

# International Conference on Machine Learning

Corvallis, Oregon, June 20-24 2007

Summary Presentation:  
Statistics, Awards, Comments

Zoubin Ghahramani  
ICML 2007 Program Chair

# Links

- Conference:  
<http://oregonstate.edu/conferences/icml2007/>
- Proceedings are Online:  
<http://www.machinelearning.org/>
- Videos are Online! :  
[http://videlectures.net/icml07\\_corvallis/](http://videlectures.net/icml07_corvallis/)

# ICML 2007: some statistics

- 522 submissions
- 150 accepted
- 29% acceptance rate

...can we predict acceptance?

# ML on ICML

Thanks to Ricardo Silva

- Extracted stems of words from titles and abstracts of all submitted papers.
- Computed  $P(\text{word} \mid \text{accepted})$  by counting the proportion of accepted papers that included that particular word.
- $P(\text{word} \mid \text{not accepted})$  computed analogously.
- Ranked each word by the odds of acceptance:  
 $P(\text{accepted} \mid \text{word}) / P(\text{not accepted} \mid \text{word})$
- Prepositions and adverbs were removed.
- The top 20 stems according to this are...

# ICML 2007

## Best Keyword Awards

Sponsored by the ???

# Top 20 words

- |              |                |
|--------------|----------------|
| 1. graphic   | 11. track      |
| 2. laplacian | 12. explain    |
| 3. unit      | 13. commun     |
| 4. exponenti | 14. intract    |
| 5. dirichlet | 15. walk       |
| 6. share     | 16. three      |
| 7. research  | 17. large-scal |
| 8. contrast  | 18. suit       |
| 9. view      | 19. degree     |
| 10. faster   | 20. go         |

# Bottom 20 words

- |              |              |
|--------------|--------------|
| 1. investig  | 11. deal     |
| 2. superior  | 12. expect   |
| 3. classifi  | 13. normal   |
| 4. origin    | 14. learner  |
| 5. ensembl   | 15. identifi |
| 6. individu  | 16. type     |
| 7. amount    | 17. subset   |
| 8. reward    | 18. unknown  |
| 9. version   | 19. recognit |
| 10. reinforc | 20. extract  |

# ICML 2007

## Best Student Paper Awards

Sponsored by the  
Machine Learning journal



# Selection Process

- Papers were nominated by the SPCs + top scoring papers: total 10 nominations
- Panel of 5 SPCs from different areas voted on the papers (+ and - votes) based on papers and reviews.
- Checked student status!
- Selected top 4 papers for award.
- The winners are (in alphabetical order):

# Winners are:

- Solving MultiClass Support Vector Machines with LaRank
  - \* **Antoine Bordes** - LIP, Universite de Paris 6
  - Leon Bottou - NEC Laboratories America
  - Patrick Gallinari - LIP, Universite de Paris 6
  - Jason Weston - NEC Laboratories America

# Winners are:

- Information-Theoretic Metric Learning
  - \* **Jason V. Davis** - University of Texas at Austin
  - \* **Brian Kulis** - University of Texas at Austin
  - \* **Prateek Jain** - University of Texas at Austin
  - \* **Suvrit Sra** - University of Texas at Austin
  - Inderjit S. Dhillon - University of Texas at Austin

# Winners are:

- Supervised Clustering of Streaming Data for Email Batch Detection
  - \* **Peter Haider** - MPI for Computer Science
  - \* **Ulf Brefeld** - MPI for Computer Science
  - Tobias Scheffer - MPI for Computer Science

# Winners are:

- Conditional Random Fields for Multi-agent Reinforcement Learning
  - \* **Xinhua Zhang** - CSL, RSISE, Australian National University, and SML NICTA
  - Douglas Aberdeen - NICTA, Australian National University
  - S.V.N. Vishwanathan - SML NICTA, and CSL, RSISE, Australian National University

