

Xiaojin Zhu

Contact

Department of Computer Sciences
University of Wisconsin, Madison
1210 West Dayton Street
Madison, WI 53706

Phone: 608-890-0129
Fax: 608-262-9777

jerryzhu@cs.wisc.edu

Research Interests

Statistical machine learning, natural language processing

Education

Ph.D. Language Technologies Carnegie Mellon University, Pittsburgh, PA Dissertation: Semi-Supervised Learning with Graphs Advisors: John Lafferty, Ronald Rosenfeld	May, 2005
M.Sc. Knowledge Discovery and Data Mining Carnegie Mellon University, Pittsburgh, PA	December, 2002
M.Sc. Language and Information Technologies Carnegie Mellon University, Pittsburgh, PA	May, 2000
M.Sc., Computer Science Shanghai Jiao Tong University, Shanghai, China	March, 1996
B.Sc., Computer Science Shanghai Jiao Tong University, Shanghai, China	July, 1993

Professional Positions

Assistant Professor Department of Computer Sciences, University of Wisconsin, Madison Madison, WI	2005–present
Research Scientist IBM China Research Laboratory Beijing, China	1996–1998

Publications

Co-Authors noted as (s) for student under my direction, (p) for post-doctoral associate under my direction, (o) for students or post-docs under the direction of others, and (a) for my thesis Advisors.

Refereed Conference Papers

Andrew B. Goldberg^(s), Ming Li^(s), and Xiaojin Zhu. Online manifold regularization: A new learning setting and empirical study. In *The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2008. (acceptance rate 98/521=18.8%).

Andrew B. Goldberg^(s), Xiaojin Zhu, Charles R. Dyer, Mohamed Eldawy^(o), and Lijie Heng^(s). Easy as ABC? Facilitating pictorial communication via semantically enhanced layout. In *Twelfth Conference on Computational Natural Language Learning (CoNLL)*, 2008. (acceptance rate 20/85=23.5%).

Xiaojin Zhu, Michael Coen, Shelley Prudom, Ricki Colman, and Joseph Kemnitz. Online learning in monkeys. In *Twenty-Third AAAI Conference on Artificial Intelligence (AAAI-08)*, 2008. (short paper, overall acceptance rate (23+227)/958=26%).

Nathan Rosenblum^(o), Xiaojin Zhu, Barton Miller, and Karen Hunt. Learning to analyze binary computer code. In *Twenty-Third AAAI Conference on Artificial Intelligence (AAAI-08)*, 2008. (full paper, acceptance rate 227/937=24%; also selected for additional poster highlight, 5%).

Xiaojin Zhu, Andrew B. Goldberg^(s), Michael Rabbat, and Robert Nowak. Learning bigrams from unigrams. In *The 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL)*, 2008. (acceptance rate 25%).

Pedro DeRose^(o), Xiaoyong Chai^(o), Byron Gao^(o), Warren Shen^(o), AnHai Doan, Philip Bohannon^(o), and Xiaojin Zhu. Building community Wikipedias: A machine-human partnership approach. In *IEEE International Conference on Data Engineering (ICDE)*, 2008. (acceptance rate 12.1%).

David Andrzejewski^(s), Anne Mulhern^(o), Ben Liblit, and Xiaojin Zhu. Statistical debugging using latent topic models. In *Proceedings of the 18th European Conference on Machine Learning (ECML)*, 2007. (acceptance rate 20%).

Gregory Druck^(o), Chris Pal^(o), Xiaojin Zhu, and Andrew McCallum. Semi-supervised classification with hybrid generative/discriminative methods. In *The Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2007. (acceptance rate 20%).

Jordan Boyd-Graber^(o), David Blei, and Xiaojin Zhu. A topic model for word sense disambiguation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP-CoNLL)*, 2007. (acceptance rate 27%).

Xiaojin Zhu, Timothy Rogers, Ruichen Qian^(s), and Chuck Kalish. Humans perform semi-supervised classification too. In *Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07)*, 2007. (full paper, acceptance rate 27%; also selected for additional poster highlight, 5%).

Xiaojin Zhu, Andrew Goldberg^(s), Mohamed Eldawy^(o), Charles Dyer, and Bradley Strock^(s). A Text-to-Picture synthesis system for augmenting communication. In *Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07)*, pages 1590–1595, 2007. (acceptance rate 27%).

Xiaojin Zhu and Andrew Goldberg^(s). Kernel regression with order preferences. In *Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07)*, 2007. (acceptance rate 27%).

