account of configural learning phenomena in classical conditioning. Configural learning experiments probe how animals discriminate and generalize between patterns of simultaneously presented stimuli (such as tones and lights) that are differentially predictive of reinforcement. Previous models of these issues have been successful more on a phenomenological than an explanatory level: they reproduce experimental findings but, lacking formal foundations, provide scant basis for understanding why animals behave as they do. We present a theory that clarifies seemingly arbitrary aspects of previous models while also capturing a broader set of data. Key patterns of data, e.g. concerning animals' readiness to distinguish patterns with varying degrees of overlap, are shown to follow from statistical inference.

****

Boosting on Manifolds: Adaptive Regularization of Base Classifiers

Ligen Wang, Balázs Kégl

In this paper we propose to combine two powerful ideas, boosting and manifold learning. On the one hand, we improve ADABOOST by incorporating knowledge on the structure of the data into base classifier design and selection. On the other hand, we use ADABOOST's efficient learning mechanism to significantly improve supervised and semi-supervised

algorithms proposed in the context of manifold learning. Beside the specific manifold-based penalization, the resulting algorithm also accommodates the boosting of a large family of regularized learning algorithms.

****

Surface Reconstruction using Learned Shape Models

Jan Solem, Fredrik Kahl

We consider the problem of geometrical surface reconstruction from one or several images using learned shape models. While humans can effortlessly retrieve 3D shape information, this inverse problem has turned out to be difficult to perform automatically. We introduce a framework based on level set surface reconstruction and shape models for achieving this goal. Through this merging, we obtain an efficient and robust method for reconstructing surfaces of an object category of interest. The shape model includes surface cues such as point, curve and silhouette features. Based on ideas from Active Shape Models, we show how both the geometry and the appearance of these features can be modelled consistently in a multi-view context. The complete surface is obtained by evolving a level set driven by a PDE, which tries to fit the surface to the inferred 3D features. In addition, an a priori 3D surface model is used to regularize the solution, in particular, where surface features are sparse. Experiments are demonstrated on a database of real face images.

****

Reducing Spike Train Variability: A Computational Theory Of Spike-Timing Dependent Plasticity

Sander Bohte, Michael C. Mozer

Experimental studies have observed synaptic potentiation when a presynaptic neuron fires shortly before a postsynaptic neuron, and synaptic depression when the presynaptic neuron fires shortly after. The dependence of synaptic modulation on the precise timing of the two action potentials is known as spike-timing dependent plasticity or STDP. We derive STDP from a simple computational principle: synapses adapt so as to minimize the postsynaptic neuron's variability to a given presynaptic input, causing the neuron's output to become more reliable in the face of noise. Using an entropy-minimization objective function and the biophysically realistic spike-response model of Gerstner (2001), we simulate neurophysiological experiments and obtain the characteristic STDP curve along with other phenomena including the reduction in synaptic plasticity as synaptic efficacy increases. We compare our account to other efforts to derive STDP from computational principles, and argue that our account provides the most comprehensive coverage of the phenomena. Thus, reliability of neural response in the face of noise may be a key goal of cortical adaptation.

****

Maximum Likelihood Estimation of Intrinsic Dimension

Elizaveta Levina, Peter Bickel

We propose a new method for estimating intrinsic dimension of a dataset derived by applying the principle of maximum likelihood to the distances between close neighbors. We derive the estimator by a Poisson process approximation, assess its bias and variance theo- retically and by simulations, and apply it to a number of simulated and real datasets. We also show it has the best overall performance compared with two other intrinsic dimension estimators.

****

Distributed Information Regularization on Graphs

Adrian Corduneanu, Tommi Jaakkola

We provide a principle for semi-supervised learning based on optimizing the rate of communicating labels for unlabeled points with side information. The side information is expressed in terms of identities of sets of points or regions with the purpose of biasing the labels in each region to be the same. The resulting regularization objective is convex, has a unique solution, and the solution can be found with a pair of local prop

- agation operations on graphs induced by the regions. We analyze the properties of the algorithm and demonstrate its performance on document classification tasks.

****

The Convergence of Contrastive Divergences

Alan L. Yuille

This paper analyses the Contrastive Divergence algorithm for learning statistical parameters. We relate the algorithm to the stochastic approximation literature. This enables us to specify conditions under which the algorithm is guaranteed to converge to the optimal solution (with probability 1). This includes necessary and sufficient conditions for the solution to be unbiased.

****

Log-concavity Results on Gaussian Process Methods for Supervised and Unsupervised Learning

Liam Paninski

Log-concavity is an important property in the context of optimization, Laplace approximation, and sampling; Bayesian methods based on Gaussian process priors have become quite popular recently for classification, regression, density estimation, and point process intensity estimation. Here we prove that the predictive densities corresponding to each of these applications are log-concave, given any observed data. We also prove that the likelihood is log-concave in the hyperparameters controlling the mean function of the Gaussian prior in the density and point process intensity estimation cases, and the mean, covariance, and observation noise parameters in the classification and regression cases; this result leads to a useful parameterization of these hyperparameters, indicating a suitably large class of priors for which the corresponding maximum a posteriori problem is log-concave.

****

Incremental Algorithms for Hierarchical Classification

Nicolò Cesa-bianchi, Claudio Gentile, Andrea Tironi, Luca Zaniboni

We study the problem of hierarchical classification when labels corresponding to partial and/or multiple paths in the underlying taxonomy are allowed. We introduce a new hierarchical loss function, the H-loss, implementing the simple intuition that additional mistakes in the subtree of a mistaken class should not be charged for. Based on a probabilistic data model introduced in earlier work, we derive the Bayes-optimal classifier for the H-loss. We then empirically compare two incremental approximations of the Bayes-optimal classifier with a flat SVM classifier and with classifiers obtained by using hierarchical versions of the Perceptron and SVM algorithms. The experiments show that our simplest incremental approximation of the Bayes-optimal classifier performs, after just one training epoch, nearly as well as the hierarchical SVM classifier (which performs best). For the same incremental algorithm we also derive an H-loss bound showing, when data are generated by our probabilistic data model, exponentially fast convergence to the H-loss of the hierarchical classifier based on the true model parameters.

****

On Semi-Supervised Classification

Balaji Krishnapuram, David Williams, Ya Xue, Lawrence Carin, Mário Figueiredo, Alexander Hartemink

A graph-based prior is proposed for parametric semi-supervised classification. The prior utilizes both labelled and unlabelled data; it also integrates features from multiple views of a given sample (e.g., multiple sensors), thus implementing a Bayesian form of co-training. An EM algorithm for training the classifier automatically adjusts the tradeoff between the contributions of: (a) the labelled data; (b) the unlabelled data; and (c) the co-training information. Active label query

selection is per- formed using a mutual information based criterion t
hat explicitly uses the unlabelled data and the co-training informati
on. Encouraging results are presented on public benchmarks and on mea
sured data from single and multiple sensors.
****

Joint MRI Bias Removal Using Entropy Minimization Across Images
Erik Learned-miller, Parvez Ahammad
The correction of bias in magnetic resonance images is an important pro
blem in medical image processing. Most previous approaches have used a
maximum likelihood method to increase the likelihood of the pix- els in
a single image by adaptively estimating a correction to the unknown im
age bias field. The pixel likelihoods are defined either in terms of a
pre-existing tissue model, or non-parametrically in terms of the image's
own pixel values. In both cases, the specific location of a pixel in the im-
age is not used to calculate the likelihoods. We suggest a new approach
in which we simultaneously eliminate the bias from a set of images of
the same anatomy, but from different patients. We use the statistics fro
m the same location across different images, rather than within an imag
e, to eliminate bias fields from all of the images simultaneously. The
method builds a "multi-resolution" non-parametric tissue model conditio
ned on image location while eliminating the bias fields associated with
the orig- inal image set. We present experiments on both synthetic and
real MR data sets, and present comparisons with other methods.
****

Chemosensory Processing in a Spiking Model of the Olfactory Bulb: Chemotopic Con
vergence and Center Surround Inhibition
Baranidharan Raman, Ricardo Gutierrez-osuna
This paper presents a neuromorphic model of two olfactory signal- proce
ssing primitives: chemotopic convergence of olfactory receptor neurons
, and center on-off surround lateral inhibition in the olfactory bulb.
A self-organizing model of receptor convergence onto glomeruli is use
d to generate a spatially organized map, an olfactory image. This map
serves as input to a lattice of spiking neurons with lateral connectio
ns. The dynamics of this recurrent network transforms the initial olfa
ctory image into a spatio-temporal pattern that evolves and stabilizes
into odor- and intensity-coding attractors. The model is validated us
ing experimental data from an array of temperature-modulated gas
sensors. Our results are consistent with recent neurobiological
findings on the antennal lobe of the honeybee and the locust.
****

Using Random Forests in the Structured Language Model
Peng Xu, Frederick Jelinek
In this paper, we explore the use of Random Forests (RFs) in the struc- tured la
nguage model (SLM), which uses rich syntactic information in predicting the next
word based on words already seen. The goal in this work is to construct RFs by
randomly growing Decision Trees (DTs) us- ing syntactic information and investig
ate the performance of the SLM modeled by the RFs in automatic speech recognitio
n. RFs, which were originally developed as classi■ers, are a combination of deci
sion tree classi■ers. Each tree is grown based on random training data sampled i
ndependently and with the same distribution for all trees in the forest, and a r
andom selection of possible questions at each node of the decision tree. Our app
roach extends the original idea of RFs to deal with the data sparseness problem
encountered in language modeling. RFs have been studied in the context of n-gram
language modeling and have been shown to generalize well to unseen data. We sho
w in this paper that RFs using syntactic information can also achieve better per
formance in both perplexity (PPL) and word error rate (WER) in a large vocabular
y speech recognition system, compared to a baseline that uses Kneser-Ney smoothi
ng.
****

Exploration-Exploitation Tradeoffs for Experts Algorithms in Reactive Environmen

ts
Daniela Farias, Nimrod Megiddo
A reactive environment is one that responds to the actions of an agent rather th an evolving obliviously. In reactive environments, experts algorithms must balance exploration and exploitation of experts more carefully th an in oblivious ones. In addition, a more subtle definition of a learn able value of an expert is required. A general exploration-exploitatio n experts method is presented along with a proper definition of value. The method is shown to asymptotically perform as well as the best ava ilable expert. Several variants are analyzed from the viewpoint of the exploration-exploitation tradeoff, including explore-then-exploit, polynomially vanishing exploration, constant-frequency exploration, and constant-s ize explo- ration phases. Complexity and performance bounds are proven .

****

Unsupervised Variational Bayesian Learning of Nonlinear Models
Antti Honkela, Harri Valpola
In this paper we present a framework for using multi-layer per- ceptron (MLP) ne tworks in nonlinear generative models trained by variational Bayesian learning. The nonlinearity is handled by linearizing it using a Gauss–Hermite quadrature a t the hidden neu- rons. This yields an accurate approximation for cases of large pos- terior variance. The method can be used to derive nonlinear coun- terparts for linear algorithms such as factor analysis, independent component/factor ana lysis and state-space models. This is demon- strated with a nonlinear factor ana lysis experiment in which even 20 sources can be estimated from a real world spe ech data set.

****

VDCBPI: an Approximate Scalable Algorithm for Large POMDPs
Pascal Poupart, Craig Boutilier
Existing algorithms for discrete partially observable Markov decision p rocesses can at best solve problems of a few thousand states due to two important sources of intractability: the curse of dimensionality and t he policy space complexity. This paper describes a new algorithm (V DCBPI) that mitigates both sources of intractability by combining the V alue Directed Compression (VDC) technique [13] with Bounded Pol- icy It eration (BPI) [14]. The scalability of VDCBPI is demonstrated on synthe tic network management problems with up to 33 million states.