ICML 2007

Business Meeting

# ICML 2007: some statistics

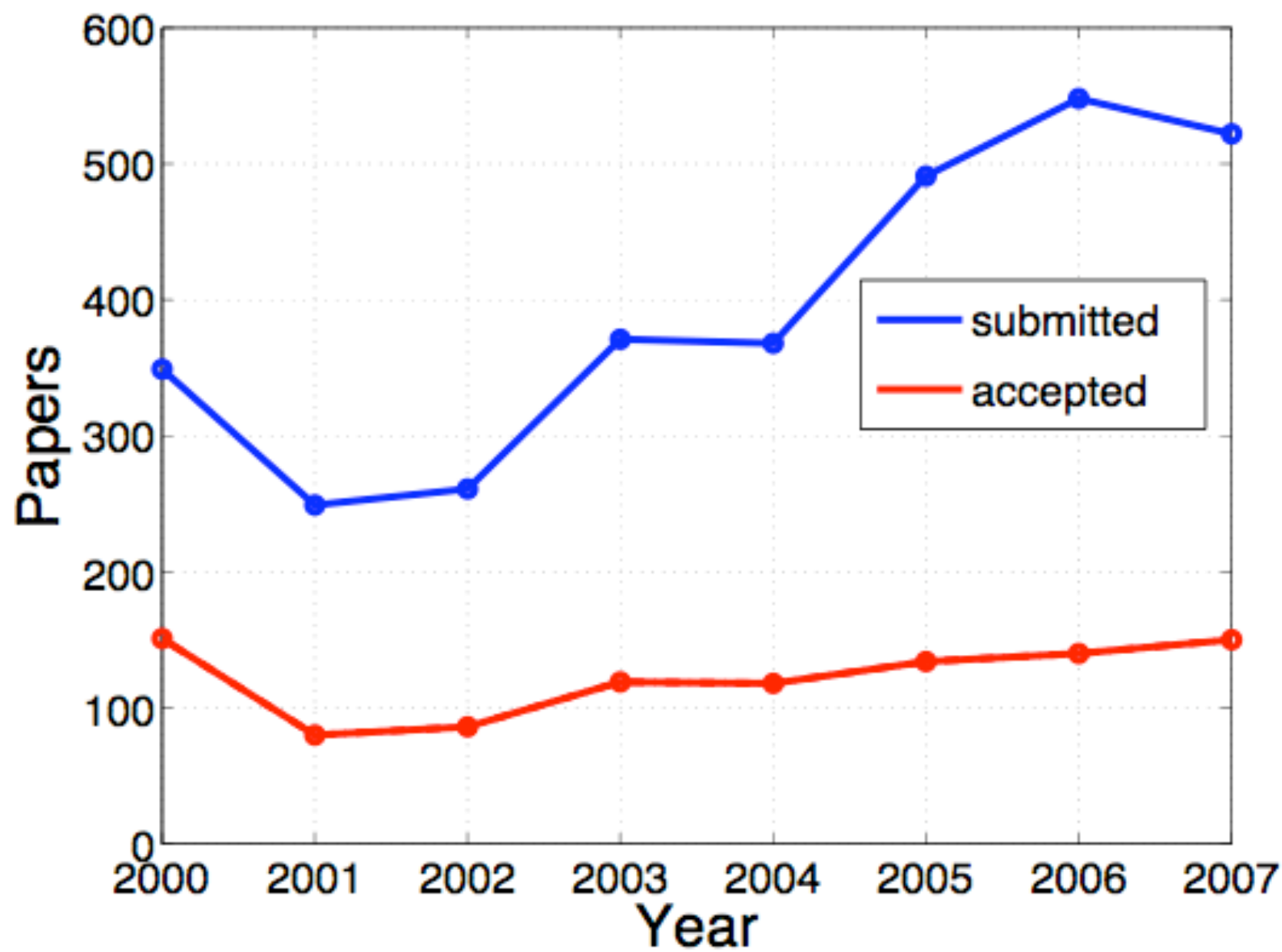
- 522 submissions
- 150 accepted
- 29% acceptance rate
- 2006: 140/548
- 2005: 134/491
- 2004: 118/368
- 2003: 119/371
- 2002: 86/261



doubling time = 5 years ← 261 x 2 = 522 !

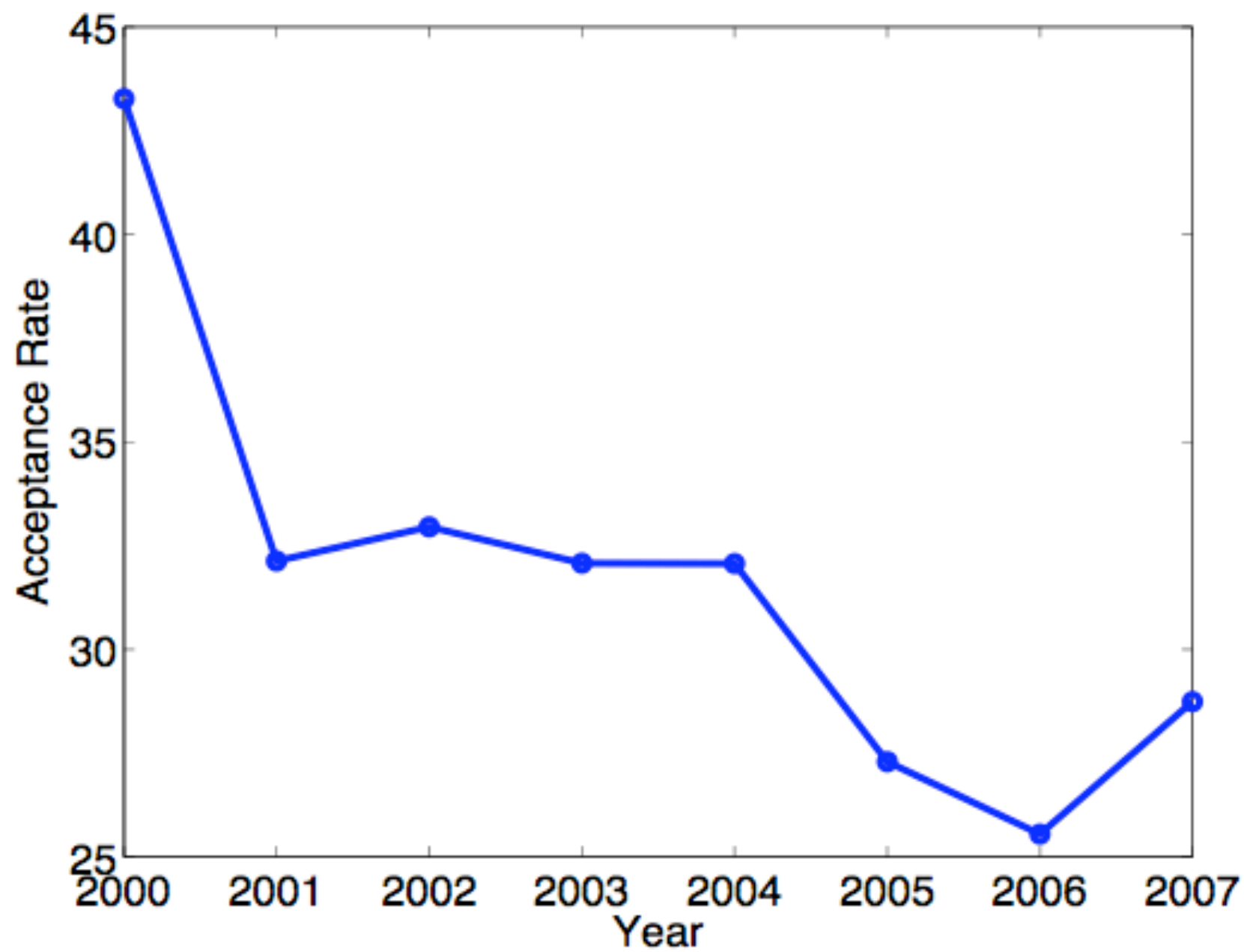
ICML 2062 will have 1,000,000 submissions :-)

