

KEVIN SMALL

Tufts Medical Center
800 Washington Street
Box 63
Boston, MA 02111
kevinssmall@gmail.com
<http://www.kevinssmall.org>

P.O. Box 120738
Boston, MA 02112
(217) 979-1992

Citizenship: USA

RESEARCH INTERESTS

Machine learning with a focus on natural language processing and medical informatics applications. More specifically, machine learning scenarios where there is interaction with a domain expert (active learning, interactive knowledge acquisition), utilization of unlabeled data, and interdependencies between learned classifiers (structured learning).

EDUCATION

Ph.D. in Computer Science, University of Illinois at Urbana-Champaign, 2009
Dissertation: *Interactive Learning Protocols for Natural Language Applications*
Committee: Dan Roth (Advisor - UIUC), Gerald DeJong (UIUC),
Julia Hockenmaier (UIUC), Andrew McCallum (UMass)

M.S. in Computer Science, University of Illinois at Urbana-Champaign, 2002
Thesis: *Passage Retrieval Using Structured Natural Language Concepts*
Advisor: Dan Roth

Graduate Studies in Electrical Engineering, Stanford University, 1998-2001
Part-time graduate coursework in signal processing while at Lucent Technologies.

B.S. in Electrical Engineering, University of Illinois at Urbana-Champaign, 1997
Minor in Computer Science

PROFESSIONAL EXPERIENCE

Special & Scientific Staff Institute for Clinical Research and Health Policy Studies, Tufts Medical Center (11/2010-Present)

Comparative effectiveness methodology research (mostly machine learning and natural language processing techniques for semi-automation of systematic reviews) with Professor Joseph Lau and Professor Thomas Trikalinos

Postdoctoral Associate Department of Computer Science, Tufts University, (9/2009-11/2010)
Machine learning research under the direction of Professor Carla Brodley and Professor Roni Khardon

Research Assistant Department of Computer Science, University of Illinois, (2004-2009)
Machine learning and natural language processing research in the Cognitive Computation Group under the direction of Professor Dan Roth

Summer Intern Motorola Labs, Schaumburg, IL (Summer 2006, Summer 2007)
Research in the natural language processing group under the direction of Paul Davis, Ph.D.

Teaching Assistant Department of Computer Science, University of Illinois

| <i>Semester</i> | <i>Course</i> | <i>Instructor</i> |
|-----------------|--|-------------------|
| Fall 2006 | Computer System Organization (<i>CS433</i>) | Josep Torrellas |
| Spring 2004 | Introduction to Artificial Intelligence (<i>CS 440</i>) | Steve Levinson |
| Fall 2003 | Introduction to Artificial Intelligence (<i>CS 440</i>) | Gerald DeJong |
| Spring 2003 | Machine Learning and Pattern Recognition (<i>CS 446</i>) | Dan Roth |
| Fall 2002 | Algorithms (<i>CS 473</i>) | Jeff Erickson |
| Summer 2001 | Introduction to Computer Science (<i>CS125</i>) | Mike Hunter |
| Spring 2001 | Introduction to Computer Science (<i>CS125</i>) | Jason Zych |

Student Mentor Data Sciences Summer Institute (Summer 2007)

The Data Sciences Summer Institute is an education program sponsored by the Multimodal Information Access and Synthesis lab. I was co-manager for a team of four graduate and three undergraduate students in the development of a named-entity recognition (NER) system.

Research Assistant Beckman Institute for Advanced Science and Research, UIUC (2001-2002)
Development of a system for conducting a virtual music ensemble under the direction of Professor Guy Garnett

Member of Technical Staff Lucent Technologies - Bell Laboratories, (1998-2001)
Member of the Broadband Switching Concept Center. Hardware circuit design and firmware development for the 7R/E Internet Telephony Switch (Digital Access Controller) and hardware circuit design for the Broadband Fiber Access Concentrator product lines.

AWARDS

- Outstanding Teaching Assistant, Department of Computer Science, UIUC, 2004.
Given to an outstanding teaching assistant each year, based both on ranking by students and recognition by faculty.
- UIUC Incomplete List of Teachers Ranked as Excellent by Their Students, UIUC, 2001.
- Induction into Eta Kappa Nu - Electrical Engineering Honor Society, 1996.

JOURNAL PUBLICATIONS

- [1] Kevin Small and Dan Roth. Margin-based active learning for structured predictions. *International Journal of Machine Learning and Cybernetics (IJMLC)*, 1(1-4):3–25, 2010.

