Summary of ICML’04
International Conference on Machine Learning

Russ Greiner
ICML’04 Program CoChair

Co-ProgramChair: Dale Schuurmans
GeneralChair: Carla Brodley

In the beautiful Canadian Rockies...

Twenty-First International Conference on Machine Learning
Banff, Alberta
4 – 8 July 2004
Outline

- Statistics
- High Points
- Comments

Submissions

- 368 submissions
- 118 acceptances
  - 32% acceptance rate
Support vector machines
Bayesian learning
Kernels
Semi-supervised learning
Reinforcement learning
Feature selection
Empirical evaluation
Statistical learning

Top 10 Submitted Topics: 2004

<table>
<thead>
<tr>
<th>Rank</th>
<th>Topic</th>
<th># Submissions</th>
<th># Accepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Statistical learning</td>
<td>49</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>Empirical evaluation</td>
<td>38</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Feature selection</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>Reinforcement learning</td>
<td>36</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>Semi-supervised learning</td>
<td>29</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>Kernels</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>On-line learning</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>Clustering</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>Bayesian learning</td>
<td>24</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>Support vector machines</td>
<td>24</td>
<td>15</td>
</tr>
</tbody>
</table>

Top 10 Submitted Topics: 2003

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical learning</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>Reinforcement learning</td>
<td>58</td>
<td>4</td>
</tr>
<tr>
<td>SVM + kernel</td>
<td>49</td>
<td>≈6,10</td>
</tr>
<tr>
<td>Applications</td>
<td>48</td>
<td>-</td>
</tr>
<tr>
<td>Clustering</td>
<td>47</td>
<td>8</td>
</tr>
<tr>
<td>Evaluation</td>
<td>46</td>
<td>2</td>
</tr>
<tr>
<td>Learning w/unlabeled data</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td>Feature selection</td>
<td>36</td>
<td>3</td>
</tr>
<tr>
<td>Text classification</td>
<td>32</td>
<td>-</td>
</tr>
<tr>
<td>Markov models (HMM, POMDPs,...)</td>
<td>25</td>
<td>≈9</td>
</tr>
</tbody>
</table>

Way back... 1994 ...

ICML 1994 Papers...
Both Then (1994) and Now (2004)

- Classification
  - Supervised learning
  - Boosting
- Reinforcement Learning
  - MDP, POMDP, …
- Applications
  - Bioinformatics
  - Planning
  - …

Then vs Now …

1994:
- Dimensionality Reduction
- Feature Selection

2004:
- Dimensionality Reduction
  - PCA, …

Then vs Now …

1994:
- Probabilistic Models
  - Rare…
  - Bayesian nets

2004:
- Probabilistic Models
- Everywhere!
  - Bayesian nets, Markov Random Fields, Conditional Random Fields,

Then vs Now …

1994:
- Symbolic learning
  - Decision Tree
  - Inductive Logic Programming

2004:
- Statistical Learning
  - 49/118!
  - Lasso, EM, MCMC,
  - Robust,
  - …
Then vs Now ...

1994:
- Concept Learning
- Neural Nets

2004:
- SVMs
- Kernel
- Numeric Methods
  - Linear Algebra
  - ...
  - Clustering
  - 12 papers, vs 0

1994:
- Validation:
  - Better than C4.5 on some UCI

2004:
- Validation:
  - Better than SVM on MANY UCI
- Some Real Applications
  - Most involve Real applications

How to get a paper accepted in ICML 2004???

| W     | P(accept|w) | W           | P(rejected|w) |
|-------|----------|-------------|-----------|
| random | 0.83     | comparison  | 0.90      |
| svms   | 0.80     | unlabeled   | 0.89      |
| fields | 0.80     | time        | 0.89      |
| under  | 0.75     | programming | 0.89      |
| skewing| 0.75     | series      | 0.87      |
| sequences | 0.75 | framework   | 0.87      |
| segmenting | 0.75 | error       | 0.87      |
| relevance| 0.75    | empirical   | 0.87      |
| redundant| 0.75    | application | 0.87      |
| multiplicative | 0.75 | noise       | 0.86      |
| motion | 0.75     | new         | 0.86      |

Attendance

- ICML’04 Conference: 397
- + 30 for Workshop/Tutorial
- Collocated with
  - UAI’04
  - COLT ’04

397
Workshops
- Statistical Relational Learning: 84
- Relational Reinforcement Learning: 28
- Predictive Representations of World Knowledge: 23
- Physiological Data Modeling: 9

Tutorials
- Bayesian Methods for Machine Learning: 144
- Kernels for Structured Data: 127
- Data Structures for Fast Statistics: 90
- Game-theoretic Learning: 81
- Spectral Clustering: 61
- Probabilistic Logic Learning: 60
- Junk E-mail Filtering: 50
- The Many Faces of ROC Analysis: 42

Outline
- Statistics
- High Points
  - Invited Speakers
  - Best Papers
  - Kernel Day
- Comments