# ICML





# ICML



# Geographical Distribution

The 150 accepted papers by first author:

- USA: 66
- Europe: 32
- China and Hong Kong: 19
- Canada: 11
- India: 6
- Australia: 5
- Japan: 5
- Israel: 3
- Korea, Russia and Taiwan: 1 each

# Topic Distribution

Top topics by number of submissions:

- Kernel methods and support vector machines 94
- Unsupervised learning, clustering 91
- Probabilistic approaches, graphical models 87
- Dimensionality reduction, manifolds and embedding 85
- Statistical models 82
- Reinforcement learning 64
- Semi-supervised learning 62
- Learning from structured data 54
- Bayesian methods 45
- Ensemble methods 37
- Applications and case studies 34
- Learning in vision 30

# Topic Distribution

Smallest topics by number of submissions:

- Cognitive aspects of learning 0
- Grammatical inference 1
- Collaborative filtering 4
- Evolutionary computation 6
- Scientific discovery 7
- Agent learning 12
- Learning in robotics 12
- Density estimation 12

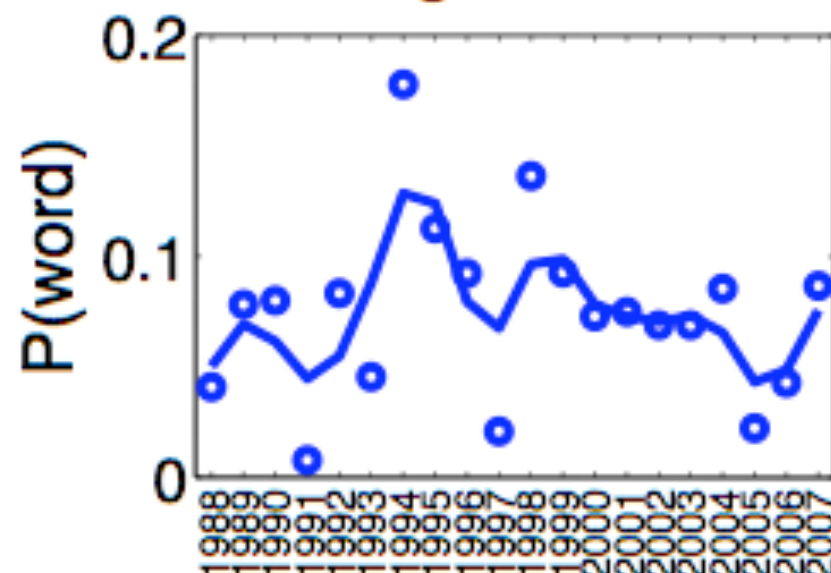
# New Topics (Keywords)

- 29: Gaussian processes 14
- 30: Bayesian methods 45
- 31: Multi-task and transfer learning 23
- 32: Semi-supervised learning 62
- 33: Dimensionality reduction, manifolds and embedding 85
- 34: Ranking and preference learning 20
- 35: Collaborative filtering 4
- 36: Active learning and experiment design 16
- 37: Density estimation 12
- 38: Sampling methods and MCMC 15

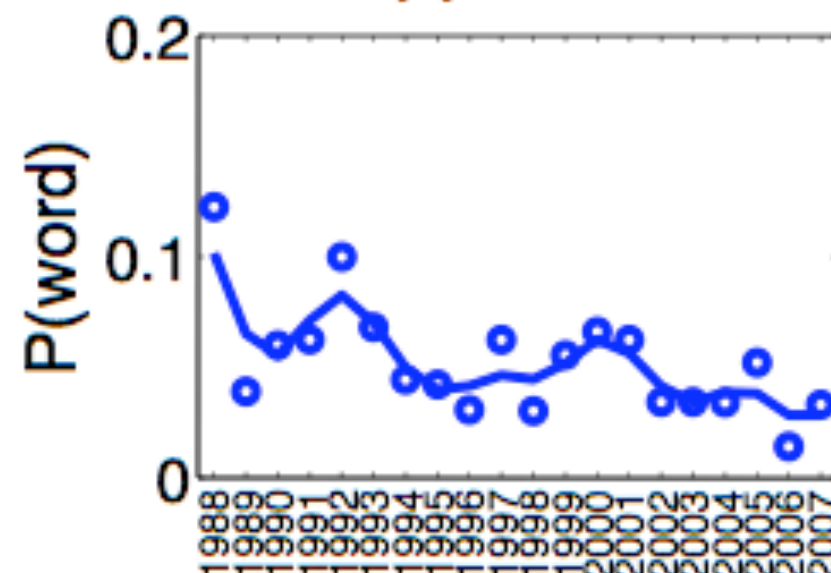
# ICML Topic evolution 1988-2007

- Thanks to Mark Reid for these statistics of ICML titles
- stemmed, picked ~20 most common words in titles for last 20 years
- smoothed and plotted