****

A Generalized Bradley-Terry Model: From Group Competition to Individual Skill
Tzu-kuo Huang, Chih-jen Lin, Ruby Weng
The Bradley-Terry model for paired comparison has been popular in many areas. We propose a generalized version in which paired individual comparisons are extend ed to paired team comparisons. We introduce a simple algorithm with convergence proofs to solve the model and obtain individual skill. A useful application to m ulti-class probability estimates using error-correcting codes is demonstrated.

****

Rate- and Phase-coded Autoassociative Memory
Máté Lengyel, Peter Dayan
Areas of the brain involved in various forms of memory exhibit patterns of neural activity quite unlike those in canonical computational models. We show how to use well-founded Bayesian probabilistic autoassociative recall to derive biologically reasonable neuronal dynamics in recurrently coupled models, together with appropriate values for parameters such a s the membrane time constant and inhibition. We explicitly treat two c ases. One arises from a standard Hebbian learning rule, and involves a ctivity patterns that are coded by graded firing rates. The other aris es from a spike timing dependent learning rule, and involves patterns coded by the phase of spike times relative to a coherent local field p otential oscillation. Our model offers a new and more complete underst anding of how neural dynamics may support autoassociation.

****
Constraining a Bayesian Model of Human Visual Speed Perception
Alan A. Stocker, Eero Simoncelli
It has been demonstrated that basic aspects of human visual motion per-
ception are qualitatively consistent with a Bayesian estimation frame-
work, where the prior probability distribution on velocity favors slow
speeds. Here, we present a refined probabilistic model that can account
for the typical trial-to-trial variabilities observed in psychophysical speed
perception experiments. We also show that data from such experiments
can be used to constrain both the likelihood and prior functions of the
model. Specifically, we measured matching speeds and thresholds in a
two-alternative forced choice speed discrimination task. Parametric fits
to the data reveal that the likelihood function is well approximated by
a LogNormal distribution with a characteristic contrast-dependent vari-
ance, and that the prior distribution on velocity exhibits significant
ly heavier tails than a Gaussian, and approximately follows a power-law
function.
****
Efficient Kernel Machines Using the Improved Fast Gauss Transform
Changjiang Yang, Ramani Duraiswami, Larry S. Davis
The computation and memory required for kernel machines with N train- i
ng samples is at least O(N 2). Such a complexity is significant even for
moderate size problems and is prohibitive for large datasets. We present
an approximation technique based on the improved fast Gauss transform
to reduce the computation to O(N). We also give an error bound for the
approximation, and provide experimental results on the UCI datasets.
****
Responding to Modalities with Different Latencies
Fredrik Bissmarck, Hiroyuki Nakahara, Kenji Doya, Okihide Hikosaka
Motor control depends on sensory feedback in multiple modalities with different
latencies. In this paper we consider within the framework of re- inforcement lea
rning how different sensory modalities can be combined and selected for real-tim
e, optimal movement control. We propose an actor-critic architecture with multip
le modules, whose output are com- bined using a softmax function. We tested our
architecture in a simu- lation of a sequential reaching task. Reaching was initi
ally guided by visual feedback with a long latency. Our learning scheme allowed
the agent to utilize the somatosensory feedback with shorter latency when the ha
nd is near the experienced trajectory. In simulations with different latencies f
or visual and somatosensory feedback, we found that the agent depended more on f
eedback with shorter latency.
****
Two-Dimensional Linear Discriminant Analysis
Jieping Ye, Ravi Janardan, Qi Li
Linear Discriminant Analysis (LDA) is a well-known scheme for feature extraction
and dimension reduction. It has been used widely in many ap- plications involvi
ng high-dimensional data, such as face recognition and image retrieval. An intri
nsic limitation of classical LDA is the so-called singularity problem, that is,
it fails when all scatter matrices are singu- lar. A well-known approach to deal
with the singularity problem is to apply an intermediate dimension reduction st
age using Principal Com- ponent Analysis (PCA) before LDA. The algorithm, called
PCA+LDA, is used widely in face recognition. However, PCA+LDA has high costs in
time and space, due to the need for an eigen-decomposition involving the scatte
r matrices. In this paper, we propose a novel LDA algorithm, namely 2DLDA, which
stands for 2-Dimensional Linear Discriminant Analysis. 2DLDA over- comes the si
ngularity problem implicitly, while achieving ef■ciency. The key difference betw
een 2DLDA and classical LDA lies in the model for data representation. Classical
LDA works with vectorized representa- tions of data, while the 2DLDA algorithm
works with data in matrix representation. To further reduce the dimension by 2DL
DA, the combi- nation of 2DLDA and classical LDA, namely 2DLDA+LDA, is studied,
where LDA is preceded by 2DLDA. The proposed algorithms are ap- plied on face re

cognition and compared with PCA+LDA. Experiments show that 2DLDA and 2DLDA+LDA a chieve competitive recognition accuracy, while being much more ef■cient.
****

Assignment of Multiplicative Mixtures in Natural Images
Odelia Schwartz, Terrence J. Sejnowski, Peter Dayan
In the analysis of natural images, Gaussian scale mixtures (GSM) have been used to account for the statistics of (cid:2)lter responses, and to inspire hi- erarc hical cortical representational learning schemes. GSMs pose a crit- ical assignm ent problem, working out which (cid:2)lter responses were gen- erated by a commo n multiplicative factor. We present a new approach to solving this assignment pr oblem through a probabilistic extension to the basic GSM, and show how to perfor m inference in the model using Gibbs sampling. We demonstrate the ef(cid:2)cacy of the approach on both synthetic and image data.
****

Convergence and No-Regret in Multiagent Learning
Michael Bowling
Learning in a multiagent system is a challenging problem due to two key factors. First, if other agents are simultaneously learning then the environment is no longer stationary, thus undermining convergence guarantees. Second, learning is often susceptible to deception, where the other agents may be able to exploit a learner's particular dynamics. In the worst case, this could result in poorer performance than if the agent was not learning at all. These challenges are identifiable in the two most common evaluation criteria for multiagent learning algorithms: convergence and regret. Algorithms focusing on convergence or regret in isolation are numerous. In this paper, we seek to address both criteria in a single algorithm by introducing GIGA-WoLF, a learning algorithm for normal-form games. We prove the algorithm guarantees at most zero average regret, while demonstrating the algorithm converges in many situations of self-play. We prove convergence in a limited setting and give empirical results in a wider variety of situations. These results also suggest a third new learning criterion combining convergence and regret, which we call negative non-convergence regret (NNR).
****

Learning Efficient Auditory Codes Using Spikes Predicts Cochlear Filters
Evan Smith, Michael Lewicki
The representation of acoustic signals at the cochlear nerve must serve a wide range of auditory tasks that require exquisite sensitivity in both time and frequency. Lewicki (2002) demonstrated that many of the filtering properties of the cochlea could be explained in terms of efficient coding of natural sounds. This model, however, did not account for properties such as phase-locking or how sound could be encoded in terms of action potentials. Here, we extend this theoretical approach with algorithm for learning efficient auditory codes using a spiking population code. Here, we propose an algorithm for learning efficient auditory codes using a theoretical model for coding sound in terms of spikes. In this model, each spike encodes the precise time position and magnitude of a localized, time varying kernel function. By adapting the kernel functions to the statistics natural sounds, we show that, compared to conventional signal representations, the spike code achieves far greater coding efficiency. Furthermore, the inferred kernels show both striking similarities to measured cochlear filters and a similar bandwidth versus frequency dependence.
****

Active Learning for Anomaly and Rare-Category Detection
Dan Pelleg, Andrew Moore
We introduce a novel active-learning scenario in which a user wants to work with a learning algorithm to identify useful anomalies. These are distinguished from the traditional statistical definition of anomalies as outliers or merely ill-modeled points. Our distinction is that the useful-

ness of anomalies is categorized subjectively by the user. We make two additional assumptions. First, there exist extremely few useful anomalies to be hunted down within a massive dataset. Second, both useful and useless anomalies may sometimes exist within tiny classes of similar anomalies. The challenge is thus to identify "rare category" records in an unlabeled noisy set with help (in the form of class labels) from a human expert who has a small budget of datapoints that they are prepared to categorize. We propose a technique to meet this challenge, which assumes a mixture model fit to the data, but otherwise makes no assumptions on the particular form of the mixture components. This property promises wide applicability in real-life scenarios and for various statistical models. We give an overview of several alternative methods, highlighting their strengths and weaknesses, and conclude with a detailed empirical analysis. We show that our method can quickly zoom in on an anomaly set containing a few tens of points in a dataset of hundreds of thousands.

****

A Feature Selection Algorithm Based on the Global Minimization of a Generalization Error Bound

Dori Peleg, Ron Meir

A novel linear feature selection algorithm is presented based on the global minimization of a data-dependent generalization error bound. Feature selection and scaling algorithms often lead to non-convex optimization problems, which in many previous approaches were addressed through gradient descent procedures that can only guarantee convergence to a local minimum. We propose an alternative approach, whereby the global solution of the non-convex optimization problem is derived via an equivalent optimization problem. Moreover, the convex optimization task is reduced to a conic quadratic programming problem for which efficient solvers are available. Highly competitive numerical results on both artificial and real-world data sets are reported.

****

Semi-supervised Learning with Penalized Probabilistic Clustering

Zhengdong Lu, Todd Leen

While clustering is usually an unsupervised operation, there are circumstances in which we believe (with varying degrees of certainty) that items A and B should be assigned to the same cluster, while items A and C should not. We would like such pairwise relations to influence cluster assignments of out-of-sample data in a manner consistent with the prior knowledge expressed in the training set. Our starting point is probabilistic clustering based on Gaussian mixture models (GMM) of the data distribution. We express clustering preferences in the prior distribution over assignments of data points to clusters. This prior penalizes cluster assignments according to the degree with which they violate the preferences. We fit the model parameters with EM. Experiments on a variety of data sets show that PPC can consistently improve clustering results.

****

A Direct Formulation for Sparse PCA Using Semidefinite Programming

Alexandre D'aspremont, Laurent Ghaoui, Michael Jordan, Gert Lanckriet

We examine the problem of approximating, in the Frobenius-norm sense, a positive, semidefinite symmetric matrix by a rank-one matrix, with an upper bound on the cardinality of its eigenvector. The problem arises in the decomposition of a covariance matrix into sparse factors, and has wide applications ranging from biology to finance. We use a modification of the classical variational representation of the largest eigenvalue of a symmetric matrix, where cardinality is constrained, and derive a semidefinite programming based relaxation for our problem.

****

Stable adaptive control with online learning

H. Kim, Andrew Ng

Learning algorithms have enjoyed numerous successes in robotic control tasks. In problems with time-varying dynamics, online learning methods have also proved to be a powerful tool for automatically tracking and/or adapting to the changing circumstances. However, for safety-critical applications such as airplane flight, the adoption of these algorithms has been significantly hampered by their lack of safety, such as "stability," guarantees. Rather than trying to show difficult, a priori, stability guarantees for specific learning methods, in this paper we propose a method for "monitoring" the controllers suggested by the learning algorithm online, and rejecting controllers leading to instability. We prove that even if an arbitrary online learning method is used with our algorithm to control a linear dynamical system, the resulting system is stable.

