Kevin Small

National Institutes of Health
9000 Rockville Pike
Building 2, Room 1E16G
Bethesda, MD 20892
kevinsmall@gmail.com
http://www.kevinsmall.org

17914 Mill Branch Place
Bowie, MD 20716
(217) 979-1992
Citizenship: USA

RESEARCH INTERESTS

Machine learning/data mining with a focus on scenarios where there is interaction with a domain expert during training (active learning, interactive knowledge acquisition), interdependencies between learned classifiers (structured learning), and efficient use of unlabeled data. I am particularly interested in natural language processing, information extraction, health informatics, and scientific evidence aggregation/analysis applications.

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 Clinical Translational Sciences, Tufts University, 2011-2012
    Graduate coursework in epidemiology and decision analysis as part of my KM1 award.
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

RESEARCH EXPERIENCE

National Institutes of Health (7/2012-Present)
Designing data-driven methodologies and software tools for informing science policy decisions. While employed by the Office of the Director (OD), I also collaborate with the Center for Information Technology (CIT), and National Library of Medicine (NLM). Responsible for reviewing common fund proposals in the areas of data-driven methods and citizen science.

Tufts Medical Center, Institute for Clinical Research and Health Policy Studies
Special & Scientific Staff (Research Faculty) (11/2010-7/2012)
Comparative effectiveness methodology research (mostly machine learning and natural language processing techniques for semi-automation of systematic reviews) in collaboration with Professors Thomas Trikalinos and Joseph Lau

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

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

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

University of Illinois, Beckman Institute for Advanced Science and Research
Research Assistant (2001-2002)
Development of a system for conducting a virtual music ensemble with Professor Guy Garnett

TEACHING EXPERIENCE

Boston University, Department of Computer Science
Visiting Lecturer (Spring 2012) – *Machine Learning (CS 542)*
Introduction to modern machine learning concepts, techniques, and algorithms. Topics include regression, kernels, support vector machines, feature selection, boosting, clustering, hidden Markov models, and Bayesian networks. Programming assignments emphasize taking theory into practice, through applications on real-world data sets.

Tufts University, Department of Computer Science
Visiting Lecturer (Fall 2011) – *Artificial Intelligence in Health Informatics (COMP 150AIH)*
Proposed, developed, and taught a course *de novo* (in partnership with Byron Wallace) on artificial intelligence and machine learning techniques as applied to health and medical informatics.

Guest Lecturer, Various
Provided a lecture on *Active Learning* for the Machine Learning course at Tufts (Fall 2010, Fall 2011) and *Interactive Learning in NLP* for the Machine Learning in NLP at UIUC (Spring 2009).

University of Illinois, Department of Computer Science
Teaching Assistant
Awarded *Outstanding Teaching Assistant* designation from the department of computer science and ranked as *excellent* for every eligible semester (those where I had lecturing responsibilities)

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

University of Illinois, Data Sciences Summer Institute
Student Mentor (Summer 2007)
The Data Sciences Summer Institute is an education program sponsored by the Multimodal Information Access and Synthesis (MIAS) lab. I was the team leader for a team of four graduate and three undergraduate students in the development of a named-entity recognition (NER) system.

PROFESSIONAL EXPERIENCE

Lucent Technologies - Bell Laboratories
Technical Staff, (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

- Career Development Award (KM1) in Comparative Effectiveness Research, Tufts Clinical and Translational Science Institute, 2011.
  NIH funded career development award for junior faculty.

- 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 ARTICLES & REFEREED REPORTS

[1] Carla E. Brodley, Umaa Rebbapragada, Kevin Small*, and Byron C. Wallace. Challenges and opportunities in applied machine learning. *Artificial Intelligence Magazine*, 33(1):11–24, 2012.

[2] Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau, Christopher H. Schmid, Lars Bertram, Christina M. Lill, Joshua T. Cohen, and Thomas A. Trikalinos. Towards modernizing the systematic review pipeline in genetics: Updating via data mining. *Genetics in Medicine*, 14:663–669, 2012.

[3] Thomas A. Trikalinos, David C. Hoaglin, Kevin M. Small, and Christopher H. Schmid. Evaluating practices ad developing tools for comparative effectiveness reviews of diagnostic test accuracy, Task 3: Methods for the joint meta-analysis of multiple tests. *Agency for Healthcare Research and Quality (AHRQ) - Technical Report*[1], 2012.

[4] 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.

---

*Authors listed in alphabetical order.

[1]AHRQ reports are rigorously peer-reviewed (typically 5-10 reviewers), subjected to open public review, and carry substantial weight. They are often commissioned to inform national committees and healthcare policy decisions.

CONFERENCE PUBLICATIONS[2]

[5] Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau, and Thomas A. Trikalinos. Deploying an interactive machine learning system in an evidence-based practice center. In *Proc. of the ACM International Health Informatics Symposium (IHI)*, pages 819–824, 2012.