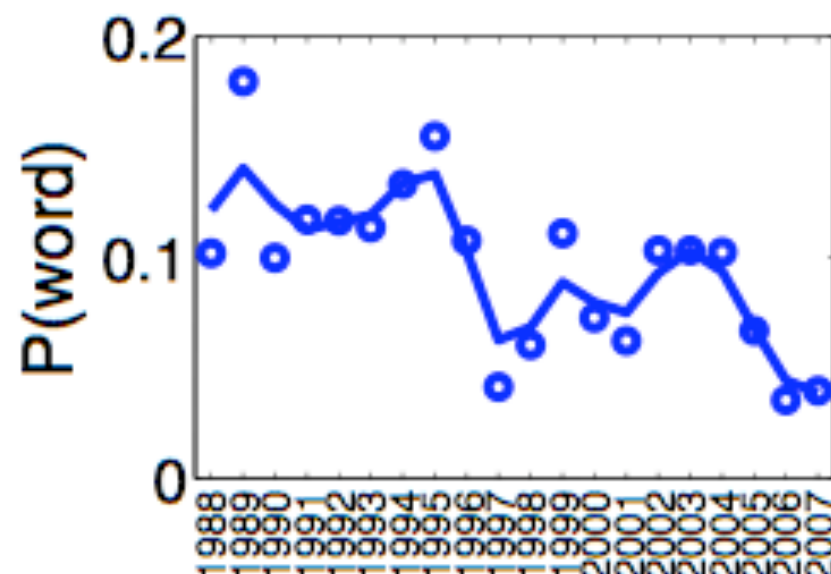
# algorithm



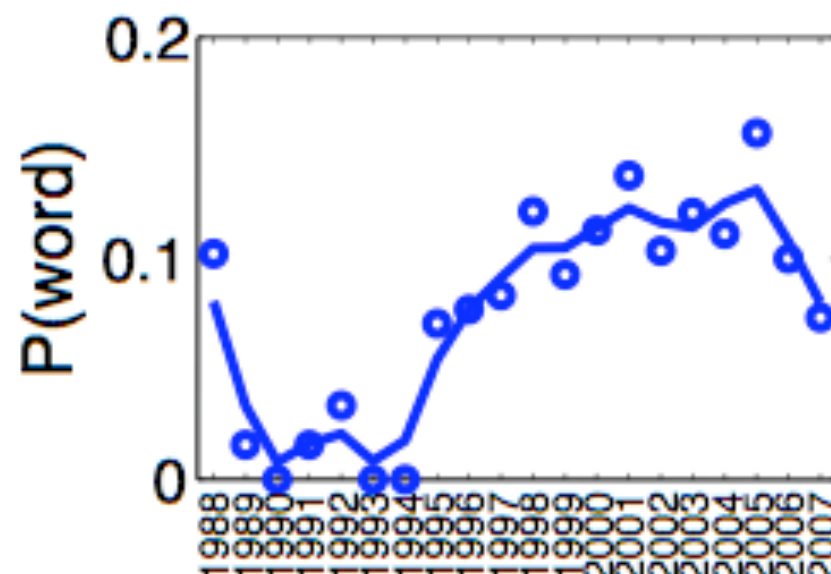
# approach



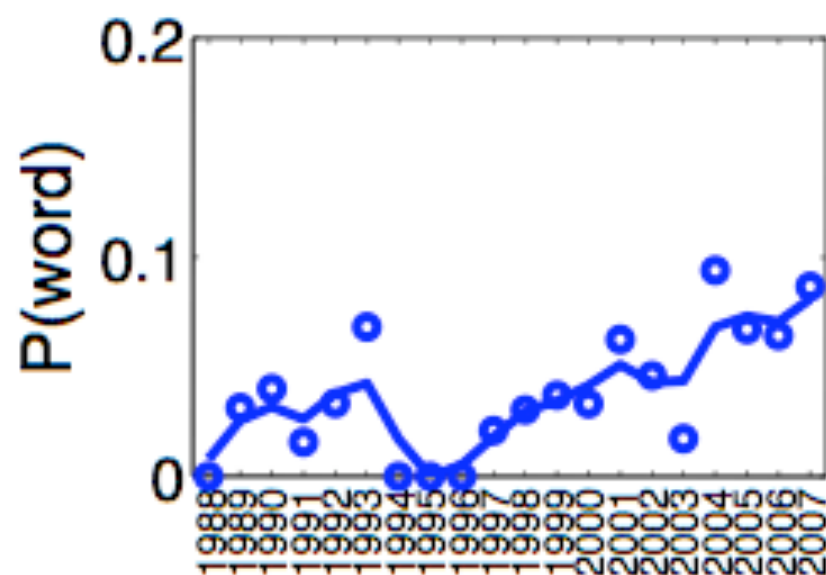
# base



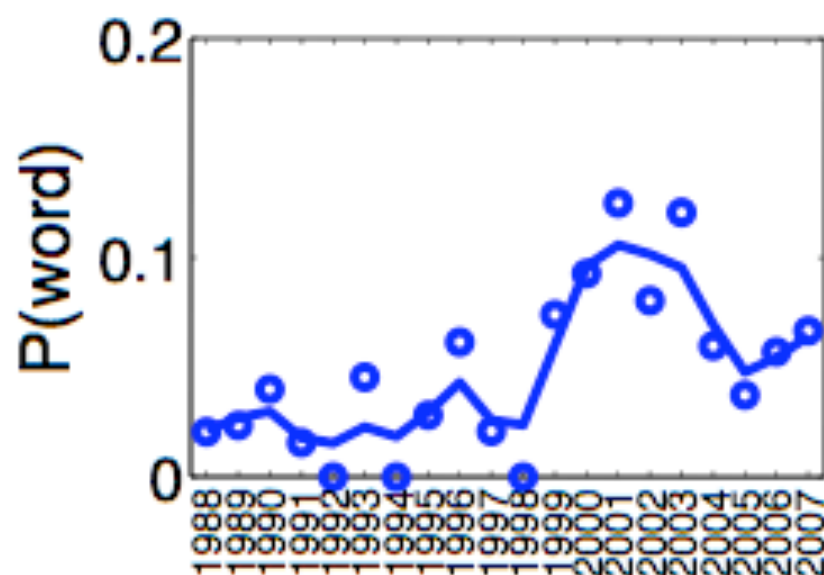
# classif



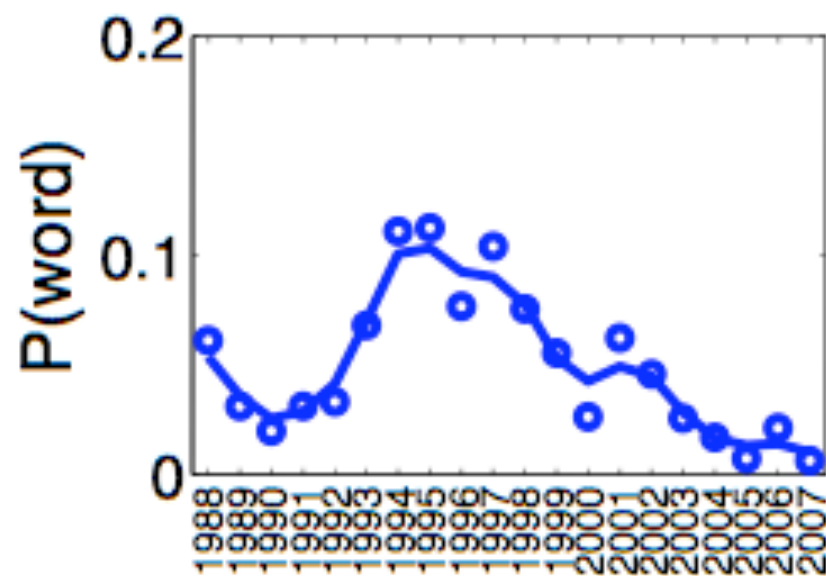
cluster



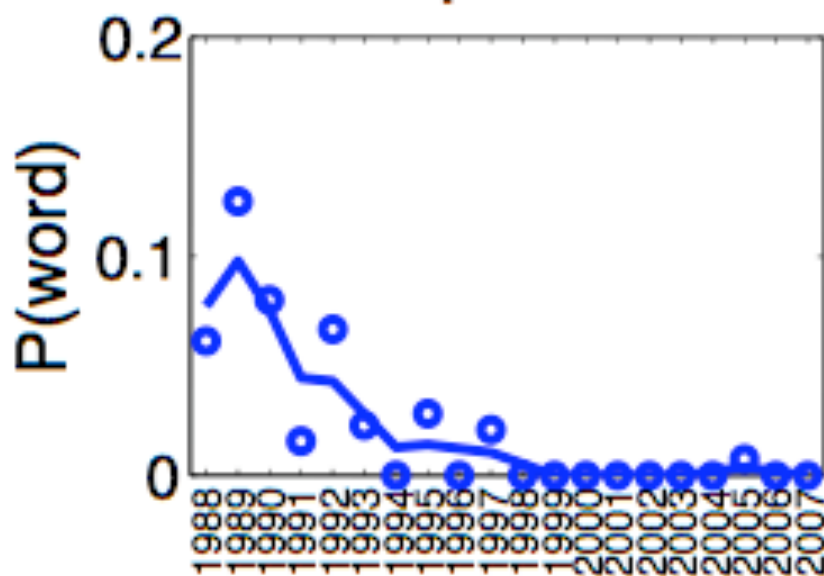
data



decis

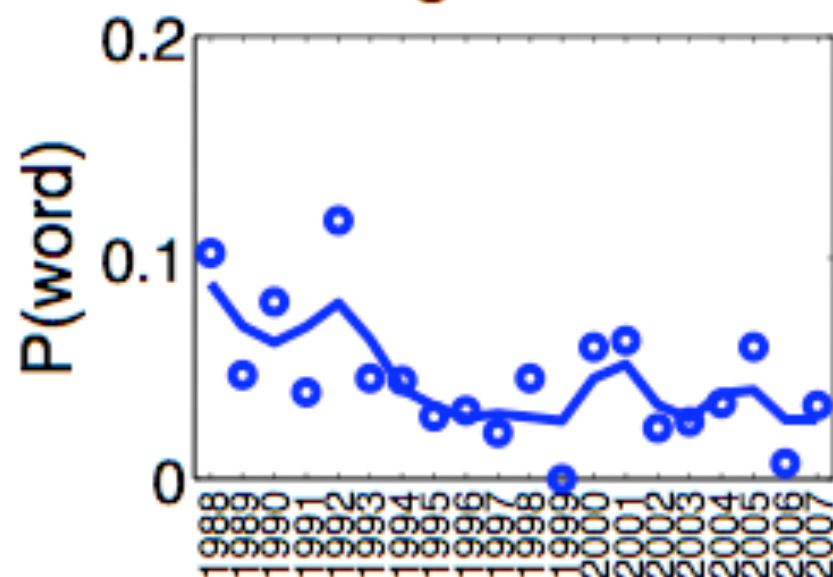


explanat

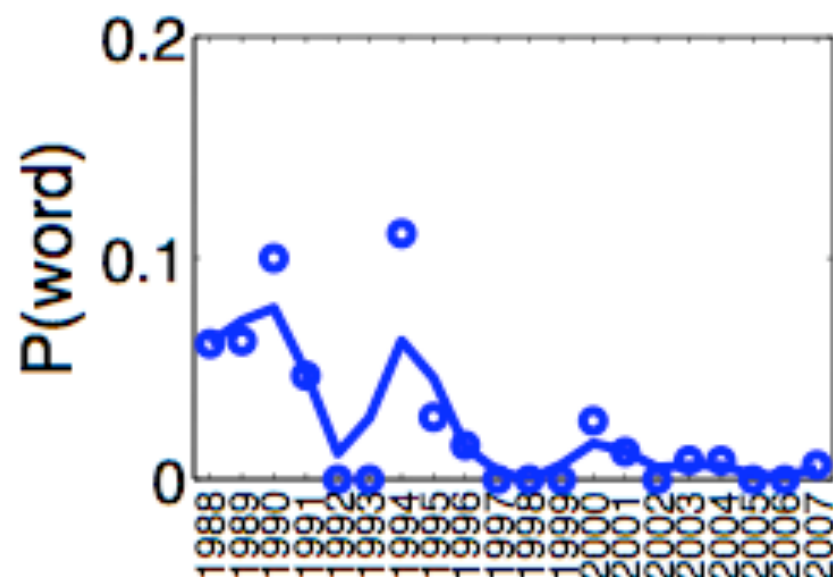




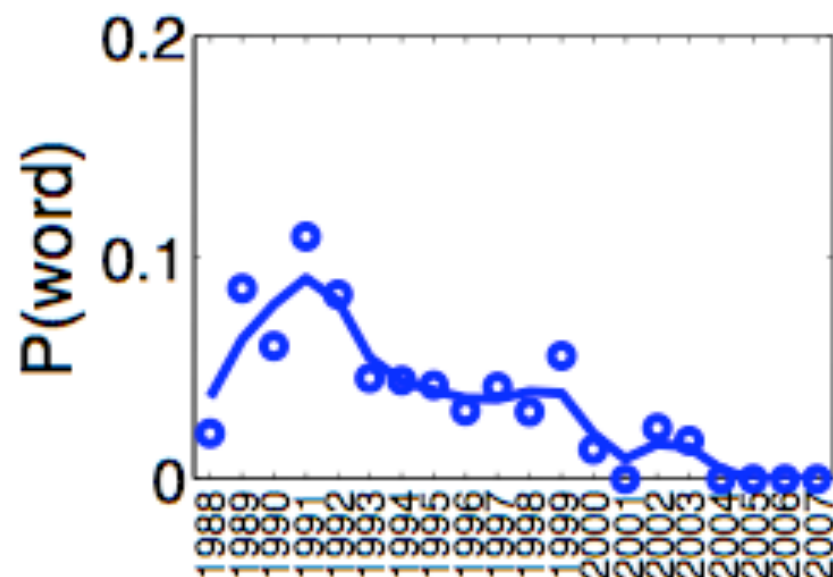
gener



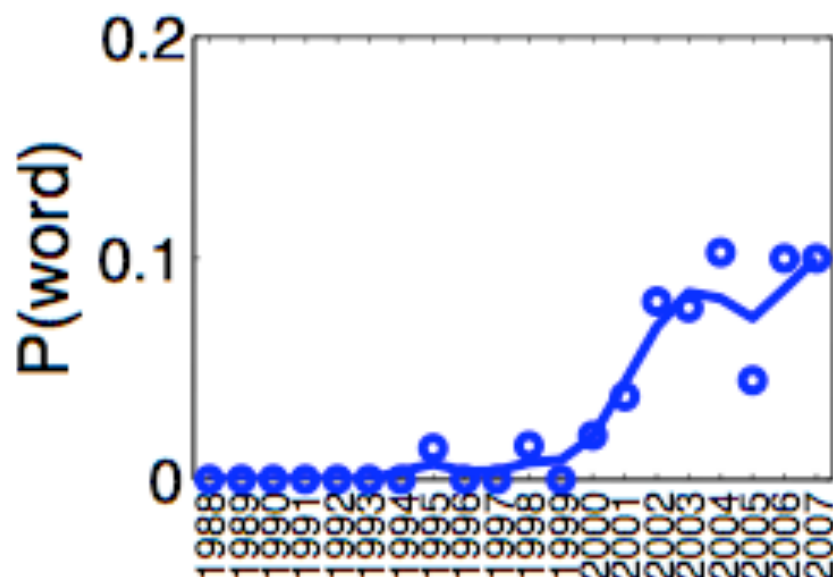
increment



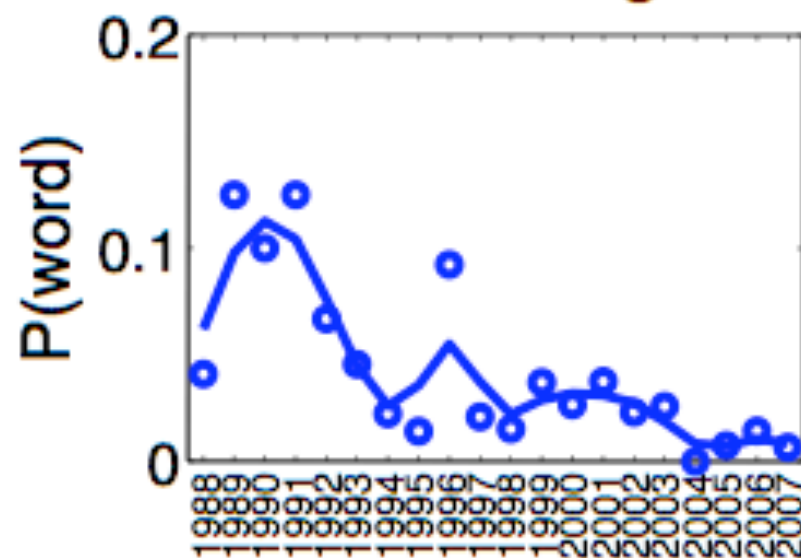
induction



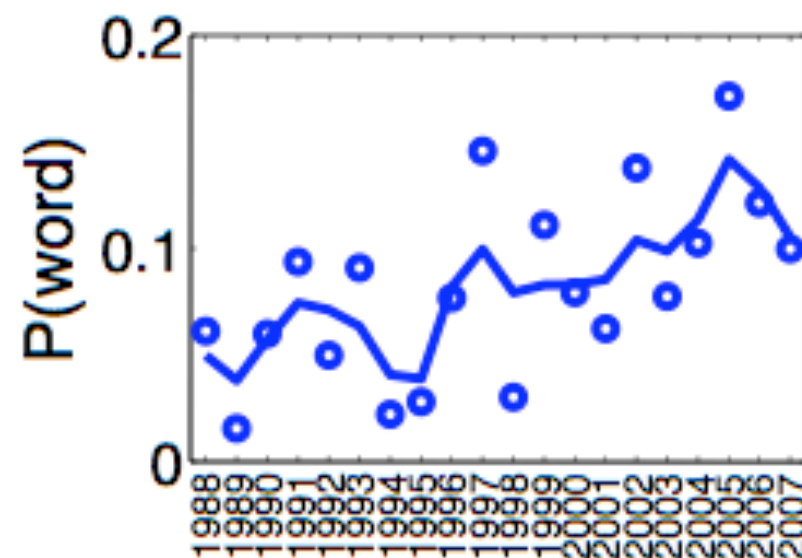
kernel



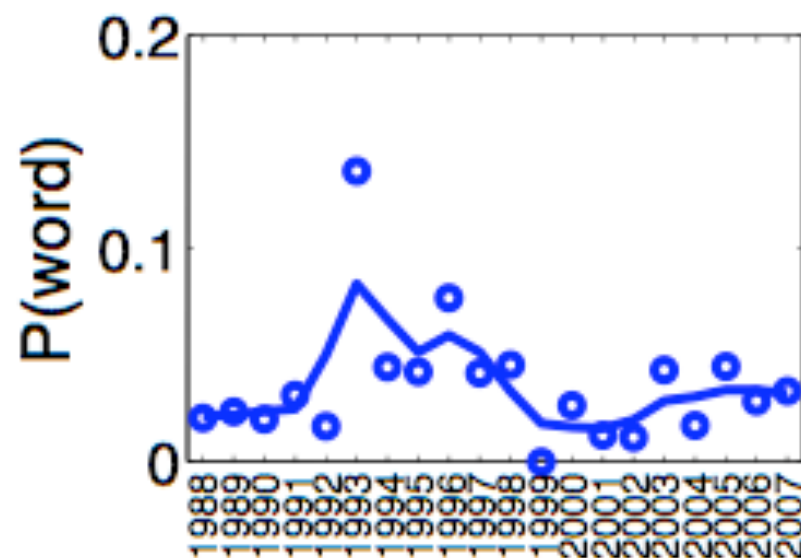
knowledg



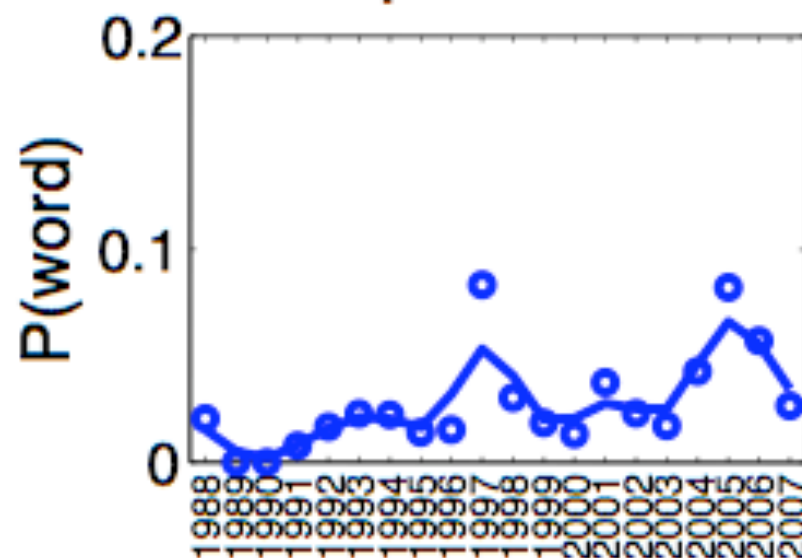
model



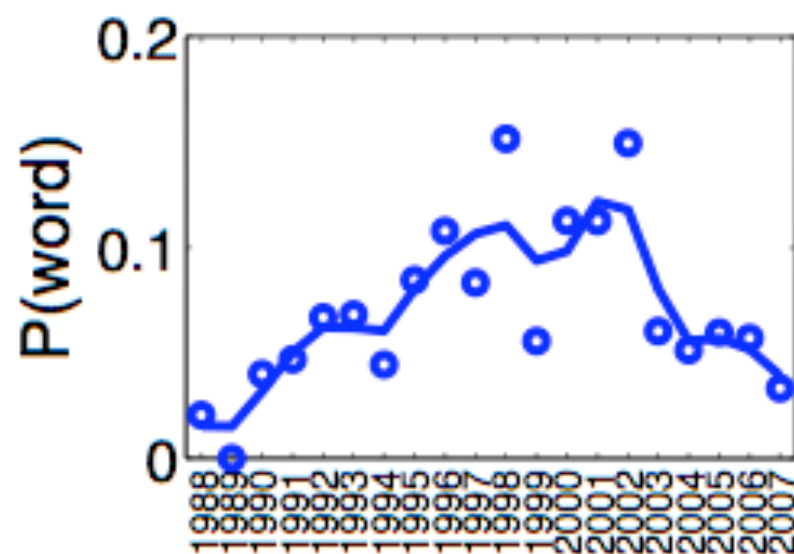
network



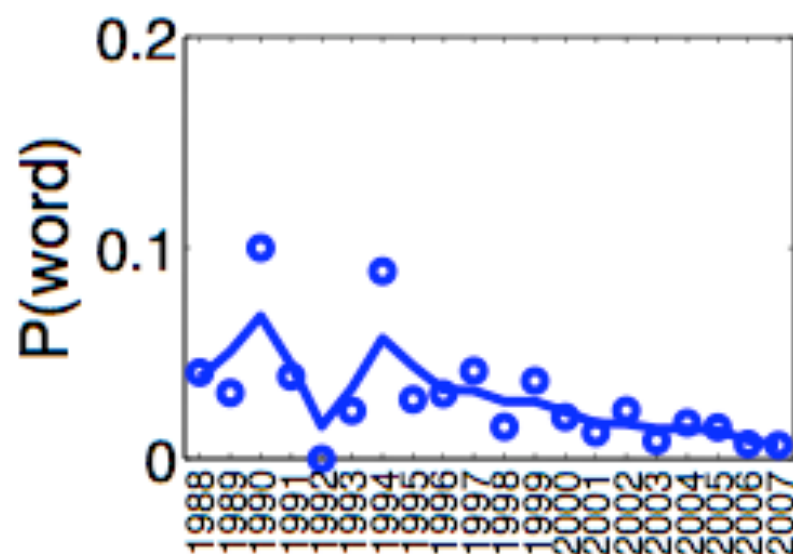