****

The Rescorla-Wagner Algorithm and Maximum Likelihood Estimation of Causal Parameters

Alan L. Yuille

This paper analyzes generalization of the classic Rescorla-Wagner (R-W) learning algorithm and studies their relationship to Maximum Likelihood estimation of causal parameters. We prove that the parameters of two popular causal models, P and P C, can be learnt by the same generalized linear Rescorla-Wagner (GLRW) algorithm provided genericity conditions apply. We characterize the fixed points of these GLRW algorithms and calculate the fluctuations about them, assuming that the input is a set of i.i.d. samples from a fixed (unknown) distribution. We describe how to determine convergence conditions and calculate convergence rates for the GLRW algorithms under these conditions.

****

Contextual Models for Object Detection Using Boosted Random Fields

Antonio Torralba, Kevin P. Murphy, William Freeman

We seek to both detect and segment objects in images. To exploit both local image data as well as contextual information, we introduce Boosted Random Fields (BRFs), which uses Boosting to learn the graph structure and local evidence of a conditional random field (CRF). The graph structure is learned by assembling graph fragments in an additive model. The connections between individual pixels are not very informative, but by using dense graphs, we can pool information from large regions of the image; dense models also support efficient inference. We show how contextual information from other objects can improve detection performance, both in terms of accuracy and speed, by using a computational cascade. We apply our system to detect stuff and things in office and street scenes. 1 Introduction

****

An Auditory Paradigm for Brain-Computer Interfaces

N. Hill, Thomas Lal, Karin Bierig, Niels Birbaumer, Bernhard Schölkopf

Motivated by the particular problems involved in communicating with "locked-in" paralysed patients, we aim to develop a brain-computer interface that uses auditory stimuli. We describe a paradigm that allows a user to make a binary decision by focusing attention on one of two concurrent auditory stimulus sequences. Using Support Vector Machine classification and Recursive Channel Elimination on the independent components of averaged event-related potentials, we show that an untrained user's EEG data can be classified with an encouragingly high level of accuracy. This suggests that it is possible for users to modulate EEG signals in a single trial by the conscious direction of attention, well enough to be useful in BCI.

****

Worst-Case Analysis of Selective Sampling for Linear-Threshold Algorithms

Nicolò Cesa-bianchi, Claudio Gentile, Luca Zaniboni

We provide a worst-case analysis of selective sampling algorithms for learning l

inear threshold functions. The algorithms considered in this paper are Perceptro n-like algorithms, i.e., algorithms which can be ef■- ciently run in any reprodu cing kernel Hilbert space. Our algorithms ex- ploit a simple margin-based random ized rule to decide whether to query the current label. We obtain selective samp ling algorithms achieving on average the same bounds as those proven for their d eterministic coun- terparts, but using much fewer labels. We complement our theo retical ■ndings with an empirical comparison on two text categorization tasks. T he outcome of these experiments is largely predicted by our theoreti- cal result s: Our selective sampling algorithms tend to perform as good as the algorithms r eceiving the true label after each classi■cation, while observing in practice su bstantially fewer labels.

****

Markov Networks for Detecting Overalpping Elements in Sequence Data

Mark Craven, Joseph Bockhorst

Many sequential prediction tasks involve locating instances of pat- te rns in sequences. Generative probabilistic language models, such as hi dden Markov models (HMMs), have been successfully applied to many of t hese tasks. A limitation of these models however, is that they cannot naturally handle cases in which pattern instances overlap in arbitrary ways. We present an alternative approach, based on conditional Markov networks, that can naturally repre- sent arbitrarily overlapping elem ents. We show how to efficiently train and perform inference with thes e models. Experimental re- sults from a genomics domain show that our models are more accu- rate at locating instances of overlapping patter ns than are baseline models based on HMMs.

****

Co-Validation: Using Model Disagreement on Unlabeled Data to Validate Classifica tion Algorithms

Omid Madani, David Pennock, Gary Flake

In the context of binary classification, we define disagreement as a mea- sure of how often two independently-trained models differ in their clas- sification of unlabeled data. We explore the use of disagreement for error estimation and model selection. We call the procedure co-validation, since the two models effectively (in)validate one another by comparing results on unlabeled data, which we assume is relatively cheap and plen- tiful compared to labeled data. We show that per-instance disagreement is an unbiased estimate of the variance of error for that instance. We also show that disagreement provides a lower bound on the prediction (g en- eralization) error, and a tight upper bound on the "variance of pre diction error", or the variance of the average error across instances, where vari- ance is measured across training sets. We present experimen tal results on several data sets exploring co-validation for error esti mation and model selection. The procedure is especially effective in ac tive learning set- tings, where training sets are not drawn at random a nd cross validation overestimates error.

****

Co-Training and Expansion: Towards Bridging Theory and Practice

Maria-florina Balcan, Avrim Blum, Ke Yang

Ke Yang

****

Probabilistic Inference of Alternative Splicing Events in Microarray Data

Ofer Shai, Brendan J. Frey, Quaid Morris, Qun Pan, Christine Misquitta, Benjamin Blencowe

Alternative splicing (AS) is an important and frequent step in mammalian gene expression that allows a single gene to specify multiple products, and is crucial for the regulation of fundamental biological processes. The extent of AS regulation, and the mechanisms involved, are not well un- derstood. We have developed a custom DNA microarray platform for surveying AS levels on a large scale. We present here a generative model for the AS Array Platform (GenASAP) and demonstrate its utility for

quantifying AS levels in different mouse tissues. Learning is performed using a variational expectation maximization algorithm, and the parameters are shown to correctly capture expected AS trends. A comparison of the results obtained with a well-established but low through-put experimental method demonstrate that AS levels obtained from GenASAP are highly predictive of AS levels in mammalian tissues.

****

Semi-supervised Learning by Entropy Minimization

Yves Grandvalet, Yoshua Bengio

We consider the semi-supervised learning problem, where a decision rule is to be learned from labeled and unlabeled data. In this framework, we motivate minimum entropy regularization, which enables to incorporate unlabeled data in the standard supervised learning. Our approach includes other approaches to the semi-supervised problem as particular or limiting cases. A series of experiments illustrates that the proposed solution benefits from unlabeled data. The method challenges mixture models when the data are sampled from the distribution class spanned by the generative model. The performances are definitely in favor of minimum entropy regularization when generative models are misspecified, and the weighting of unlabeled data provides robustness to the violation of the "cluster assumption". Finally, we also illustrate that the method can also be far superior to manifold learning in high dimension spaces.

****

Detecting Significant Multidimensional Spatial Clusters

Daniel Neill, Andrew Moore, Francisco Pereira, Tom M. Mitchell

Assume a uniform, multidimensional grid of bivariate data, where each cell of the grid has a count ci and a baseline bi. Our goal is to find spatial regions (d-dimensional rectangles) where the ci are significantly higher than expected given bi. We focus on two applications: detection of clusters of disease cases from epidemiological data (emergency department visits, over-the-counter drug sales), and discovery of regions of increased brain activity corresponding to given cognitive tasks (from fMRI data). Each of these problems can be solved using a spatial scan statistic (Kulldorff, 1997), where we compute the maximum of a likelihood ratio statistic over all spatial regions, and find the significance of this region by randomization. However, computing the scan statistic for all spatial regions is generally computationally infeasible, so we introduce a novel fast spatial scan algorithm, generalizing the 2D scan algorithm of (Neill and Moore, 2004) to arbitrary dimensions. Our new multidimensional multiresolution algorithm allows us to find spatial clusters up to 1400x faster than the naive spatial scan, without any loss of accuracy.

****

Message Errors in Belief Propagation

Alexander Ihler, John Fisher, Alan Willsky

Belief propagation (BP) is an increasingly popular method of performing approximate inference on arbitrary graphical models. At times, even further approximations are required, whether from quantization or other simplified message representations or from stochastic approximation methods. Introducing such errors into the BP message computations has the potential to adversely affect the solution obtained. We analyze this effect with respect to a particular measure of message error, and show bounds on the accumulation of errors in the system. This leads both to convergence conditions and error bounds in traditional and approximate BP message passing.

****

Methods Towards Invasive Human Brain Computer Interfaces

Thomas Lal, Thilo Hinterberger, Guido Widman, Michael Schröder, N. Hill, Wolfgang Rosenstiel, Christian Elger, Niels Birbaumer, Bernhard Schölkopf

During the last ten years there has been growing interest in the development of Brain Computer Interfaces (BCIs). The field has mainly been driven by the needs of completely paralyzed patients to communicate. With a few exceptions, most human BCIs are based on extracranial electroencephalography (EEG). However, reported bit rates are still low. One reason for this is the low signal-to-noise ratio of the EEG [16]. We are currently investigating if BCIs based on electrocorticography (ECoG) are a viable alternative. In this paper we present the method and examples of intracranial EEG recordings of three epilepsy patients with electrode grids placed on the motor cortex. The patients were asked to repeatedly imagine movements of two kinds, e.g., tongue or finger movements. We analyze the classifiability of the data using Support Vector Machines (SVMs) [18, 21] and Recursive Channel Elimination (RCE) [11].
****

A Three Tiered Approach for Articulated Object Action Modeling and Recognition
Le Lu, Gregory Hager, Laurent Younes
Visual action recognition is an important problem in computer vision. In this paper, we propose a new method to probabilistically model and recognize actions of articulated objects, such as hand or body gestures, in image sequences. Our method consists of three levels of representation. At the low level, we first extract a feature vector invariant to scale and in-plane rotation by using the Fourier transform of a circular spatial histogram. Then, spectral partitioning [20] is utilized to obtain an initial clustering; this clustering is then refined using a temporal smoothness constraint. Gaussian mixture model (GMM) based clustering and density estimation in the subspace of linear discriminant analysis (LDA) are then applied to thousands of image feature vectors to obtain an intermediate level representation. Finally, at the high level we build a temporal multi-resolution histogram model for each action by aggregating the clustering weights of sampled images belonging to that action. We discuss how this high level representation can be extended to achieve temporal scaling invariance and to include Bi-gram or Multi-gram transition information. Both image clustering and action recognition/segmentation results are given to show the validity of our three tiered representation.
****

Temporal-Difference Networks
Richard S. Sutton, Brian Tanner
We introduce a generalization of temporal-difference (TD) learning to networks of interrelated predictions. Rather than relating a single prediction to itself at a later time, as in conventional TD methods, a TD network relates each prediction in a set of predictions to other predictions in the set at a later time. TD networks can represent and apply TD learning to a much wider class of predictions than has previously been possible. Using a random-walk example, we show that these networks can be used to learn to predict by a fixed interval, which is not possible with conventional TD methods. Secondly, we show that if the inter-predictive relationships are made conditional on action, then the usual learning-efficiency advantage of TD methods over Monte Carlo (supervised learning) methods becomes particularly pronounced. Thirdly, we demonstrate that TD networks can learn predictive state representations that enable exact solution of a non-Markov problem. A very broad range of inter-predictive temporal relationships can be expressed in these networks. Overall we argue that TD networks represent a substantial extension of the abilities of TD methods and bring us closer to the goal of representing world knowledge in entirely predictive, grounded terms.
****

Methods for Estimating the Computational Power and Generalization Capability of Neural Microcircuits
Wolfgang Maass, Robert Legenstein, Nils Bertschinger

What makes a neural microcircuit computationally powerful? Or more precisely, which measurable quantities could explain why one microcircuit C is better suited for a particular family of computational tasks than another microcircuit C ? We propose in this article quantitative measures for evaluating the computational power and generalization capability of a neural microcircuit, and apply them to generic neural microcircuit models drawn from different distributions. We validate the proposed measures by comparing their prediction with direct evaluations of the computational performance of these microcircuit models. This procedure is applied first to microcircuit models that differ with regard to the spatial range of synaptic connections and with regard to the scale of synaptic efficacies in the circuit, and then to microcircuit models that differ with regard to the level of background input currents and the level of noise on the membrane potential of neurons. In this case the proposed method allows us to quantify differences in the computational power and generalization capability of circuits in different dynamic regimes (UP- and DOWN-states) that have been demonstrated through intracellular recordings in vivo.