[6] Byron C. Wallace, Kevin Small, Carla E. Brodley, and Thomas A. Trikalinos. Class imbalance, redux. In *Proc. of the International Conference on Data Mining (ICDM)*, pages 754–763, 2011.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Mark Sammons, V.G. Vinod Vydiswaran ad Tim Vieira, Nikhil Johri, Ming-Wei Chang, Dan Goldwasser, Vivek Srikumar, Gourab Kundu, Yuancheng Tu, Kevin Small, Josh Rule, Quang Do, and Dan Roth. Relation alignment for textual entailment recognition. In *Text Analysis Conference (TAC)*, 2009.

[13] 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.

[14] Alexandre Klementiev, Dan Roth, Kevin Small*, and Ivan Titov. Unsupervised rank aggregation with domain-specic expertise. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1101–1106, 2009.

[15] 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.

---

[2]For my biomedical research friends, please be aware that conference publications are the primary vehicle for disseminating research results to the computer science community, represent the publication of record, and are generally considered more prestigious than journal articles.

[16] 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.

[17] Dan Roth and Kevin Small*. Active learning for pipeline models. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 683–688, 2008.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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

[24] 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.

[25] 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.

[26] 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.

[27] Alexandre Klementiev, Dan Roth, Kevin Small*, and Ivan Titov. Unsupervised prediction aggregation. In *NIPS Workshop on Learning with Orderings*, 2009.

[28] Alexandre Klementiev, Dan Roth, Kevin Small*, and Ivan Titov. Unsupervised rank aggregation with domain-specific expertise. In *The Learning Workshop (Snowbird)*, 2009.

[29] Dan Roth and Kevin Small*. Interactive introduction of semantic information for discriminative learning. In *NSF Symposium on Semantic Knowledge Discovery, Organization, and Use*, 2008.

[30] 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.

[31] Dan Roth and Kevin Small*. Active learning for pipeline models. In *The Learning Workshop (Snowbird)*, 2008.

[32] Alexandre Klementiev, Dan Roth, and Kevin Small*. Unsupervised rank aggregation with distance-based models. In *The Learning Workshop (Snowbird)*, 2008.

[33] 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.

MISCELLANY

[34] Lisa R. McCluskey, Alexandra Tellier, Elizabeth L. Fay, Jeffrey J. Klunk, and Kevin Small. Using a canine companion training curriculum to reduce return rates in a no-kill animal shelter environment *(extended abstract and oral presentation)*. In *The Annual Conference of the Association of Pet Dog Trainers (APDT)*, 2012.

[35] Lisa R. McCluskey and Kevin Small. An evidence-based canine companion training class curriculum assessment methodology *(extended abstract and poster presentation)*. In *The Annual Conference of the Association of Pet Dog Trainers (APDT)*, 2011.

[36] Lisa R. McCluskey, Sarah A. Larson, Emilie C. Bess, and Kevin Small. Utilizing the companion sniffer dog training method to develop aptitude in locating target scents *(extended abstract and poster presentation)*. In *The Annual Conference of the Association of Pet Dog Trainers (APDT)*, 2010.

PRESENTATIONS

- Poster Presentation, "Semi-automated Data Extraction from Biomedical Literature for Systematic Reviews" at Tufts CTSI Graduate Program Symposium, May, 2012.

- Invited Talk, "Interactive Machine Learning: Methods and Applications"
  – National Institutes of Health (Bethesda, MD), April, 2012.

- Poster Presentation, "An Interactive Machine Learning System for Biomedical Citation Screening" at NIH Workshop on Natural Language Processing: State of the Art, Future Directions and Applications for Enhancing Clinical Decision-Making (Bethesda, MD), April, 2012.

- Invited Talk, "Interactive Machine Learning for Biomedical Abstract Screening"
  – Boston University (Boston, MA), September, 2011.

- 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.

FUNDING

Grant Title: Tufts CTSI Career Development Program in Comparative Effectiveness Research
Funding Agency: KM1 CA156726 NIH/NCI
Amount: $232,056 (updated to $128,550 as I left early to work at NIH)
Period: 7/1/2011 to 6/30/2013 (updated to 8/3/2012)
Role: Scholar

PROFESSIONAL SERVICE

- Senior Program Committee: National Conference on Artificial Intelligence (AAAI) 2011

- Program Committee: National Conference on Artificial Intelligence (AAAI) 2010, 2012, 2013; Annual Meeting of the Association of Computational Linguistics (ACL) 2007, 2013; International Conference on Information and Knowledge Management (CIKM) 2012; International Conference on Computational Linguistics (COLING) 2008; European Conference on Machine Learning (ECML) 2012; Conference on Empirical Methods on Natural Language Processing (EMNLP) 2011; International Conference on Machine Learning (ICML) 2009, 2010; ICDM International Workshop on Cost Sensitive Data Mining (COSTS) 2012; 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, 2012; 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; BMC Bioinformatics; 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.