predict



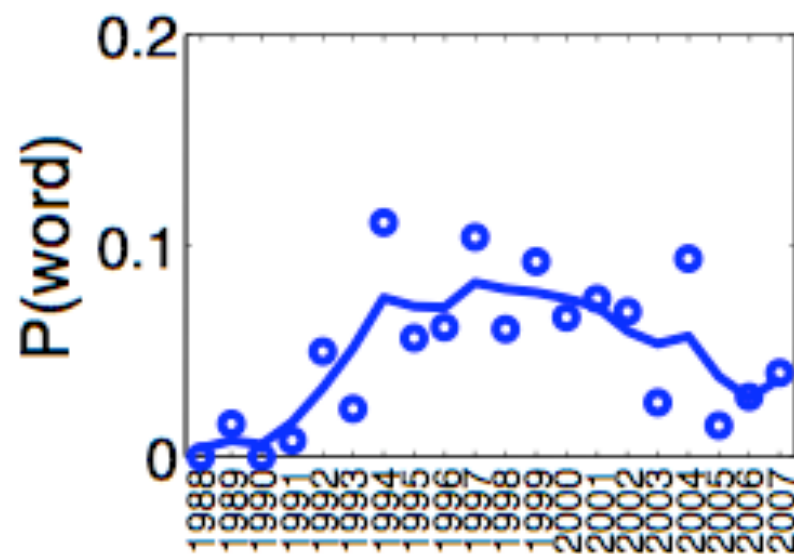
reinforc



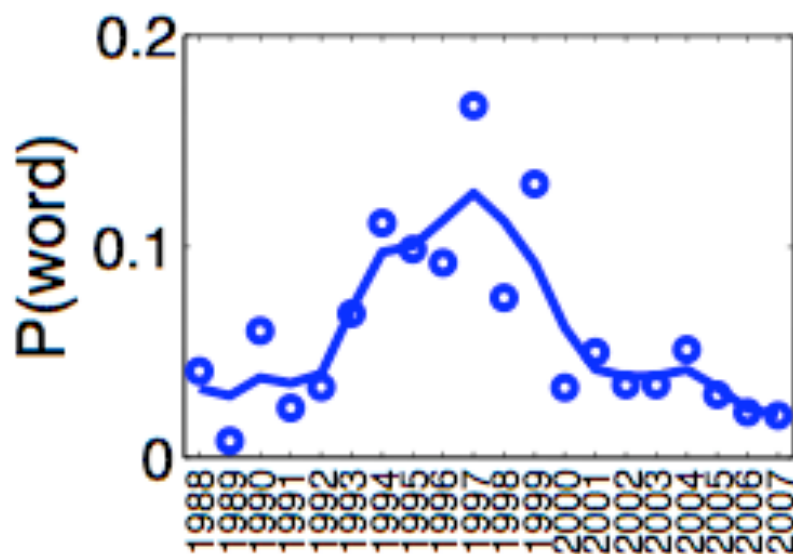
rule



select



tree



# Do High Bids Predict Accept?

- 8 papers with >30 high bids
- How many of these were accepted?
- 0!
- $p=0.064$  under null hyp of  $\text{accept}=0.29$
- Papers submitted early tend to have more bids.
- Lesson: don't make your paper popular for bidding :-)

# Reviewing Process Innovations

- Reviewer assignment process:
  - (1) automatic assignment (PCs and SPCs) based on bids.
  - (2) SPCs suggest changes to PC assignments,
  - (3) manual implementation of these suggestions + rebalancing
- Two discussion periods led by the senior program committee (SPC), one just before and one after the submission of author responses. At the end of the second discussion