Mariyam Mirza^(o), Joel Sommers^(o), Paul Barford, and Xiaojin Zhu. A machine learning approach to TCP throughput prediction. In *The International Conference on Measurement and Modeling of Computer Systems (ACM SIGMETRICS)*, 2007. (acceptance rate 17%).

Xiaojin Zhu, Andrew Goldberg^(s), Jurgen Van Gael^(s), and David Andrzejewski^(s). Improving diversity in ranking using absorbing random walks. In *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, 2007. (acceptance rate 24%).

Andrew Goldberg^(s), Xiaojin Zhu, and Stephen Wright. Dissimilarity in graph-based semi-supervised classification. In *Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2007. (acceptance rate 50%).

Jurgen Van Gael^(s) and Xiaojin Zhu. Correlation clustering for crosslingual link detection. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2007. (acceptance rate 34%).

Xiaojin Zhu and John Lafferty^(a). Harmonic mixtures: combining mixture models and graph-based methods for inductive and scalable semi-supervised learning. In *The 22nd International Conference on Machine Learning (ICML)*. ACM Press, 2005. (acceptance rate 27%).

Xiaojin Zhu, Jaz Kandola, Zoubin Ghahramani, and John Lafferty^(a). Nonparametric transforms of graph kernels for semi-supervised learning. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems (NIPS) 17*. MIT Press, Cambridge, MA, 2005. (acceptance rate 25%).

John Lafferty^(a), Xiaojin Zhu, and Yan Liu. Kernel conditional random fields: Representation and clique selection. In *The 21st International Conference on Machine Learning (ICML)*, 2004. (acceptance rate 32%).

Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty^(a). Semi-supervised learning using Gaussian fields and harmonic functions. In *The 20th International Conference on Machine Learning (ICML)*, 2003. (acceptance rate 32%).

Stefanie Shriver, Arthur Toth, Xiaojin Zhu, Alex Rudnicky, and Roni Rosenfeld^(a). A unified design for human-machine voice interaction. In *Human Factors in Computing Systems (CHI)*. ACM Press, 2001. (acceptance rate 20%).

Xiaojin Zhu and Ronald Rosenfeld^(a). Improving trigram language modeling with the World Wide Web. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2001. (acceptance rate 51%).

Ronald Rosenfeld^(a), Xiaojin Zhu, Stefanie Shriver, Arthur Toth, Kevin Lenzo, and Alan Black. Towards a universal speech interface. In *International Conference on Spoken Language Processing (ICSLP)*, 2000.

Xiaojin Zhu, Jie Yang, and Alex Waibel. Segmenting hands of arbitrary color. In *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000. (acceptance rate 52%).

Jie Yang, Xiaojin Zhu, Ralph Gross, John Kominek, Yue Pan, and Alex Waibel. Multimodal people ID for multimedia meeting browser. In *The Seventh ACM International Multimedia Conference*, 1999. (acceptance rate 21%).

Xiaojin Zhu, Stanley F. Chen, and Ronald Rosenfeld^(a). Linguistic features for whole sentence maximum entropy language models. In *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech)*, 1999.

Journal Papers

Ronald Rosenfeld^(a), Stanley Chen, and Xiaojin Zhu. Whole-sentence exponential language models: a vehicle for linguistic-statistical integration. *Computers Speech and Language*, 15(1), 2001.

Book and Book Chapters

Xiaojin Zhu, Jaz Kandola, John Lafferty^(a), and Zoubin Ghahramani. Graph kernels by spectral transforms. In O. Chapelle, B. Schölkopf, and A. Zien, editors, *Semi-Supervised Learning*. MIT Press, 2006.

Refereed Workshop Papers

Nathan Rosenblum^(o), Xiaojin Zhu, Barton Miller, and Karen Hunt. Machine learning-assisted binary code analysis. In *NIPS 2007 workshop on Machine Learning in Adversarial Environments for Computer Security*, 2007.

SaiSuresh Krishnakumaran^(s) and Xiaojin Zhu. Hunting elusive metaphors using lexical resources. In *NAACL 2007 Workshop on Computational Approaches to Figurative Language*, 2007.

Andrew Goldberg^(s), Dave Andrzejewski^(s), Jurgen Van Gael^(s), Burr Settles^(o), Xiaojin Zhu, and Mark Craven. Ranking biomedical passages for relevance and diversity: University of Wisconsin, Madison at TREC genomics 2006. In *Proceedings of the Fifteenth Text Retrieval Conference (TREC)*, 2006.

Andrew Goldberg^(s) and Xiaojin Zhu. Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. In *HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing*, New York, NY, 2006.

