Where will malicious traffic on the Internet originate from?

We adaptively prune a decision tree over IP address space (2^{32} leaves!) ... 

... combining “experts” algorithms and online paging...

...and find dynamic maps of Internet malicious activity!

Legend:
- Good IP
- Bad IP
Skill Discovery in Continuous Reinforcement Learning Domains using Skill Chaining

George Konidaris and Andrew Barto, *University of Massachusetts Amherst*

- Skill discovery: autonomous acquisition of new high-level actions by RL agents.
- Skill chaining is a skill discovery method for continuous RL domains.
- We construct chains of skills leading to initial target event (e.g., end-of-episode).
- We demonstrate significant performance improvements in *Pinball*, a challenging dynamic continuous domain.

Skill chaining breaks the solution of a difficult dynamic control problem into a chain of subskills (left). We thus obtain better policies (below) using dedicated FAs for each skill.
Efficient Match Kernels (EMK) between Sets of Features for Visual Recognition

Liefeng Bo (TTI-Chicago), Cristian Sminchisescu (University of Bonn)

Bag-of-Words+Linear Kernel = Special Match Kernel however, the quantization is too coarse

Accurate Quantization: Gaussian kernel can be used as a local kernel, but this is computationally prohibitive

Our Solutions: Approximate the Gaussian kernel with low dimensional features (linear complexity)

(1) Constrained Kernel Singular Value Decomposition
(2) Random Fourier Set Features

Caltech-101: highly competitive results compared with ten recent algorithms.
Consider OMP recovery of sparse vector $x$ from linear measurements:

$$y = \Phi x + w$$

**Sufficient** condition for detection of $k$-sparse $x$, $k_{\text{min}} \leq k \leq k_{\text{max}}$, with large random $\Phi \in \mathbb{R}^{m \times n}$:

$$m > 2k_{\text{max}} \log(n - k_{\text{min}})$$

- Improves on previous scaling
  $$m > 4k \log(n)$$
- Allows noise and uncertainty in $k$
- Same scaling law as lasso method [Wainwright 06]

Simulated $P($misdetection$)$ for OMP at $n = 100$. Line: theoretical sufficient condition for no errors at infinite $n$

**OMP can have similar performance as more complex lasso**
Fast nearest neighbor problems arise in many domains

Goal of this paper: learn hash functions for nearest neighbor search
- Minimize reconstruction error between input distances and hashed Hamming distances
- Scalable coordinate-descent algorithm
- Unlike previous methods (e.g. LSH, spectral hashing, semantic hashing), no assumptions about data distribution, and it can easily be kernelized

Results show competitive performance as compared to existing methods
Perceptual Multistability as Markov Chain Monte Carlo Inference (#991)

Samuel J. Gershman  Edward Vul & Joshua B. Tenenbaum
Princeton University  Massachusetts Institute of Technology

Why does perception alternate?
Posterior over image interpretations is multimodal. Perceptual system approximates posterior by drawing samples from a Markov chain.

Necker cube

Markov random field image models

Markov chain
Information-Theoretic Lower Bounds on the Oracle Complexity of Convex Optimization

Alekh Agarwal, Peter Bartlett, Pradeep Ravikumar, Martin Wainwright

Statistical Learning

Stochastic Convex Optimization

\[ \ell(h_T) \leq \ell(h^*) + \varepsilon \]

\[ f(x_T) \leq f(x^*) + \varepsilon \]

How big can T be so no method can achieve an error \( \varepsilon \)?
How do humans learn separable dimensions?

We model dimensional bias by learning the base distribution of a Dirichlet Process.

We correctly model children’s learning of dimensional biases without knowing the dimensional basis a priori!
Lower Bounds on Minimax Rates for Non-parametric Sparse Additive Models
Garvesh Raskutti, Martin Wainwright, Bin Yu

SPAM:
\[ y = h(X) + \alpha + \text{noise} \]
\[ |S| \leq s \ll p \]
\( \alpha \) is sparse

What is minimum \( n \) required?

Two separate problems!!

Subset Recovery
Find subset \( S \)
Rate: \( \frac{s \log(p/s)}{n} \)

Non-parametric Regression
Estimate \( h_S \)
Rate: \( \sigma^2_n \)
Neural networks for hierarchical Bayesian inference by importance sampling

Building block: importance sampler

Hierarchical inference by recursive importance sampling

Applications in neuroscience

Combination of cues from sensory modalities

Explaining oblique effect in visual perception
Posterior Sparsity vs. Parameter Sparsity

João Graça  Kuzman Ganchev  Ben Taskar  Fernando Pereira
L2F INESC-ID  University of Pennsylvania  Google Research
Lisboa, Portugal  Philadelphia, PA, USA  Mountain View, CA, USA

**Motivation**

E[degree] = 10000  E[degree] = 1.5

- JJ: car
- VB: object
- NN: offensive
- JJ: romantic
- NN: being

**Posterior vs. Parameter**

\[ p(t \mid w) \quad p(w \mid t) \]

\[ \begin{array}{c}
W_1 \\
T_1 \\
W_2 \\
T_2 \\
W_3 \\
T_3
\end{array} \]

**EM**

**Sparse**

**Mutual Information**

- None
- Param
- Post

**Sparsity Method**
Structural inference affects depth perception in the context of potential occlusion

Ian H. Stevenson and Konrad P. Körding
Departments of Physiology and Physical Medicine and Rehabilitation
Northwestern University, Chicago, Illinois, USA

An ordinal model of probabilistic cue combination...

Explains depth perception during occlusion...

Nakayama & Shimojo (1990)
New Experiment
Main Ideas:
- Treat image features and templates as matrices or tensors rather than simply “vectorizing” them.
- Learn a low rank template by performing alternating minimization using standard linear SVM solver.

Advantages:
1. Rank constraint provides natural regularization.
2. Learn shared subspaces across multiple classes or training sets to allow for transfer learning.
3. 10x runtime speedup with no loss in performance.

Mathematical Formulation:
\[
\frac{1}{2} Tr(W^T W) + C \sum_n \max(0, 1 - y_n Tr(W^T X^n))
\]

\[
W = W_{xy} W_{xf}^T
\]