****

The Variational Ising Classifier (VIC) Algorithm for Coherently Contaminated Data

Oliver Williams, Andrew Blake, Roberto Cipolla

There has been substantial progress in the past decade in the development of object classifiers for images, for example of faces, humans and vehicles. Here we address the problem of contaminations (e.g. occlusion, shadows) in test images which have not explicitly been encountered in training data. The Variational Ising Classifier (VIC) algorithm models contamination as a mask (a field of binary variables) with a strong spatial coherence prior. Variational inference is used to marginalize over contamination and obtain robust classification. In this way the VIC approach can turn a kernel classifier for clean data into one that can tolerate contamination, without any specific training on contaminated positives.

****

Generalization Error and Algorithmic Convergence of Median Boosting

Balázs Kégl

We have recently proposed an extension of ADABOOST to regression that uses the median of the base regressors as the final regressor. In this paper we extend theoretical results obtained for ADABOOST to median boosting and to its localized variant. First, we extend recent results on efficient margin maximizing to show that the algorithm can converge to the maximum achievable margin within a preset precision in a finite number of steps. Then we provide confidence-interval-type bounds on the generalization error.

****

Supervised Graph Inference

Jean-philippe Vert, Yoshihiro Yamanishi

We formulate the problem of graph inference where part of the graph is known as a supervised learning problem, and propose an algorithm to solve it. The method involves the learning of a mapping of the vertices to a Euclidean space where the graph is easy to infer, and can be formulated as an optimization problem in a reproducing kernel Hilbert space. We report encouraging results on the problem of metabolic network reconstruction from genomic data.

****

Confidence Intervals for the Area Under the ROC Curve

Corinna Cortes, Mehryar Mohri

In many applications, good ranking is a highly desirable performance for a classifier. The criterion commonly used to measure the ranking quality of a classification algorithm is the area under the ROC curve (AUC). To

report it properly, it is crucial to determine an interval of confidence for its value. This paper provides confidence intervals for the AUC based on a statistical and combinatorial analysis using only simple parameters such as the error rate and the number of positive and negative examples. The analysis is distribution-independent, it makes no assumption about the distribution of the scores of negative or positive examples. The results are of practical use and can be viewed as the equivalent for AUC of the standard confidence intervals given in the case of the error rate. They are compared with previous approaches in several standard classification tasks demonstrating the benefits of our analysis.

****

Maximising Sensitivity in a Spiking Network

Anthony Bell, Lucas Parra

We use unsupervised probabilistic machine learning ideas to try to ex- plain the kinds of learning observed in real neurons, the goal being to connect abstract principles of self-organisation to known biophysi- cal processes. For example, we would like to explain Spike Timing- Dependent Plasticity (see [5,6] and Figure 3A), in terms of information theory. Starting out, we explore the optimisation of a network sensitiv- ity measure related to maximising the mutual information between input spike timings and output spike timings. Our derivations are analogous to those in ICA, except that the sensitivity of output timings to input tim- ings is maximised, rather than the sensitivity of output '■ring rates' to inputs. ICA and related approaches have been successful in explaining the learning of many properties of early visual receptive ■elds in rate cod- ing models, and we are hoping for similar gains in understanding of spike coding in networks, and how this is supported, in principled probabilistic ways, by cellular biophysical processes. For now, in our initial simula- tions, we show that our derived rule can learn synaptic weights which can unmix, or demultiplex, mixed spike trains. That is, it can recover inde- pendent point processes embedded in distributed correlated input spike trains, using an adaptive single-layer feedforward spiking network.

****

Probabilistic Computation in Spiking Populations

Richard Zemel, Rama Natarajan, Peter Dayan, Quentin Huys

As animals interact with their environments, they must constantly update estimates about their states. Bayesian models combine prior probabilities, a dynamical model and sensory evidence to update estimates optimally. These models are consistent with the results of many diverse psychophysical studies. However, little is known about the neural representation and manipulation of such Bayesian information, particularly in populations of spiking neurons. We consider this issue, suggesting a model based on standard neural architecture and activations. We illustrate the approach on a simple random walk example, and apply it to a sensorimotor integration task that provides a particularly compelling example of dynamic probabilistic computation.

****

Theory of localized synfire chain: characteristic propagation speed of stable spike pattern

Kosuke Hamaguchi, Masato Okada, Kazuyuki Aihara

Repeated spike patterns have often been taken as evidence for the synfire chain, a phenomenon that a stable spike synchrony propagates through a feedforward network. Inter-spike intervals which represent a repeated spike pattern are influenced by the propagation speed of a spike packet. However, the relation between the propagation speed and network structure is not well understood. While it is apparent that the propagation speed depends on the excitatory synapse strength, it might also be related to spike patterns. We analyze a feedforward network with Mexican-Hat-type connectivity (FMH) using the Fokker-Planck equation. We show that both a uniform and a localized spike packet are stable in the FMH in a certain parameter region. We also demonstrate that the propagation

speed depends on the distinct firing patterns in the same network.

****

Semi-supervised Learning on Directed Graphs

Dengyong Zhou, Thomas Hofmann, Bernhard Schölkopf

Given a directed graph in which some of the nodes are labeled, we inves- tigate the question of how to exploit the link structure of the graph to infer the labels of the remaining unlabeled nodes. To that extent we propose a regularization framework for functions de(cid:2)ned over nodes of a directed graph that forces the classi(cid:2)cation function to change slowly on densely linked subgraphs. A powerful, yet computationally simple classi(cid:2)cation algorithm is derived within the proposed framework. The experimental evaluation on real-world Web classi(cid:2)cation problems demonstrates en- couraging results that validate our approach.

****

Class-size Independent Generalization Analsysis of Some Discriminative Multi-Category Classification

Tong Zhang

We consider the problem of deriving class-size independent generalization bounds for some regularized discriminative multi-category classification methods. In particular, we obtain an expected generalization bound for a standard formulation of multi-category support vector machines. Based on the theoretical result, we argue that the formulation over-penalizes misclassification error, which in theory may lead to poor generalization performance. A remedy, based on a generalization of multi-category logistic regression (conditional maximum entropy), is then proposed, and its theoretical properties are examined.

****

■■-norm Minimization for Basis Selection

David Wipf, Bhaskar Rao

Finding the sparsest, or minimum ■0-norm, representation of a signal given an overcomplete dictionary of basis vectors is an important prob- lem in many application domains. Unfortunately, the required optimiza- tion problem is often intractable because there is a combinatorial increase in the number of local minima as the number of candidate basis vectors increases. This de■ciency has prompted most researchers to instead min- imize surrogate measures, such as the ■1-norm, that lead to more tractable computational methods. The downside of this procedure is that we have now introduced a mismatch between our ultimate goal and our objective function. In this paper, we demonstrate a sparse Bayesian learning-based method of minimizing the ■0-norm while reducing the number of trou- blesome local minima. Moreover, we derive necessary conditions for local minima to occur via this approach and empirically demonstrate that there are typically many fewer for general problems of interest.

****

The Power of Selective Memory: Self-Bounded Learning of Prediction Suffix Trees

Ofer Dekel, Shai Shalev-shwartz, Yoram Singer

Prediction suffix trees (PST) provide a popular and effective tool for tasks such as compression, classification, and language modeling. In this paper we take a decision theoretic view of PSTs for the task of sequence prediction. Generalizing the notion of margin to PSTs, we present an online PST learning algorithm and derive a loss bound for it. The depth of the PST generated by this algorithm scales linearly with the length of the input. We then describe a self-bounded enhancement of our learning algorithm which automatically grows a bounded-depth PST. We also prove an analogous mistake-bound for the self-bounded algorithm. The result is an efficient algorithm that neither relies on a-priori assumptions on the shape or maximal depth of the target PST nor does it require any parameters. To our knowledge, this is the first provably-correct PST learning algorithm which generates a bounded-depth PST while being competitive with any fixed PST determined in hindsight.

****

Modelling Uncertainty in the Game of Go
David Stern, Thore Graepel, David MacKay
Go is an ancient oriental game whose complexity has defeated at- tempts to automate it. We suggest using probability in a Bayesian sense to model the uncertainty arising from the vast complexity of the game tree. We present a simple conditional Markov random field model for predicting the pointwise territory outcome of a game. The topology of the model reflects the spatial structure of the Go board. We describe a version of the Swendsen-Wang process for sampling from the model during learning and apply loopy belief propagation for rapid inference and prediction. The model is trained on several hundred records of professional games. Our experimental results indicate that the model successfully learns to predict territory despite its simplicity.
****
Spike Sorting: Bayesian Clustering of Non-Stationary Data
Aharon Bar-hillel, Adam Spiro, Eran Stark
Spike sorting involves clustering spike trains recorded by a microelectrode according to the source neuron. It is a complicated problem, which requires a lot of human labor, partly due to the non-stationary nature of the data. We propose an automated technique for the clustering of non-stationary Gaussian sources in a Bayesian framework. At a first search stage, data is divided into short time frames and candidate descriptions of the data as a mixture of Gaussians are computed for each frame. At a second stage transition probabilities between candidate mixtures are computed, and a globally optimal clustering is found as the MAP solution of the resulting probabilistic model. Transition probabilities are computed using local stationarity assumptions and are based on a Gaussian version of the Jensen-Shannon divergence. The method was applied to several recordings. The performance appeared almost indistinguishable from humans in a wide range of scenarios, including movement, merges, and splits of clusters.
****
Harmonising Chorales by Probabilistic Inference
Moray Allan, Christopher Williams
We describe how we used a data set of chorale harmonisations composed by Johann Sebastian Bach to train Hidden Markov Models. Using a probabilistic framework allows us to create a harmonisation system which learns from examples, and which can compose new harmonisations. We make a quantitative comparison of our system's harmonisation performance against simpler models, and provide example harmonisations.
****
Nearly Tight Bounds for the Continuum-Armed Bandit Problem
Robert Kleinberg
In the multi-armed bandit problem, an online algorithm must choose from a set of strategies in a sequence of n trials so as to minimize the total cost of the chosen strategies. While nearly tight upper and lower bounds are known in the case when the strategy set is finite, much less is known when there is an infinite strategy set. Here we consider the case when the set of strategies is a subset of Rd, and the cost functions are continuous. In the d = 1 case, we improve on the best-known upper and lower bounds, closing the gap to a sublogarithmic factor. We also consider the case where d > 1 and the cost functions are convex, adapting a recent online convex optimization algorithm of Zinkevich to the sparser feedback model of the multi-armed bandit problem.
****
Mistake Bounds for Maximum Entropy Discrimination
Philip Long, Xinyu Wu
We establish a mistake bound for an ensemble method for classification based on maximizing the entropy of voting weights subject to margin constraints. The bound is the same as a general bound proved for the Weigh

ted Majority Algorithm, and similar to bounds for other variants of Win
now. We prove a more refined bound that leads to a nearly opti- mal alg
orithm for learning disjunctions, again, based on the maximum entropy p
rinciple. We describe a simplification of the on-line maximum entropy m
ethod in which, after each iteration, the margin constraints are replac
ed with a single linear inequality. The simplified algorithm, which tak
es a similar form to Winnow, achieves the same mistake bounds.
****