Maria-Florina Balcan, Avrim Blum, Patrick Pakyang Choi, John Lafferty^(a), Brian Pantano, Mugizi Robert Rwebangira, and Xiaojin Zhu. Person identification in webcam images: An application of semi-supervised learning. In *ICML 2005 Workshop on Learning with Partially Classified Training Data*, 2005.

Xiaojin Zhu, John Lafferty^(a), and Zoubin Ghahramani. Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions. In *ICML 2003 workshop on The Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, 2003.

Technical Reports

Xiaojin Zhu and Andrew Goldberg^(s). Semi-supervised regression with order preferences. Technical Report 1578, Department of Computer Sciences, University of Wisconsin-Madison, 2006.

Xiaojin Zhu, David Blei, and John Lafferty^(a). TagLDA: Bringing document structure knowledge into topic models. Technical Report 1553, Department of Computer Sciences, University of Wisconsin-Madison, 2006.

Xiaojin Zhu. Semi-supervised learning literature survey. Technical Report 1530, Department of Computer Sciences, University of Wisconsin, Madison, 2005.

Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty^(a). Time-sensitive Dirichlet process mixture models. Technical Report CMU-CALD-05-104, Carnegie Mellon University, 2005.

Xiaojin Zhu. *Semi-Supervised Learning with Graphs*. PhD thesis, Carnegie Mellon University, 2005. CMU-LTI-05-192.

Xiaojin Zhu, John Lafferty^(a), and Zoubin Ghahramani. Semi-supervised learning: From Gaussian fields to Gaussian processes. Technical Report CMU-CS-03-175, Carnegie Mellon University, 2003.

Xiaojin Zhu and Zoubin Ghahramani. Learning from labeled and unlabeled data with label propagation. Technical Report CMU-CALD-02-107, Carnegie Mellon University, 2002.

Xiaojin Zhu and Zoubin Ghahramani. Towards semi-supervised classification with Markov random fields. Technical Report CMU-CALD-02-106, Carnegie Mellon University, 2002.

Xiaojin Zhu and Ronald Rosenfeld^(a). Improving trigram language modeling with the World Wide Web. Technical Report CMU-CALD-00-171, Carnegie Mellon University, 2000.

Teaching Experience

Computer Sciences Department, University of Wisconsin, Madison, WI 2005–present
Assistant Professor
CS 769 – Advanced Natural Language Processing
CS 838 – Topics in Advanced Natural Language Processing
CS 540 – Introduction to Artificial Intelligence

Center of Automatic Learning and Discovery, Carnegie Mellon University June 16, 2004

Instructor

Learning from Labeled and Unlabeled Data, CALD Summer School.

School of Computer Science, Carnegie Mellon University, Pittsburgh, PA

Fall 2000

Teaching Assistant

15-681 – Machine Learning

Talks and Media Coverage

Tutorial (with J. Blitzer), to be presented at the 46th Association for Computational Linguistics meeting (ACL), Columbus, OH, 2008.

Invited talk, The 40th Interface Symposium (annual conference on the interface of computing science and statistics), Durham, NC, 2008.

Presentation and demo, University of Wisconsin Cognitive Science Conference Hertz Foundation poster session, Madison, WI, 2008.

Invited talk, Computer Science and Engineering Department Colloquium, Michigan State University, Lansing, MI, 2007.

Invited talk, Department of Statistics, University of Michigan, Ann Arbor, MI, 2007.

“Picture This”, NewScientist, August 18-24, p.22, 2007.

Invited tutorial, International Conference on Machine Learning (ICML), Corvallis, OR, 2007.

Invited alumnus talk, Language Technology Institute Retreat, Carnegie Mellon University, Pittsburgh, PA, 2007.

Invited talk, Psychology Department, University of Wisconsin, Madison, WI, 2007.

Invited participant, BIRS workshop of mathematical programming in machine learning and data mining, Banff, Canada, 2007.

Invited talk, Joint Statistical Meetings (JSM), Seattle, WA, 2006.

Invited talk, AERFAI Summer School on Action and Object Classification Techniques in Digital Images, University of Granada, Spain, 2006.

Invited talk, Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI, 2006.

Invited talk, Computer Science and Engineering Department, Washington University in St. Louis, MO, 2006.

Invited talk, Statistics Department, University of Wisconsin, Madison, WI, 2006.

Invited talk, University of Cambridge, UK, 2004.

Invited talk, Gatsby Computational Neuroscience Unit, University College London, UK, 2004.

Invited talk, Microsoft Research Cambridge, UK, 2004.