Invited Presentations
- Identify areas ADJACENT to ML
  - BioInformatics
  - Vision
  - Computational Finance
- Invite leading exponent of that area
Invited Talk: Bioinformatics

- **Gene Myers**
  - Celera Genomics
  - **BLAST**
    - ACM Kanellakis Prize in 2001
    - Member: National Academy of Engineering

- Whole Genome Sequencing, Comparative Genomics, and Systems Biology

Invited Talk: Vision

- **Michael Black**
  - leading researcher in computer vision & AI
  - *Learning to See People*

  - Motion Capture
    - Loss of 3D in 2D projection
    - Unusual poses
    - Self occlusion
    - Low contrast

Invited Presentations: Computational Finance

- **#1:** 2003 Nobel Prize Winner, Economics
  - for methods of analyzing economic time series with time-varying volatility (ARCH)
  - Robert Engle

- **#2:** ...
- **#3:** ...
- **#4:** ...
- ...
Models of Dynamic Uncertainty in Univariate and Multivariate Systems

ARCH Model...

- Autoregressive Conditional Heteroskedasticity
  - Predictive (conditional)
  - Uncertainty (heteroskedasticity)
  - That fluctuates over time (autoregressive)

- Common use ...
  - Matlab package...

HOW ARE OPTIONS PRICED?
Outstanding Student Papers

- Generalized Low Rank Approximations of Matrices
  - Jieping Ye

- Decentralized Detection and Classification using Kernel Method
  - XuanLong Nguyen, Martin Wainwright, Michael Jordan

- Learning a Kernel Matrix for Nonlinear Dimensionality Reduction
  - Kilian Weinberger, Fei Sha, Lawrence Saul

---

Generalized Low Rank Approximations of Matrices

- Jieping Ye

Useful to find Low-Rank approximation of information
Most view Matrix as Vector
Why not view Matrix as Matrix?
- GLRAM

---

Decentralized Detection and Classification using Kernel Method

- XuanLong Nguyen
- Martin Wainwright
- Michael Jordan

Problem: Given training data \((x_i, y_i)\) for \(i = 1, n\)
Find: decision rules \((\gamma_1, \ldots, \gamma_s; \gamma)\) that minimize misclassification probability: \(P(Y \neq \gamma(Z_1, \ldots, Z_s))\)
Learning a Kernel Matrix for Nonlinear Dimensionality Reduction

- Kilian Weinberger
- Fei Sha
- Lawrence Saul

Honorable Mention for Outstanding Paper Award

- Multiple Kernel Learning, Conic Duality, and the SMO Algorithm
  - Francis Bach, Gert Lanckriet, Michael Jordan
- Efficient Hierarchical MCMC for Policy Search
  - Malcolm Strens
- Authorship Verification as a One-Class Classification Problem
  - Moshe Koppel, Jonathan Schler

Kernel Trick

\[ \phi : (x_1, x_2) \rightarrow (x_1^2, \sqrt{2}x_1x_2, x_2^2) \]

\[ (\frac{\gamma}{2})^2 + (\frac{\eta}{2})^2 = 1 \rightarrow \frac{\gamma}{2} + \beta = 1 \]

Kernel Day

COLT Session: Kernels

- Bayesian Networks and Inner Product Spaces
- Inequality for Nearly Log-concave Distributions with Applications to Learning
- Bayes and Tukey Meet at the Center Point
- Sparseness vs Estimating Conditional Probabilities: Some Asymptotic Results
- A Statistical Mechanics Analysis of Gram Matrix Eigenvalue Spectra
- Statistical Properties of Kernel Principal Component Analysis
- Kernelizing Sorting, Permutation and Alignment for Minimum Volume PCA
- Regularization and Semisupervised Learning on Large Graphs

ICML Session: Kernels

- Support Vector Machine Learning for Interdependent and Structured Output Spaces
- Probabilistic Tangent Subspace: A Unified View
- Bayesian Inference for Transductive Learning of Kernel Matrix Using the Tanner-Wong Data Augmentation Algorithm
- Kernel-Based Discriminative Learning Algorithms for Labeling Sequences, Trees and Graphs
- A Kernel View of the Dimensionality Reduction of Manifolds
- Multiple Kernel Learning, Conic Duality, and the SMO Algorithm
- Learning with Non Positive Kernels
- Extensions of Marginalized Graph Kernels
**Distribution of Papers**

- Lots of numerical methods
  - Linear algebra
  - Dimensionality reduction techniques
  - …
- Better know the “kernel trick”!
- Is ICML becoming more like NIPS?

**Similarity (cosine) of**

- counts of all words in NIPS-xx
- counts of all words in ICML-xx

---

**Field of Machine Learning**

- Conference went VERY VERY well.
- Because the field is ... Alive and Well!
- Lots of new Ideas and Insights
- Many new challenges!
- Interacting with + Contributing to...
  - NLU, BioInfo, Vision, Robotics, Web, ...