A Harmonic Excitation State-Space Approach to Blind Separation of Speech
Rasmus Olsson, Lars K. Hansen
We discuss an identi■cation framework for noisy speech mixtures. A
block-based generative model is formulated that explicitly incorporates
the time-varying harmonic plus noise (H+N) model for a number of latent
sources observed through noisy convolutive mixtures. All parameters
including the pitches of the source signals, the amplitudes and phases of
the sources, the mixing ■lters and the noise statistics are estimated by
maximum likelihood, using an EM-algorithm. Exact averaging over the
hidden sources is obtained using the Kalman smoother. We show that
pitch estimation and source separation can be performed simultaneously.
The pitch estimates are compared to laryngograph (EGG) measurements.
Arti■cial and real room mixtures are used to demonstrate the viability
of the approach. Intelligible speech signals are re-synthesized from the
estimated H+N models.
****

Matrix Exponential Gradient Updates for On-line Learning and Bregman Projection
Koji Tsuda, Gunnar Rätsch, Manfred K. K. Warmuth
We address the problem of learning a symmetric positive definite matrix.
The central issue is to design parameter updates that preserve positive
definiteness. Our updates are motivated with the von Neumann diver-
gence. Rather than treating the most general case, we focus on two key
applications that exemplify our methods: On-line learning with a simple
square loss and finding a symmetric positive definite matrix subject to
symmetric linear constraints. The updates generalize the Exponentiated
Gradient (EG) update and AdaBoost, respectively: the parameter is now
a symmetric positive definite matrix of trace one instead of a probabil
ity vector (which in this context is a diagonal positive definite matr
ix with trace one). The generalized updates use matrix logarithms and
exponen- tials to preserve positive definiteness. Most importantly, we
show how the analysis of each algorithm generalizes to the non-diagon
al case. We apply both new algorithms, called the Matrix Exponentiated
Gradient (MEG) update and DefiniteBoost, to learn a kernel matrix fro
m distance measurements.
****

Blind One-microphone Speech Separation: A Spectral Learning Approach
Francis Bach, Michael Jordan
We present an algorithm to perform blind, one-microphone speech sep- ar
ation. Our algorithm separates mixtures of speech without modeling indi
vidual speakers. Instead, we formulate the problem of speech sep- arati
on as a problem in segmenting the spectrogram of the signal into two or
more disjoint sets. We build feature sets for our segmenter using clas
sical cues from speech psychophysics. We then combine these fea- tures
into parameterized affinity matrices. We also take advantage of the fac
t that we can generate training examples for segmentation by artifi- ci
ally superposing separately-recorded signals. Thus the parameters of th
e affinity matrices can be tuned using recent work on learning spectral
clustering [1]. This yields an adaptive, speech-specific segmentation al-
gorithm that can successfully separate one-microphone speech mixtures.
****

Resolving Perceptual Aliasing In The Presence Of Noisy Sensors
Guy Shani, Ronen Brafman

Agents learning to act in a partially observable domain may need to ove
rcome the problem of perceptual aliasing i.e., different states that a
ppear similar but require different responses. This problem is exacer-
bated when the agent's sensors are noisy, i.e., sensors may produce dif-
ferent observations in the same state. We show that many well-known r
einforcement learning methods designed to deal with perceptual alias- i
ng, such as Utile Suffix Memory, finite size history windows, eligibility
traces, and memory bits, do not handle noisy sensors well. We suggest
a new algorithm, Noisy Utile Suffix Memory (NUSM), based on USM, that
uses a weighted classification of observed trajectories. We compare NU
SM to the above methods and show it to be more robust to noise.
****

Kernels for Multi--task Learning
Charles Micchelli, Massimiliano Pontil
This paper provides a foundation for multitask learning using reproducing ker-
nel Hilbert spaces of vectorvalued functions. In this setting, the kerne
l is a matrixvalued function. Some explicit examples will be described
which go be- yond our earlier results in [7]. In particular, we charact
erize classes of matrix valued kernels which are linear and are of the
dot product or the translation invari- ant type. We discuss how these k
ernels can be used to model relations between the tasks and present lin
ear multitask learning algorithms. Finally, we present a novel proof of
the representer theorem for a minimizer of a regularization func- tion
al which is based on the notion of minimal norm interpolation.
****

Variational Minimax Estimation of Discrete Distributions under KL Loss
Liam Paninski
We develop a family of upper and lower bounds on the worst-case ex- pec
ted KL loss for estimating a discrete distribution on a finite number m
of points, given N i.i.d. samples. Our upper bounds are approximation-
theoretic, similar to recent bounds for estimating discrete entropy; the
lower bounds are Bayesian, based on averages of the KL loss under Di
richlet distributions. The upper bounds are convex in their parameters
and thus can be minimized by descent methods to provide estimators with
low worst-case error; the lower bounds are indexed by a one-dimensional
parameter and are thus easily maximized. Asymptotic analysis of the b
ounds demonstrates the uniform KL-consistency of a wide class of es- ti
mators as c = N/m (no matter how slowly), and shows that no estimator
is consistent for c bounded (in contrast to entropy estima- tion). Mor
eover, the bounds are asymptotically tight as c 0 or , and are shown n
umerically to be tight within a factor of two for all c. Finally, in th
e sparse-data limit c 0, we find that the Dirichlet-Bayes (add-constan
t) estimator with parameter scaling like -c log(c) optimizes both the u
pper and lower bounds, suggesting an optimal choice of the "add-constan
t" parameter in this regime.
****

Online Bounds for Bayesian Algorithms
Sham M. Kakade, Andrew Ng
We present a competitive analysis of Bayesian learning algorithms in the online
learning setting and show that many simple Bayesian algorithms (such as Gaussian
linear regression and Bayesian logistic regression) per- form favorably when co
mpared, in retrospect, to the single best model in the model class. The analysis
does not assume that the Bayesian algo- rithms' modeling assumptions are "corre
ct," and our bounds hold even if the data is adversarially chosen. For Gaussian
linear regression (us- ing logloss), our error bounds are comparable to the best
bounds in the online learning literature, and we also provide a lower bound sho
wing that Gaussian linear regression is optimal in a certain worst case sense. W
e also give bounds for some widely used maximum a posteriori (MAP) estimation al
gorithms, including regularized logistic regression.
****

Analysis of a greedy active learning strategy
Sanjoy Dasgupta
We abstract out the core search problem of active learning schemes, to better understand the extent to which adaptive labeling can improve sample complexity. We give various upper and lower bounds on the number of labels which need to be queried, and we prove that a popular greedy active learning rule is approximately as good as any other strategy for minimizing this number of labels.
****
A Cost-Shaping LP for Bellman Error Minimization with Performance Guarantees
Daniela Farias, Benjamin Roy
We introduce a new algorithm based on linear programming that approximates the differential value function of an average-cost Markov decision process via a linear combination of pre-selected basis functions. The algorithm carries out a form of cost shaping and minimizes a version of Bellman error. We establish an error bound that scales gracefully with the number of states without imposing the (strong) Lyapunov condition required by its counterpart in [6]. We propose a path-following method that automates selection of important algorithm parameters which represent counterparts to the "state-relevance weights" studied in [6].
****
Validity Estimates for Loopy Belief Propagation on Binary Real-world Networks
Joris M. Mooij, Hilbert Kappen
We introduce a computationally efficient method to estimate the validity of the BP method as a function of graph topology, the connectivity strength, frustration and network size. We present numerical results that demonstrate the correctness of our estimates for the uniform random model and for a real-world network ("C. Elegans"). Although the method is restricted to pair-wise interactions, no local evidence (zero "biases") and binary variables, we believe that its predictions correctly capture the limitations of BP for inference and MAP estimation on arbitrary graphical models. Using this approach, we find that BP always performs better than MF. Especially for large networks with broad degree distributions (such as scale-free networks) BP turns out to significantly outperform MF.
****
Bayesian inference in spiking neurons
Sophie Deneve
We propose a new interpretation of spiking neurons as Bayesian integrators accumulating evidence over time about events in the external world or the body, and communicating to other neurons their certainties about these events. In this model, spikes signal the occurrence of new information, i.e. what cannot be predicted from the past activity. As a result, firing statistics are close to Poisson, albeit providing a deterministic representation of probabilities. We proceed to develop a theory of Bayesian inference in spiking neural networks, recurrent interactions implementing a variant of belief propagation.
****
Computing regularization paths for learning multiple kernels
Francis Bach, Romain Thibaux, Michael Jordan
The problem of learning a sparse conic combination of kernel functions or kernel matrices for classification or regression can be achieved via the regularization by a block 1-norm [1]. In this paper, we present an algorithm that computes the entire regularization path for these problems. The path is obtained by using numerical continuation techniques, and involves a running time complexity that is a constant times the complexity of solving the problem for one value of the regularization parameter. Working in the setting of kernel linear regression and kernel logistic regression, we show empirically that the effect of the block 1-norm

reg- ularization differs notably from the (non-block) 1-norm regulariza tion commonly used for variable selection, and that the regularization path is of particular value in the block case.

****

Saliency-Driven Image Acuity Modulation on a Reconfigurable Array of Spiking Sil icon Neurons

R. Vogelstein, Udayan Mallik, Eugenio Culurciello, Gert Cauwenberghs, Ralph Etie nne-Cummings

We have constructed a system that uses an array of 9,600 spiking sili- con neuro ns, a fast microcontroller, and digital memory, to implement a recon■gurable net work of integrate-and-■re neurons. The system is designed for rapid prototyping of spiking neural networks that require high-throughput communication with exter nal address-event hardware. Arbitrary network topologies can be implemented by s electively rout- ing address-events to speci■c internal or external targets acco rding to a memory-based projective ■eld mapping. The utility and versatility of the system is demonstrated by con■guring it as a three-stage network that accept s input from an address-event imager, detects salient regions of the image, and performs spatial acuity modulation around a high-resolution fovea that is center ed on the location of highest salience.

****

Common-Frame Model for Object Recognition

Pierre Moreels, Pietro Perona

A generative probabilistic model for objects in images is presented. An object consists of a constellation of features. Feature appearance and pose are modeled probabilistically. Scene images are generated by draw- ing a set of objects from a given database, with random clutter sprinkled on the remaining image surface. Occlusion is allowed. We study the case where features from the same object share a common reference fram e. Moreover, parameters for shape and appearance den- sities are shared across features. This is to be contrasted with previous work on probab ilistic `constellation' models where features depend on each other, and each feature and model have different pose and appear- ance statistics [1, 2]. These two differences allow us to build models containing hund reds of features, as well as to train each model from a single example. Our model may also be thought of as a probabilistic revisitation of Lo we's model [3, 4]. We propose an efficient entropy-minimization inferen ce algorithm that constructs the best interpretation of a scene as a co llection of objects and clutter. We test our ideas with experiments on two image databases. We compare with Lowe's algorithm and demonstrate b etter performance, in particular in presence of large amounts of backgr ound clutter.

****

Brain Inspired Reinforcement Learning

Françcois Rivest, Yoshua Bengio, John Kalaska

Successful application of reinforcement learning algorithms often involves cons iderable hand-crafting of the necessary non-linear features to reduce the compl exity of the value functions and hence to promote convergence of the algorithm. In contrast, the human brain readily and autonomously finds the complex featur es when provided with sufficient training. Recent work in machine learning and neurophysiology has demonstrated the role of the basal ganglia and the frontal cortex in mammalian reinforcement learning. This paper develops and explores n ew learning algorithms that provides potential new approaches to the feature construction problem. The algorithms are compared and evaluated on the Acrobot task.