CONFERENCE PUBLICATIONS

- [2] Kevin Small, Byron C. Wallace, Carla E. Brodley, and Thomas A. Trikalinos. The constrained weight space SVM: Learning with ranked features. In *Proc. of the International Conference on Machine Learning (ICML)*, pages 865–872, 2011.
- [3] Byron C. Wallace, Kevin Small, Carla E. Brodley, , and Thomas A. Trikalinos. Who should label what? instance allocation in multiple expert active learning. In *Proc. of the SIAM International Conference on Data Mining (SDM)*, pages 176–187, 2011.
- [4] Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau, and Thomas A. Trikalinos. Modeling annotation time to reduce workload in comparative effectiveness reviews. In *Proc. of the ACM International Health Informatics Symposium (IHI)*, pages 28–35, 2010.
- [5] Byron C. Wallace, Kevin Small, Carla E. Brodley, and Thomas A. Trikalinos. Active learning for biomedical citation screening. In *Proc. of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 173–182, 2010.
- [6] Ivan Titov, Alexandre Klementiev, Kevin Small, and Dan Roth. Unsupervised aggregation for classification problems with large numbers of categories. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 836–843, 2010.
- [7] Mark Sammons, V.G. Vinod Vydiswaran ad Tim Vieira, Nikhil Johri, Ming-Wei Chang, Dan Goldwasser, Vivek Srikumar, Gourab Kundi, Yuancheng Tu, Kevin Small, Josh Rule, Quang Do, and Dan Roth. Relation alignment for textual entailment recognition. In *Text Analysis Conference (TAC)*, 2009.
- [8] Dan Roth and Kevin Small*. Interactive feature space construction using semantic information. In *Proc. of the Conference on Computational Natural Language Learning (CoNLL)*, pages 66–74, 2009.
- [9] Alexandre Klementiev, Dan Roth, Kevin Small*, and Ivan Titov. Unsupervised rank aggregation with domain-specific expertise. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1101–1106, 2009.
- [10] Dan Roth, Kevin Small*, and Ivan Titov. Sequential learning of classifiers for structured prediction problems. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 440–447, 2009.
- [11] Alexandre Klementiev, Dan Roth, and Kevin Small*. Unsupervised rank aggregation with distance-based models. In *Proc. of the International Conference on Machine Learning (ICML)*, pages 472–479, 2008.
- [12] Dan Roth and Kevin Small*. Active learning for pipeline models. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 683–688, 2008.
- [13] Alexandre Klementiev, Dan Roth, and Kevin Small*. An unsupervised learning algorithm for rank aggregation. In *Proc. of the European Conference on Machine Learning (ECML)*, pages 616–623, 2007.

*Authors listed in alphabetical order.

- [14] Paul Davis, Kevin Small*, and Zhuli Xie. All links are not the same: Evaluating word alignments for statistical machine translation. In *Proc. of the Machine Translation Summit (MT Summit)*, pages 119–126, 2007.
- [15] Dan Roth and Kevin Small*. Margin-based active learning for structured output spaces. In *Proc. of the European Conference on Machine Learning (ECML)*, pages 413–424, 2006.
- [16] Xin Li, Dan Roth, and Kevin Small*. The role of semantic information in learning question classifiers. In *Proc. of the International Joint Conference on Natural Language Processing (IJCNLP)*, 2004.
- [17] Dan Roth, Chad Cumby, Xin Li, Paul Morie, Ramya Nagarajan, Nick Rizzolo, Kevin Small*, and Wentau Yih. Question answering via enhanced understanding of questions. In *Text Retrieval Conference (TREC)*, pages 592–601, 2002.
- [18] Guy Garnett, Mangesh Jonnalagadda, Ivan Elezovic, Timothy Johnson, and Kevin Small. Technological advances for conducting a virtual ensemble. In *International Computer Music Conference*, 2001.

WORKSHOPS AND SYMPOSIA

- [19] Kim Cuong Pham, Nicholas Rizzolo, Kevin Small, Kevin Chen-Chuan Cheng, and Dan Roth. Object search: Supporting structured queries in web search engines. In *Proc. of the NAACL Workshop on Semantic Search (SemanticSearch)*, 2010.
- [20] Byron C. Wallace, Kevin Small, Carla E. Brodley, and Thomas A. Trikalinos. Active learning for biomedical citation screening. In *Northeast Student Colloquium on Artificial Intelligence (NESCAI)*, 2010.
- [21] Kevin Small and Dan Roth. Interactive feature space construction. In *NIPS Workshop on Analysis and Design of Algorithms for Interactive Machine Learning (ADA-IML)*, 2009.
- [22] Alexandre Klementiev, Dan Roth, Kevin Small*, and Ivan Titov. Unsupervised prediction aggregation. In *NIPS Workshop on Learning with Orderings*, 2009.
- [23] Alexandre Klementiev, Dan Roth, Kevin Small*, and Ivan Titov. Unsupervised rank aggregation with domain-specific expertise. In *The Learning Workshop (Snowbird)*, 2009.
- [24] Dan Roth and Kevin Small*. Interactive introduction of semantic information for discriminative learning. In *NSF Symposium on Semantic Knowledge Discovery, Organization, and Use*, 2008.
- [25] Alexandre Klementiev, Dan Roth, and Kevin Small*. A framework for unsupervised rank aggregation. In *Proc. of the SIGIR Workshop on Learning to Rank for Information Retrieval (LR4IR)*, pages 32–39, 2008.
- [26] Dan Roth and Kevin Small*. Active learning for pipeline models. In *The Learning Workshop (Snowbird)*, 2008.