- *If it deals with real data, it’s machine learning.*
- ... most of AI now deals with real data!

---

**ICML’04 ~ NIPS ??**

- Yes, NIPS-xx titles are getting more similar to ICML-xx each year...
- But ... ICML’04 is still very similar to earlier ICML-xx’s

---

**Sponsors**

- **Local (Alberta)**
- **Gov’t (US)**
- **Industry**
Thanks to …

- All Award Winners, Invited Speakers
  - for giving great talks then
  - sending me their slides!
- ICML’04 Committee
  - Carla Brodley, Johannes Furnkranz, Rob Holte,
  - Michael DeMarco, Kiri Wagstaff, Jennifer Dy, Stephen Scott,
  - David Jensen, David Woloschuk + web team, C-H Lee
- Area chairs, reviewers, authors, attendees, ...
- My funders
  - Alberta Ingenuity Centre for Machine Learning
  - NSERC
  - UofAlberta Dept of Computing Science

Questions?

- ICML 2004: 4-8/July Banff, Alberta
  - http://www.aicml.cs.uaicerta.ca/_banff04/icml/
- ICML 2005: 7-11/Aug Bonn, Germany
- ICML 2006: 24-30/June Pittsburgh, PA
  - http://icml2006.org/
- Pictures…
Scenes of ICML’04
International Conference on Machine Learning

Thanks to David Woloschuk
In the beautiful Canadian Rockies...

Twenty-First International Conference on Machine Learning
Banff, Alberta
4 – 8 July 2004
Gene Myers

Whole Genome Sequencing, Comparative Genomics, & Systems Biology

Gene Myers
University of California
Berkeley
Computation of marginalized kernels

(Tsuda et al, 2002; Jaakkola & Haussler, 1999)

- Marginalized kernel $K_Q(x, x')$ is defined as:

$$K_Q(x, x') := \sum_{z, z'} Q(z|x)Q(z'|x') K_z(z, z'),$$

Factorized distributions Base kernel

- If $K_z(z, z')$ is decomposed into smaller components of $z$ and $z'$, then $K_Q(x, x')$ can be computed efficiently (in polynomial-time).
- Original size: \(6570 \times 413 \times 798 \times 8 = 17.3\) GB.
- Reduced size: \((6570 \times 25 \times 50 + 413 \times 25 + 798 \times 50) \times 8 = 66.2\) MB.
- Compression ratio = 17.3 GB/66.2 MB = 262.
- SVD is not applicable for this dataset.
- GLRAM scales to large datasets.
GLRAM: The main results

- **Theorem 2** Let \( L, R \), and \( \{M_i\}_{i=1}^n \) minimize \( \sum_{i=1}^n \| A_i - LM_i R^T \|_F^2 \) subject to \( L^T L = I_{\ell_2} \) and \( R^T R = I_{\ell_2} \). Then for every \( i \), \( M_i = L^T A_i R \).

  - Expand \( \sum_{i=1}^n \| A_i - LM_i R^T \|_F^2 \) as
    \[
    \sum_{i=1}^n \| A_i - LM_i R^T \|_F^2 = \sum_{i=1}^n \text{trace} \left( (A_i - LM_i R^T)(A_i - LM_i R^T)^T \right) \\
    = \sum_{i=1}^n \text{trace}(A_i A_i^T) + \sum_{i=1}^n \text{trace}(M_i M_i^T) - 2 \sum_{i=1}^n \text{trace}(LM_i R^T A_i^T).
    \]

  - Take the derivative with respect to \( M_i \) and get \( 2M_i - 2L^T A_i R = 0 \iff M_i = L^T A_i R \).

- **Theorem 3** Let \( L \) and \( R \) be defined as in Theorem 2. Then \( L \) and \( R \) maximize \( \sum_{i=1}^n \| L^T A_i R \|_F^2 \).

  - Plugging \( M_i = L^T A_i R \) into \( \sum_{i=1}^n \| A_i - LM_i R^T \|_F^2 \), we obtain
    \[
    \sum_{i=1}^n \| A_i - LM_i R^T \|_F^2 = \sum_{i=1}^n \| A_i \|_F^2 - \sum_{i=1}^n \| L^T A_i R \|_F^2.
    \]

  - The minimization of \( \sum_{i=1}^n \| A_i - LM_i R^T \|_F^2 \) is equivalent to the maximization of \( \sum_{i=1}^n \| L^T A_i R \|_F^2 \).
Support Vector Machines

What about 3D Pose?

Contour Points / Shape Model

30+ dimensions

Learned mapping

K. Grauman, G. Shakhnarovich, T. Darrell, ICCV’03
Single View to 3D Pose

Given synthetic training data, learn the mapping from silhouette contours to 3D pose.

“Gaussian kernel RVM”, Agarwal and Triggs CVPR04

Problems

Accidental alignment

Motion blur.
(nothing to match)