****

Semi-supervised Learning via Gaussian Processes

Neil Lawrence, Michael Jordan

We present a probabilistic approach to learning a Gaussian Process class ifier in the presence of unlabeled data. Our approach involves a "null c ategory noise model" (NCNM) inspired by ordered cate- gorical noise mode

ls. The noise model reflects an assumption that the data density is lowe
r between the class-conditional densities. We illustrate our approach on
a toy problem and present compar- ative results for the semi-supervised
classification of handwritten digits.
****

Parallel Support Vector Machines: The Cascade SVM
Hans Graf, Eric Cosatto, Léon Bottou, Igor Dourdanovic, Vladimir Vapnik
We describe an algorithm for support vector machines (SVM) that can be parallel
ized efficiently and scales to very large problems with hundreds of thousands o
f training vectors. Instead of analyzing the whole training set in one optimiza
tion step, the data are split into subsets and optimized separately with multip
le SVMs. The partial results are combined and filtered again in a 'Cascade' of
SVMs, until the global optimum is reached. The Cascade SVM can be spread over
multiple processors with minimal communication overhead and requires far less m
emory, since the kernel matrices are much smaller than for a regular SVM. Conve
rgence to the global optimum is guaranteed with multiple passes through the Cas
cade, but already a single pass provides good generalization. A single pass is
5x – 10x faster than a regular SVM for problems of 100,000 vectors when implem
ented on a single processor. Parallel implementations on a cluster of 16 proces
sors were tested with over 1 million vectors (2-class problems), converging in
a day or two, while a regular SVM never converged in over a week.
****

Sampling Methods for Unsupervised Learning
Rob Fergus, Andrew Zisserman, Pietro Perona
We present an algorithm to overcome the local maxima problem in es-
timating the parameters of mixture models. It combines existing ap-
proaches from both EM and a robust fitting algorithm, RANSAC, to give
a data-driven stochastic learning scheme. Minimal subsets of data poi
nts,
sufficient to constrain the parameters of the model, are drawn from p
ro-
posal densities to discover new regions of high likelihood. The propo
sal
densities are learnt using EM and bias the sampling toward promising
solutions. The algorithm is computationally efficient, as well as eff
ective
at escaping from local maxima. We compare it with alternative methods
,
including EM and RANSAC, on both challenging synthetic data and the
computer vision problem of alpha-matting.
****

Limits of Spectral Clustering
Ulrike Luxburg, Olivier Bousquet, Mikhail Belkin
An important aspect of clustering algorithms is whether the partitions
constructed on finite samples converge to a useful clustering of the whole
data space as the sample size increases. This paper investigates this
question for normalized and unnormalized versions of the popular spec-
tral clustering algorithm. Surprisingly, the convergence of unnormalized
spectral clustering is more difficult to handle than the normalized case.
Even though recently some first results on the convergence of normal-
ized spectral clustering have been obtained, for the unnormalized case
we have to develop a completely new approach combining tools from
numerical integration, spectral and perturbation theory, and probability.
It turns out that while in the normalized case, spectral clustering usually
converges to a nice partition of the data space, in the unnormalized c
ase the same only holds under strong additional assumptions which are n
ot always satisfied. We conclude that our analysis gives strong evidenc
e for the superiority of normalized spectral clustering. It also provid
es a basis for future exploration of other Laplacian-based methods.
****

Using the Equivalent Kernel to Understand Gaussian Process Regression
Peter Sollich, Christopher Williams
The equivalent kernel [1] is a way of understanding how Gaussian pro-
cess regression works for large sample sizes based on a continuum limit.
In this paper we show (1) how to approximate the equivalent kernel of the
widely-used squared exponential (or Gaussian) kernel and related ke
r- nels, and (2) how analysis using the equivalent kernel helps to u
nderstand the learning curves for Gaussian processes.
****

Newscast EM
Wojtek Kowalczyk, Nikos Vlassis
We propose a gossip-based distributed algorithm for Gaussian mixture learning, N
ewscast EM. The algorithm operates on network topologies where each node observe
s a local quantity and can communicate with other nodes in an arbitrary point-to
-point fashion. The main difference between Newscast EM and the standard EM algo
rithm is that the M-step in our case is implemented in a decentralized manner: (
random) pairs of nodes repeatedly exchange their local parameter estimates and c
om- bine them by (weighted) averaging. We provide theoretical evidence and demon
strate experimentally that, under this protocol, nodes converge ex- ponentially
fast to the correct estimates in each M-step of the EM algo- rithm.
****

Dynamic Bayesian Networks for Brain-Computer Interfaces
Pradeep Shenoy, Rajesh PN Rao
We describe an approach to building brain-computer interfaces (BCI) bas
ed on graphical models for probabilistic inference and learning. We sho
w how a dynamic Bayesian network (DBN) can be used to infer probability
distributions over brain- and body-states during planning and executio
n of actions. The DBN is learned directly from observed data and allows
measured signals such as EEG and EMG to be interpreted in terms of int
ernal states such as intent to move, preparatory activity, and movement
execution. Unlike traditional classification-based approaches to BCI,
the proposed approach (1) allows continuous tracking and predic- tion o
f internal states over time, and (2) generates control signals based on
an entire probability distribution over states rather than binary yes/no
decisions. We present preliminary results of brain- and body-state es-
timation using simultaneous EEG and EMG signals recorded during a se
lf-paced left/right hand movement task.
****

Edge of Chaos Computation in Mixed-Mode VLSI - A Hard Liquid
Felix Schürmann, Karlheinz Meier, Johannes Schemmel
Computation without stable states is a computing paradigm dif- ferent fr
om Turing's and has been demonstrated for various types of simulated neu
ral networks. This publication transfers this to a hardware implemented
neural network. Results of a software im- plementation are reproduced sh
owing that the performance peaks when the network exhibits dynamics at t
he edge of chaos. The liquid computing approach seems well suited for op
erating analog computing devices such as the used VLSI neural network.
****

Breaking SVM Complexity with Cross-Training
Léon Bottou, Jason Weston, Gökhan Bakir
We propose to selectively remove examples from the training set using probabilis
tic estimates related to editing algorithms (Devijver and Kittler, 1982). This h
euristic procedure aims at creating a separable distribution of training example
s with minimal impact on the position of the decision boundary. It breaks the li
near dependency between the number of SVs and the number of training examples, a
nd sharply reduces the complexity of SVMs during both the training and predictio
n stages.
****

Linear Multilayer Independent Component Analysis for Large Natural Scenes
Yoshitatsu Matsuda, Kazunori Yamaguchi

In this paper, linear multilayer ICA (LMICA) is proposed for extracting independent components from quite high-dimensional observed signals such as large-size natural scenes. There are two phases in each layer of LMICA. One is the mapping phase, where a one-dimensional mapping is formed by a stochastic gradient algorithm which makes more highly-correlated (non-independent) signals be nearer incrementally. Another is the local-ICA phase, where each neighbor (namely, highly-correlated) pair of signals in the mapping is separated by the MaxKurt algorithm. Because LMICA separates only the highly-correlated pairs instead of all ones, it can extract independent components quite efficiently from ap- propriate observed signals. In addition, it is proved that LMICA al ways converges. Some numerical experiments verify that LMICA is quite ef- ficient and effective in large-size natural image processing.

****

Identifying Protein-Protein Interaction Sites on a Genome-Wide Scale
Haidong Wang, Eran Segal, Asa Ben-Hur, Daphne Koller, Douglas Brutlag

Protein interactions typically arise from a physical interaction of one or more small sites on the surface of the two proteins. Identifying these sit es is very important for drug and protein design. In this paper, we p ropose a computational method based on probabilistic relational model that at- tempts to address this task using high-throughput protein i nteraction data and a set of short sequence motifs. We learn the mode l using the EM algorithm, with a branch-and-bound algorithm as an app roximate infer- ence for the E-step. Our method searches for motifs w hose presence in a pair of interacting proteins can explain their obs erved interaction. It also tries to determine which motif pairs have high affinity, and can therefore lead to an interaction. We show that our method is more accurate than others at predicting new protein-pr otein interactions. More importantly, by examining solved structures of protein complexes, we find that 2/3 of the predicted active motifs correspond to actual interaction sites.

****

Proximity Graphs for Clustering and Manifold Learning
Richard Zemel, Miguel Carreira-Perpiñán

Many machine learning algorithms for clustering or dimensionality re- duction ta ke as input a cloud of points in Euclidean space, and construct a graph with the input data points as vertices. This graph is then parti- tioned (clustering) or used to rede■ne metric information (dimensional- ity reduction). There has been much recent work on new methods for graph-based clustering and dimensionality r eduction, but not much on constructing the graph itself. Graphs typically used i nclude the fully- connected graph, a local ■xed-grid graph (for image segmentati on) or a nearest-neighbor graph. We suggest that the graph should adapt locally to the structure of the data. This can be achieved by a graph ensemble that comb ines multiple minimum spanning trees, each ■t to a perturbed version of the data set. We show that such a graph ensemble usually pro- duces a better representat ion of the data manifold than standard methods; and that it provides robustness to a subsequent clustering or dimension- ality reduction algorithm based on the graph.

****

Fast Rates to Bayes for Kernel Machines
Ingo Steinwart, Clint Scovel

We establish learning rates to the Bayes risk for support vector machines (SVMs) with hinge loss. In particular, for SVMs with Gaussian RBF ke rnels we propose a geometric condition for distributions which can be u sed to determine approximation properties of these kernels. Finally, we compare our methods with a recent paper of G. Blanchard et al..

****

Discriminant Saliency for Visual Recognition from Cluttered Scenes
Dashan Gao, Nuno Vasconcelos

Saliency mechanisms play an important role when visual recognition must

be performed in cluttered scenes. We propose a computational definition of saliency that deviates from existing models by equating saliency to discrimination. In particular, the salient attributes of a given visual class are defined as the features that enable best discrimination between that class and all other classes of recognition interest. It is shown that this definition leads to saliency algorithms of low complexity, that are scalable to large recognition problems, and is compatible with existing models of early biological vision. Experimental results demonstrating success in the context of challenging recognition problems are also presented.

****

Bayesian Regularization and Nonnegative Deconvolution for Time Delay Estimation

Yuanqing Lin, Daniel Lee

Bayesian Regularization and Nonnegative Deconvolution (BRAND) is proposed for estimating time delays of acoustic signals in reverberant environments. Sparsity of the nonnegative filter coefficients is enforced using an L1-norm regularization. A probabilistic generative model is used to simultaneously estimate the regularization parameters and filter coefficients from the signal data. Iterative update rules are derived under a Bayesian framework using the Expectation-Maximization procedure. The resulting time delay estimation algorithm is demonstrated on noisy acoustic data.

****

Algebraic Set Kernels with Application to Inference Over Local Image Representations

Amnon Shashua, Tamir Hazan

This paper presents a general family of algebraic positive definite similarity functions over spaces of matrices with varying column rank. The columns can represent local regions in an image (whereby images have varying number of local parts), images of an image sequence, motion trajectories in a multibody motion, and so forth. The family of set kernels we derive is based on a group invariant tensor product lifting with parameters that can be naturally tuned to provide a cook-book of sorts covering the possible "wish lists" from similarity measures over sets of varying cardinality. We highlight the strengths of our approach by demonstrating the set kernels for visual recognition of pedestrians using local parts representations.