period, the SPC members gave their recommendations and provided a summary review for each of their papers. (296 posts in discussion 1, 1011 posts in discussion 2). SPC discussion involvement very variable.

# Reviewing Process Innovations

- Authors were asked to submit a list of changes with their final accepted papers, which was checked by the SPCs to ensure that reviewer comments had been addressed.
- Very few conditional accepts (25).

# Dual Submission Policy

- The ICML policy is that **it should be clear which parts of a submitted paper are original contributions and which parts are review of previously published or simultaneously submitted work.** If two papers contain a significant amount of material that reviewers could think is original, but that in fact also appears in a separate previous or simultaneously submitted paper, then it becomes impossible to assess the significance of the novel material in a paper. These cases will be considered dual-submissions and automatically rejected.

# Recommendations for future Program Chairs

- Coordinate with Program Chairs of simultaneous conferences to find dual submissions.
- Allow more time for paper assignment and initial review.
- Re-organize and rationalize list of topics, having more topics and keywords is useful.
- Randomize order of papers presented during bidding.



# A Possible Future Innovation in the Review Process?

- For each accepted paper, each reviewer is allowed to submit a short commentary summarizing the paper and putting it into perspective.
- These non-anonymous comments are published with the paper, along with an author reply.

# Other Innovations

- Abstract booklet
- USB drives with proceedings
- Moving towards no printed proceedings?

# Program Chair perspective

- 10,000+ emails!

# Thanks to Organizing Committee

- General Chair: Claude Sammut
- Local Arrangements Chair: Prasad Tadepalli
- Tutorials Chair: Max Welling
- Workshops Chair: Alan Fern
- Publication Chair: Ricardo Silva
- Publicity Chair: Xiaoli Fern
- Registration Chair: Soumya Ray

THANKS to SPCs, PCs, and Additional Reviewers!

# Videos of Talks are Online

<http://videolectures.net/icml07>

# Suggestions from audience..

- web site should have proceedings page numbers. Proceedings on arXiv?
- authors should be able to rate reviewers.
- randomize (or reverse) order of papers presented during bidding.
- Workshops need to be solicited, we need more workshops.