- [27] Alexandre Klementiev, Dan Roth, and Kevin Small*. Unsupervised rank aggregation with distance-based models. In *The Learning Workshop (Snowbird)*, 2008.
- [28] Dan Roth and Kevin Small*. Active learning with perceptron for structured output. In *Proc. of the ICML Workshop on Learning in Structured Output Spaces (LISOS)*, 2006.

PRESENTATIONS

- Paper Presentation, “The Constrained Weight Space SVM: Learning with Ranked Features” at the Twenty-Eighth International Conference on Machine Learning (Bellvue, WA), 2011.
- Invited Talk, “Unsupervised Rank Aggregation with Distance-based Models” at American Institute of Mathematics Workshop on The Mathematics of Ranking (Palo Alto, CA), 2010.
- Adjunct Poster Presentation, “Active Learning for Biomedical Citation Screening,” at the Sixteenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (Washington DC), 2010.
- Poster Presentation, “Interactive Feature Space Construction,” at the Workshop on Analysis and Design of Algorithms for Interactive Machine Learning at NIPS (Whistler, BC), 2009.
- Poster Presentation, “Unsupervised Prediction Aggregation,” at the Workshop on Learning with Orderings at NIPS (Whistler, BC), 2009.
- Invited Talk, “Unsupervised Aggregation of Structured Labels with Distance-based Models” – Harvard University (Cambridge, Massachusetts), October, 2009.
- Paper Presentation, “Interactive Feature Space Construction with Semantic Information,” at the Thirteenth Conference on Computational Natural Language Learning (Boulder, CO), 2009.
- Invited Talk, “Interactive Learning Protocols for Natural Language Applications”
 - Lawrence Livermore National Laboratories (Livermore, California), May, 2009.
 - Tufts University (Boston, Massachusetts), April, 2009.
 - University of Amsterdam (Amsterdam, Netherlands), March, 2009.
 - Alberta Ingenuity Centre for Machine Learning (Edmonton, Alberta), March, 2009.
- Oral Presentation, “Interactive Introduction of Semantic Information for Discriminative Learning,” at the NSF Symposium on Semantic Knowledge Discovery, Organization and Use (New York City, NY), 2008.
- Paper Presentation, “Active Learning for Pipeline Models,” at the Twenty-Third Conference on Artificial Intelligence (Chicago, IL), 2008.
- Adjunct Poster Presentation, “Unsupervised Rank Aggregation with Distance-based Models,” at the Twenty-Fifth International Conference on Machine Learning (Helsinki, Finland), 2008.
- Oral Presentation, “Active Learning for Pipeline Models,” at Snowbird Learning Workshop (Snowbird, UT), 2008.

- Poster Presentation, “Unsupervised Rank Aggregation with Distance-based Models,” at Snowbird Learning Workshop (Snowbird, UT), 2008.
- Poster presentation, “An Unsupervised Learning Algorithm for Rank Aggregation,” at the Eighteenth European Conference on Machine Learning (Warsaw, Poland), 2007.
- Paper presentation, “All Links are not the Same: Evaluating Word Alignments for Statistical Machine Translation,” at Machine Translation Summit XI (Copenhagen, Denmark), 2007
- Paper presentation, “Margin-based Active Learning for Structured Output Spaces,” at the Seventeenth European Conference on Machine Learning (Berlin, Germany), 2006.
- Paper presentation, “Active Learning with Perceptron for Structured Output,” at the International Machine Learning Conference Workshop on Learning in Structured Output Spaces (Pittsburgh, USA), 2006.

PROFESSIONAL SERVICE

- Senior Program Committee: National Conference on Artificial Intelligence (AAAI) 2011
- Program Committee: National Conference on Artificial Intelligence (AAAI) 2010; Annual Meeting of the Association of Computational Linguistics (ACL) 2007; International Conference on Computational Linguistics (COLING) 2008; Conference on Empirical Methods on Natural Language Processing (EMNLP) 2011; International Conference on Machine Learning (ICML) 2009, 2010; International Joint Conference on Artificial Intelligence (IJCAI) 2011; Annual ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2011; Meeting of the North American Association of Computational Linguistics (NAACL) 2009, 2010; NAACL Workshop on Active Learning for Natural Language Processing (ALNLP) 2010; NIPS Workshop on Advances in Ranking 2009
- Workshop Organizer: NAACL Workshop on Active Learning for Natural Language Processing (ALNLP) 2010
- Journal Reviewer: ACM Transactions on Speech and Language Processing; International Journal of Computer Vision; Journal of Artificial Intelligence Research; Machine Learning Journal
- External Reviewer: National Conference on Artificial Intelligence (AAAI), Annual Meeting of the Association of Computational Linguistics (ACL), International Conference on Computational Linguistics (COLING), Conference on Learning Theory (COLT), Conference on Computational Natural Language Learning (CoNLL), Meeting of the European Association of Computational Linguistics (EACL), Conference on Empirical Methods on Natural Language Processing (EMNLP), International Conference on Machine Learning (ICML), Meeting of the North American Association of Computational Linguistics (NAACL), Neural Information Processing Systems Conference (NIPS), International Conference and Exhibition on Computer Graphics and Interactive Technologies (SIGGRAPH)

REFERENCES

Available upon request.