****

The Correlated Correspondence Algorithm for Unsupervised Registration of Nonrigid Surfaces

Dragomir Anguelov, Praveen Srinivasan, Hoi-cheung Pang, Daphne Koller, Sebastian Thrun, James Davis

We present an unsupervised algorithm for registering 3D surface scans of an object undergoing significant deformations. Our algorithm does not need markers, nor does it assume prior knowledge about object shape, the dynamics of its deformation, or scan alignment. The algorithm registers two meshes by optimizing a joint probabilistic model over all point-to-point correspondences between them. This model enforces preservation of local mesh geometry, as well as more global constraints that capture the preservation of geodesic distance between corresponding point pairs. The algorithm applies even when one of the meshes is an incomplete range scan; thus, it can be used to automatically fill in the remaining surfaces for this partial scan, even if those surfaces were previously only seen in a different configuration. We evaluate the algorithm on several real-world datasets, where we demonstrate good results in the presence of significant movement of articulated parts and non-rigid surface deformation. Finally, we show that the output of the algorithm can be used for compelling computer graphics tasks such as interpolation between two scans of a non-rigid object and automatic recovery of articulated object models.

****
Maximum-Margin Matrix Factorization
Nathan Srebro, Jason Rennie, Tommi Jaakkola
We present a novel approach to collaborative prediction, using low-norm instead of low-rank factorizations. The approach is inspired by, and has strong connections to, large-margin linear discrimination. We show how to learn low-norm factorizations by solving a semi-de■nite program, and discuss generalization error bounds for them.
****
Real-Time Pitch Determination of One or More Voices by Nonnegative Matrix Factorization
Fei Sha, Lawrence Saul
An auditory "scene", composed of overlapping acoustic sources, can be viewed as a complex object whose constituent parts are the individual sources. Pitch is known to be an important cue for auditory scene analysis. In this paper, with the goal of building agents that operate in human environments, we describe a real-time system to identify the presence of one or more voices and compute their pitch. The signal processing in the front end is based on instantaneous frequency estimation, a method for tracking the partials of voiced speech, while the pattern-matching in the back end is based on nonnegative matrix factorization, an unsupervised algorithm for learning the parts of complex objects. While supporting a framework to analyze complicated auditory scenes, our system maintains real-time operability and state-of-the-art performance in clean speech.
****
Efficient Out-of-Sample Extension of Dominant-Set Clusters
Massimiliano Pavan, Marcello Pelillo
Dominant sets are a new graph-theoretic concept that has proven to be relevant in pairwise data clustering problems, such as image segmentation. They generalize the notion of a maximal clique to edge-weighted graphs and have intriguing, non-trivial connections to continuous quadratic optimization and spectral-based grouping. We address the problem of grouping out-of-sample examples after the clustering process has taken place. This may serve either to drastically reduce the computational burden associated to the processing of very large data sets, or to efficiently deal with dynamic situations whereby data sets need to be updated continually. We show that the very notion of a dominant set offers a simple and efficient way of doing this. Numerical experiments on various grouping problems show the effectiveness of the approach.
****
Instance-Specific Bayesian Model Averaging for Classification
Shyam Visweswaran, Gregory Cooper
Classification algorithms typically induce population-wide models that are trained to perform well on average on expected future instances. We introduce a Bayesian framework for learning instance-specific models from data that are optimized to predict well for a particular instance. Based on this framework, we present a that performs selective model averaging over a restricted class of Bayesian networks. On experimental evaluation, this algorithm shows superior performance over model selection. We intend to apply such instance-specific algorithms to improve the performance of patient-specific predictive models induced from medical data.
****
Hierarchical Eigensolver for Transition Matrices in Spectral Methods
Chakra Chennubhotla, Allan Jepson
We show how to build hierarchical, reduced-rank representation for large stochastic matrices and use this representation to design an ef■cient al- gorithm for computing the largest eigenvalues, and the corresponding eigenvectors. In particular, the eigen problem is ■rst solved at the coars- est level of the representation. The approximate eigen solution is then interpolated over successive levels

of the hierarchy. A small number of power iterations are employed at each stage to correct the eigen solution. The typical speedups obtained by a Matlab impleme ntation of our fast eigensolver over a standard sparse matrix eigensolver [13] a re at least a factor of ten for large image sizes. The hierarchical representati on has proven to be effective in a min-cut based segmentation algorithm that we proposed recently [8].

****

Exponentiated Gradient Algorithms for Large-margin Structured Classification
Peter Bartlett, Michael Collins, Ben Taskar, David McAllester

We consider the problem of structured classi■cation, where the task is to predic t a label y from an input x, and y has meaningful internal struc- ture. Our fram ework includes supervised training of Markov random ■elds and weighted context-f ree grammars as special cases. We describe an algorithm that solves the large-ma rgin optimization problem de■ned in [12], using an exponential-family (Gibbs dis tribution) representation of structured objects. The algorithm is ef■cient—even in cases where the number of labels y is exponential in size—provided that certa in expecta- tions under Gibbs distributions can be calculated ef■ciently. The me thod for structured labels relies on a more general result, speci■cally the ap- plication of exponentiated gradient updates [7, 8] to quadratic programs.

****

A Method for Inferring Label Sampling Mechanisms in Semi-Supervised Learning
Saharon Rosset, Ji Zhu, Hui Zou, Trevor Hastie

We consider the situation in semi-supervised learning, where the "label sampling" mechanism stochastically depends on the true response (as well as potentially on the features). We suggest a method of moments for estimating this stochastic dependence using the unlabeled data. This is potentially useful for two distinct purposes: a. As an input to a super- vised learning procedure which can be used to "de-bias" its results using labeled data only and b. As a potentially interesting learning task in it- self. We present several examples to illustrate the prac tical usefulness of our method.

****

Semi-Markov Conditional Random Fields for Information Extraction
Sunita Sarawagi, William W. Cohen

We describe semi-Markov conditional random ■elds (semi-CRFs), a con- ditionally trained version of semi-Markov chains. Intuitively, a semi- CRF on an input sequ ence x outputs a “segmentation” of x, in which labels are assigned to segments (i.e., subsequences) of x rather than to individual elements xi of x. Importantly , features for semi-CRFs can measure properties of segments, and transitions wit hin a segment can be non-Markovian. In spite of this additional power, exact lea rning and inference algorithms for semi-CRFs are polynomial-time—often only a sm all constant factor slower than conventional CRFs. In experiments on ■ve named e ntity recognition problems, semi-CRFs generally outper- form conventional CRFs.

****

Adaptive Manifold Learning
Jing Wang, Zhenyue Zhang, Hongyuan Zha

Recently, there have been several advances in the machine learning and pattern recognition communities for developing manifold learning algo- rithms to construct nonlinear low-dimensional manifolds from sample data points embedded in high-dimensional spaces. In this paper, we de- velop algorithms that address two key issues in manifold learning: 1) the adaptive selection of the neighborhood sizes; and 2) better fitting the local geometric structure to account for the variations in the curvature of the manifold and its interplay with the sampling density of the dat a set. We also illustrate the effectiveness of our methods on some syn thetic data sets.

****

Euclidean Embedding of Co-Occurrence Data
Amir Globerson, Gal Chechik, Fernando Pereira, Naftali Tishby

Embedding algorithms search for low dimensional structure in complex da

ta, but most algorithms only handle objects of a single type for which pairwise distances are specified. This paper describes a method for embedding objects of different types, such as images and text, into a single common Euclidean space based on their co-occurrence statistics. The joint distributions are modeled as exponentials of Euclidean distances in the low-dimensional embedding space, which links the problem to convex optimization over positive semidefinite matrices. The local structure of our embedding corresponds to the statistical correlations via random walks in the Euclidean space. We quantify the performance of our method on two text datasets, and show that it consistently and significantly outperforms standard methods of statistical correspondence modeling, such as multidimensional scaling and correspondence analysis.

****

Heuristics for Ordering Cue Search in Decision Making

Peter Todd, Anja Dieckmann

Simple lexicographic decision heuristics that consider cues one at a time in a particular order and stop searching for cues as soon as a decision can be made have been shown to be both accurate and frugal in their use of information. But much of the simplicity and success of these heuristics comes from using an appropriate cue order. For instance, the Take The Best heuristic uses validity order for cues, which requires considerable computation, potentially undermining the computational advantages of the simple decision mechanism. But many cue orders can achieve good decision performance, and studies of sequential search for data records have proposed a number of simple ordering rules that may be of use in constructing appropriate decision cue orders as well. Here we consider a range of simple cue ordering mechanisms, including tallying, swapping, and move-to-front rules, and show that they can find cue orders that lead to reasonable accuracy and considerable frugality when used with lexicographic decision heuristics.

****

Coarticulation in Markov Decision Processes

Khashayar Rohanimanesh, Robert Platt, Sridhar Mahadevan, Roderic Grupen

We investigate an approach for simultaneously committing to multiple activities, each modeled as a temporally extended action in a semi-Markov decision process (SMDP). For each activity we define a set of admissible solutions consisting of the redundant set of optimal policies, and those policies that ascend the optimal state-value function associated with them. A plan is then generated by merging them in such a way that the solutions to the subordinate activities are realized in the set of admissible solutions satisfying the superior activities. We present our theoretical results and empirically evaluate our approach in a simulated domain.

****

Integrating Topics and Syntax

Thomas Griffiths, Mark Steyvers, David Blei, Joshua Tenenbaum

Statistical approaches to language learning typically focus on either short-range syntactic dependencies or long-range semantic dependencies between words. We present a generative model that uses both kinds of dependencies, and can be used to simultaneously find syntactic classes and semantic topics despite having no representation of syntax or seman-t

ics beyond statistical dependency. This model is competitive on tasks like part-of-speech tagging and document classification with models that exclusively use short- and long-range dependencies respectively.

****

Object Classification from a Single Example Utilizing Class Relevance Metrics
Michael Fink

We describe a framework for learning an object classifier from a single example. This goal is achieved by emphasizing the relevant dimensions for classification using available examples of related classes. Learning to accurately classify objects from a single training example is often unfeasible due to overfitting effects. However, if the instance representation provides that the distance between each two instances of the same class is smaller than the distance between any two instances from different classes, then a nearest neighbor classifier could achieve perfect performance with a single training example. We therefore suggest a two stage strategy. First, learn a metric over the instances that achieves the distance criterion mentioned above, from available examples of other related classes. Then, using the single examples, define a nearest neighbor classifier where distance is evaluated by the learned class relevance metric. Finding a metric that emphasizes the relevant dimensions for classification might not be possible when restricted to linear projections. We therefore make use of a kernel based metric learning algorithm. Our setting encodes object instances as sets of locality based descriptors and adopts an appropriate image kernel for the class relevance metric learning. The proposed framework for learning from a single example is demonstrated in a synthetic setting and on a character classification task.

****

A Large Deviation Bound for the Area Under the ROC Curve
Shivani Agarwal, Thore Graepel, Ralf Herbrich, Dan Roth

The area under the ROC curve (AUC) has been advocated as an evalu- ation criterion for the bipartite ranking problem. We study large devi- ation properties of the AUC; in particular, we derive a distribution-free large deviation bound for the AUC which serves to bound the expected accuracy of a ranking function in terms of its empirical AUC on an inde- pendent test sequence. A comparison of our result with a corresponding large deviation result for the classi■cation error rate suggests that the test sample size required to obtain an -accurate estimate of the expected ac- curacy of a ranking function with δ-con■dence is larger than that required to obtain an -accurate estimate of the expected error rate of a classi■- cation function with the same con■dence. A simple application of the union bound allows the large deviation bound to be extended to learned ranking functions chosen from ■nite function classes.