Invited talk, NSF Aladdin Workshop on Graph Partitioning in Vision and Machine Learning, Pittsburgh, PA, 2003.

Grants

“Learning: Between Humans and Machines”, University of Wisconsin Cognitive Science Cluster Research Fellowship. R. Qian, X. Zhu (Project Sponsor). \$3K, 2008.

“Semi-Supervised Learning in Humans and Machines”, University of Wisconsin Graduate School Research Award. X. Zhu PI, T. Rogers Co-PI. \$35K, 2008-2009.

“RI: Text-to-Picture Synthesis”, National Science Foundation IIS-0711887. X. Zhu PI, C. Dyer Co-PI. \$400K, 2007-2010.

“Application of artificial intelligence and human computing methods to panoramic astrophysical surveys”, University of Wisconsin Graduate School Research Award. X. Zhu PI, E. Churchwell, Co-PI. \$31K, 2007-2008.

“Extracting background knowledge from the scientific literature to improve the accuracy of gene regulatory network inference”, University of Wisconsin Graduate School Research Award. M. Craven PI, X. Zhu Co-PI. \$30K, 2006-2007.

Awards and Honors

Microsoft Research Graduate Fellowship, 2000.

Research Division Award, IBM, 1998.

First Patent Application Invention Achievement Award, IBM, 1997.

Greater China Group Team Award, IBM, 1997.

Professional Activities

Program Committee, Conference on Empirical Methods in Natural Language Processing (EMNLP-CoNLL), 2008.

Program Committee, European Conference on Computer Vision (ECCV), 2008.

Program Committee, The Fifth Midwest Computational Linguistics Colloquium (MCLC-5), 2008.

Program Committee, The 6th International Workshop on Mining and Learning with Graphs (MLG), 2008.

NSF review panel, CISE/IIS, 2008.

Program Committee, International Conference on Machine Learning (ICML), 2008.

Program Committee, AAAI Conference on Artificial Intelligence (AAAI), 2008.

Program Committee, Annual Meeting of the Association for Computational Linguistics (ACL), 2008.

Program Committee, The Pacific Rim International Conference on Artificial Intelligence (PRICAI), 2008.

Senior Program Committee, International Conference on Machine Learning (ICML), 2007.

Program Committee, Conference on Empirical Methods in Natural Language Processing (EMNLP-CoNLL), 2007.

Program Committee, AAAI Conference on Artificial Intelligence (AAAI), 2007.

Program Committee, International Conference on Artificial Intelligence and Statistics (AISTATS), 2007.

Program Committee, European Conference on Machine Learning and European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), 2007.

Program Committee, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2007.

Program Committee, International Workshop on Mining and Learning on Graphs (MLG), 2007.

Program Committee, HLT-NAACL workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing, 2007.

Program Committee, International Conference on Machine Learning (ICML), 2006.

Program Committee, AAAI Conference on Artificial Intelligence (AAAI), 2006.

Program Committee, Uncertainty in Artificial Intelligence (UAI), 2006.

Program Committee, International Conference on Knowledge Discovery and Data Mining (KDD), 2006.

Program Committee, European Conference on Machine Learning and European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), 2006.

Program Committee, ICML workshop on Learning with Nonparametric Bayesian Method, 2006.

Program Committee, ECML/PKDD workshop on Mining and Learning with Graphs, 2006.

Program Committee, HLT-NAACL workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing, 2006.

Program Committee, Uncertainty in Artificial Intelligence (UAI), 2005.

Program Committee, ICML Workshop on Learning with Partially Classified Training Data, 2005.

Program Committee, International Conference on Artificial Intelligence and Statistics (AISTATS), 2005.

Program Committee, International Conference on Machine Learning (ICML), 2004.

Referee: Journal of Machine Learning Research, Machine Learning, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Information Theory, IEEE Intelligent Systems, IEEE Transactions on Neural Networks, ACM Transactions on Knowledge Discovery from Data, Pattern Recognition Letters, Optimization Method and Software, Journal of Computational and Graphical Statistics, Neurocomputing, Journal of Software (China), Neural Information Processing Systems (NIPS), International Joint Conferences on Artificial Intelligence (IJCAI)

Student Supervision

Current Students

Andrew Goldberg, 2005–, PhD., UW Computer Sciences Department

David Andrzejewski (with Mark Craven), 2005–, PhD., UW Computer Sciences Department

Nathanael Fillmore, 2008–, PhD., UW Computer Sciences Department

Jingci Meng (with Chunming Zhang), 2005–, PhD., UW Statistics Department

Former Students

Ming Li, Sept. 2007–March 2008, PhD., visiting student, Computer Science Department, Nanjing University, China

Lijie Heng, 2008, MS., UW Computer Sciences Department

Jurgen Van Gael, 2007, MS., UW Computer Sciences Department

Examining Committee Member

Soumya Ray, PhD.'05, *Learning from data with Complex Interactions and Ambiguous Labels*, UW Computer Sciences Dept., Fourth reader.