****

Adaptive Discriminative Generative Model and Its Applications
Ruei-sung Lin, David Ross, Jongwoo Lim, Ming-Hsuan Yang

This paper presents an adaptive discriminative generative model that generalizes the conventional Fisher Linear Discriminant algorithm and renders a proper probabilistic interpretation. Within the context of object tracking, we aim to find a discriminative generative model that best separates the target from the background. We present a computationally efficient algorithm to constantly update this discriminative model as time progresses. While most tracking algorithms operate on the premise that the object appearance or ambient lighting condition does not significantly change as time progresses, our method adapts a discriminative generative model to reflect appearance variation of the target and background, thereby facilitating the tracking task in ever-changing environments. Numerous experiments show that our method is able to learn a discrimina- tive generative model for tracking target objects undergoing large pose and lighting changes.

****

Optimal sub-graphical models
Mukund Narasimhan, Jeff A. Bilmes
We investigate the problem of reducing the complexity of a graphical
model (G, PG) by finding a subgraph H of G, chosen from a class of su
bgraphs H, such that H is optimal with respect to KL-divergence. We d
o this by first defining a decomposition tree representation for G, which
is closely related to the junction-tree representation for G. We then give
an algorithm which uses this representation to compute the optimal H
H. Gavril [2] and Tarjan [3] have used graph separation properties t
o solve several combinatorial optimization problems when the size of
the minimal separators in the graph is bounded. We present an extensi
on of this technique which applies to some important choices of H eve
n when the size of the minimal separators of G are arbitrarily large.
In particular, this applies to problems such as finding an optimal s
ubgraphical model over a (k - 1)-tree of a graphical model over a k-t
ree (for arbitrary k) and selecting an optimal subgraphical model wit
h (a constant) d fewer edges with respect to KL-divergence can be sol
ved in time polynomial in |V (G)| using this formulation. 1 Intr
oduction and Preliminaries
****
Modeling Conversational Dynamics as a Mixed-Memory Markov Process
Tanzeem Choudhury, Sumit Basu
influences
****
Generalization Error Bounds for Collaborative Prediction with Low-Rank Matrices
Nathan Srebro, Noga Alon, Tommi Jaakkola
We prove generalization error bounds for predicting entries in a partially
observed matrix by fitting the observed entries with a low-rank matrix. In
justifying the analysis approach we take to obtain the bounds, we pr
esent an example of a class of functions of finite pseudodimension su
ch that the sums of functions from this class have unbounded pseudodi
mension.
****
Incremental Learning for Visual Tracking
Jongwoo Lim, David Ross, Ruei-sung Lin, Ming-Hsuan Yang
Most existing tracking algorithms construct a representation of a target
object prior to the tracking task starts, and utilize invariant features to
handle appearance variation of the target caused by lighting, pose, and
view angle change. In this paper, we present an efficient and effec-
tive online algorithm that incrementally learns and adapts a low dimen
- sional eigenspace representation to reflect appearance changes of th
e tar- get, thereby facilitating the tracking task. Furthermore, our i
ncremental method correctly updates the sample mean and the eigenbasis
, whereas existing incremental subspace update methods ignore the fact
the sample mean varies over time. The tracking problem is formulated
as a state inference problem within a Markov Chain Monte Carlo framewo
rk and a particle filter is incorporated for propagating sample distri
butions over time. Numerous experiments demonstrate the effectiveness
of the pro- posed tracking algorithm in indoor and outdoor environment
s where the target objects undergo large pose and lighting changes.
****
Face Detection --- Efficient and Rank Deficient
Wolf Kienzle, Matthias Franz, Bernhard Schölkopf, Gökhan Bakir
This paper proposes a method for computing fast approximations to sup-
port vector decision functions in the field of object detection. In the
present approach we are building on an existing algorithm where the set
of support vectors is replaced by a smaller, so-called reduced set of syn-
thesized input space points. In contrast to the existing method that finds
the reduced set via unconstrained optimization, we impose a structural
constraint on the synthetic points such that the resulting approximati

ons can be evaluated via separable filters. For applications that requi
re scan- ning large images, this decreases the computational complexity
by a sig- nificant amount. Experimental results show that in face dete
ction, rank deficient approximations are 4 to 6 times faster than uncon
strained re- duced set systems.
****

A Temporal Kernel-Based Model for Tracking Hand Movements from Neural Activities
Lavi Shpigelman, Koby Crammer, Rony Paz, Eilon Vaadia, Yoram Singer
We devise and experiment with a dynamical kernel-based system for track
ing hand movements from neural activity. The state of the system corres
ponds to the hand location, velocity, and acceleration, while the syste
m's input are the instantaneous spike rates. The system's state dy- nam
ics is defined as a combination of a linear mapping from the previous e
stimated state and a kernel-based mapping tailored for modeling neural
activities. In contrast to generative models, the activity-to-state mapping
is learned using discriminative methods by minimizing a noise-robust
loss function. We use this approach to predict hand trajectories on the
basis of neural activity in motor cortex of behaving monkeys and find
that the proposed approach is more accurate than both a static approach
based on support vector regression and the Kalman filter.
****

On-Chip Compensation of Device-Mismatch Effects in Analog VLSI Neural Networks
Miguel Figueroa, Seth Bridges, Chris Diorio
Device mismatch in VLSI degrades the accuracy of analog arithmetic cir
cuits and lowers the learning performance of large-scale neural net- w
orks implemented in this technology. We show compact, low-power on-chi
p calibration techniques that compensate for device mismatch. Our tech
niques enable large-scale analog VLSI neural networks with learn- ing
performance on the order of 10 bits. We demonstrate our techniques on
a 64-synapse linear perceptron learning with the Least-Mean-Squares (L
MS) algorithm, and fabricated in a 0.35m CMOS process.
****

At the Edge of Chaos: Real-time Computations and Self-Organized Criticality in R
ecurrent Neural Networks
Nils Bertschinger, Thomas Natschläger, Robert Legenstein
In this paper we analyze the relationship between the computational ca-
pabilities of randomly connected networks of threshold gates in the time-
series domain and their dynamical properties. In particular we propose
a complexity measure which we find to assume its highest values near
the edge of chaos, i.e. the transition from ordered to chaotic dynamic
s. Furthermore we show that the proposed complexity measure predicts
the computational capabilities very well: only near the edge of chaos
are such networks able to perform complex computations on time serie
s. Ad- ditionally a simple synaptic scaling rule for self-organized c
riticality is presented and analyzed.
****

Trait Selection for Assessing Beef Meat Quality Using Non-linear SVM
Juan Coz, Gustavo Bayón, Jorge Díez, Oscar Luaces, Antonio Bahamonde, Carlos Sañ
udo
In this paper we show that it is possible to model sensory impressions
of consumers about beef meat. This is not a straightforward task; the r
eason is that when we are aiming to induce a function that maps object
descriptions into ratings, we must consider that consumers' ratings are
just a way to express their preferences about the products presented in
the same testing session. Therefore, we had to use a special purpose
SVM polynomial kernel. The training data set used collects the ratings of
panels of experts and consumers; the meat was provided by 103 bovines
of 7 Spanish breeds with different carcass weights and aging periods.
Additionally, to gain insight into consumer preferences, we used feature
subset selection tools. The result is that aging is the most important trait

for improving consumers' appreciation of beef meat.
****
Sharing Clusters among Related Groups: Hierarchical Dirichlet Processes
Yee Teh, Michael Jordan, Matthew Beal, David Blei
We propose the hierarchical Dirichlet process (HDP), a nonparametric Bayesian mo del for clustering problems involving multiple groups of data. Each group of dat a is modeled with a mixture, with the number of components being open-ended and inferred automatically by the model. Further, components can be shared across gr oups, allowing dependencies across groups to be modeled effectively as well as c onferring generaliza- tion to new groups. Such grouped clustering problems occur often in practice, e.g. in the problem of topic discovery in document corpora. We report experimental results on three text corpora showing the effective and s uperior performance of the HDP over previous models.
****
Spike-timing Dependent Plasticity and Mutual Information Maximization for a Spik ing Neuron Model
Taro Toyoizumi, Jean-pascal Pfister, Kazuyuki Aihara, Wulfram Gerstner
We derive an optimal learning rule in the sense of mutual information maximizati on for a spiking neuron model. Under the assumption of small ■uctuations of the input, we ■nd a spike-timing dependent plas- ticity (STDP) function which depend s on the time course of excitatory postsynaptic potentials (EPSPs) and the autoc orrelation function of the postsynaptic neuron. We show that the STDP function h as both positive and negative phases. The positive phase is related to the shape of the EPSP while the negative phase is controlled by neuronal refractoriness.
****
Approximately Efficient Online Mechanism Design
David C. Parkes, Dimah Yanovsky, Satinder Singh
Online mechanism design (OMD) addresses the problem of sequential dec ision making in a stochastic environment with multiple self-interested agents. The goal in OMD is to make value-maximizing decisions despite this self-interest. In previous work we presented a Markov decision pro- cess (MDP)-based approach to OMD in large-scale problem domains. In practice the underlying MDP needed to solve OMD is too large and hence the mechanism must consider approximations. This raises the pos- sibility that agents may be able to exploit the approximation for selfish gain. We adopt sparse-sampling-based MDP algorithms to implement - efficient policies, and retain truth-revelation as an approximate Bayesia n- Nash equilibrium. Our approach is empirically illustrated in the c ontext of the dynamic allocation of WiFi connectivity to users in a c offeehouse.
****
Comparing Beliefs, Surveys, and Random Walks
Erik Aurell, Uri Gordon, Scott Kirkpatrick
Survey propagation is a powerful technique from statistical physics that has been applied to solve the 3-SAT problem both in principle and in practice. We give, using only probability arguments, a common deriva- t ion of survey propagation, belief propagation and several interesting hy- brid methods. We then present numerical experiments which use WSAT (a widely used random-walk based SAT solver) to quantify the complex- it y of the 3-SAT formulae as a function of their parameters, both as ran- domly generated and after simplication, guided by survey propagation. Some properties of WSAT which have not previously been reported make it an ideal tool for this purpose its mean cost is proportional to the num- ber of variables in the formula (at a xed ratio of clauses to variables) in the easy-SAT regime and slightly beyond, and its behavior in the hard- SAT regime appears to reect the underlying structure of the solution space that has been predicted by replica symmetry-breaking arguments. An analysis of the tradeoffs between the various methods of search for satisfying assignments shows WSAT to be far more powerful than has been appreciated, and suggests some interesting new directions for prac-

tical algorithm development.

****

Theories of Access Consciousness

Michael Colagrosso, Michael C. Mozer

Theories of access consciousness address how it is that some mental states but not others are available for evaluation, choice behavior, and verbal report. Farah, O'Reilly, and Vecera (1994) argue that quality of representation is critical; Dehaene, Sergent, and Changeux (2003) argue that the ability to communicate representations is critical. We present a probabilistic information transmission or PIT model that suggests both of these conditions are essential for access consciousness. Having successfully modeled data from the repetition priming literature in the past, we use the PIT model to account for data from two experiments on subliminal priming, showing that the model produces priming even in the absence of accessibility and reportability of internal states. The model provides a mechanistic basis for understanding the dissociation of priming and awareness.

****