Mankyu Sung, PhD.'05, *Scalable, Controllable, Efficient and Convincing Crowd Simulation*, UW Computer Sciences Dept., Fourth reader.

Guodong Guo, PhD.'06, *Face, Expression, and Iris Recognition Using Learning-based Approaches*, UW Computer Sciences Dept., Third reader.

Shaohua Fan, PhD.'06, *Sequential Monte Carlo Methods for Physically Based Rendering*, UW Computer Sciences Dept., Fourth reader.

Pedro Bizarro, PhD.'06, *Adaptive Query Processing: Dealing with Incomplete and Uncertain Statistics*, UW Computer Sciences Dept., Fourth reader.

Ye Chen, PhD.'07, *A Bayesian Network Model of Knowledge-Based Authentication*, UW Operations and Information Management Dept., External reader.

Michael Wallick, PhD.'07, *Automatic Organization of Large Collections of Photographs*, UW Computer Sciences Dept., Third reader.

Jesse Davis, PhD.'07, *View Learning: A Statistical Relational Approach to Mining Biomedical Databases*, UW Computer Sciences Dept., Fifth reader.

Edward W. Wild, PhD.'08, *Optimization-Based Machine Learning and Data Mining*, UW Computer Sciences Dept., Sixth reader.

Su Zhang, PhD. ABD'05. *Network Traffic Characterization*, UW Computer Sciences Dept., Third reader, prelim. committee.

Burr Settles, PhD. ABD'06. *Active Learning for Sequence Labeling and Multiple Instance Tasks*, UW Computer Sciences Dept., Fourth reader, prelim. committee.

Louis Oliphant, PhD. ABD'06. *Using Probabilistic Models to Direct Search in Inductive Logic Programming*, UW Computer Sciences Dept., Third reader, prelim. committee.

Yong Lu, PhD. ABD'06. *A Computational Framework for Multi-Species Data Integration in Systems Biology*, Carnegie Mellon University, Computer Science Dept., External reader, prelim. committee.

Hector Corrada Bravo, PhD. ABD'07. *Algorithms for Scalable Probabilistic Inference and Regularized Kernel Estimation*, UW Computer Sciences Dept., Third reader, prelim. committee.

Mugizi Robert Rwebangira, PhD. ABD'07. *Techniques for Exploiting Unlabeled Data*, Carnegie Mellon University, Computer Science Dept., External reader, prelim. committee.

Trevor Walker, PhD. ABD'07. *Relational Methods Incorporating Domain Knowledge for Transfer in Reinforcement Learning*, UW Computer Sciences Dept., Fifth reader, prelim. committee.

Arup Dutta, PhD. ABD'07. *Artificial Neural Network Approach to Crash Modeling and Prediction*, UW Department of Civil and Environmental Engineering, External reader, prelim. committee.

Lisa Torrey, PhD. ABD'07. *Relational Transfer in Reinforcement Learning*, UW Computer Sciences Dept., Fourth reader, prelim. committee.

Aarti Singh, PhD. ABD'07. *Density Level Set Estimation: New Theory and Applications to Statistical Inference and Learning*, UW Department of Electrical and Computer Engineering, Fourth reader, prelim. committee.

Jian Liu, PhD. ABD'08. *Mapping Soil Variation with Satellite-based Observations of Surface Dynamics*, UW Department of Geography, External reader, prelim. committee.

Feng Liu, PhD. ABD'08. *Synthesizing Novel Multimedia from Images and Videos*, UW Computer Sciences Dept., Third reader, prelim. committee.

Nathan Rosenblum, PhD. ABD'08. *Recovering Program Binary Provenance*, UW Computer Sciences Dept., Second reader, prelim. committee.

Departmental and University Service

Participating as an advisor in the Undergraduate Research Scholars Program at University of Wisconsin. The program helps first- and second-year undergraduates get hands-on experience in research. 2007.

UW Computer Sciences Department, graduation admission committee, 2007.

UW Computer Sciences Department, graduation admission committee, 2006.

UW Computer Sciences Department, faculty recruiting committee, 2005.

Carnegie Mellon University Center for Automated Learning and Discovery, PhD. student speaking requirement committee, 2003-2004.

June 13, 2008