

YI ZHANG

School of Engineering, SOE3
University of California, Santa Cruz
1156 High Street, Santa Cruz, CA 95064-1077
url: <http://www.soe.ucsc.edu/~yiz>
yiz@soe.ucsc.edu

Research Interests

Information Retrieval, Natural Language Processing, Applied Machine Learning, Data Mining,

Education

| | | |
|-----------------------------------|---|------------|
| Carnegie Mellon University | School of Computer Science Specializing in Language Technologies Dissertation: Bayesian Graphical Models for Adaptive Information Filtering | Ph.D. 2005 |
| Carnegie Mellon University | School of Computer Science | M.S. 2001 |
| Tsinghua University | Department of Computer Science and Technology | B.S. 1999 |

Academic Appointments

| | |
|-----------|--|
| 2005- | <i>Assistant Professor</i> Jack Baskin School of Engineering, University of California, Santa Cruz Other affiliations: Computer Science, Computer Engineering, Applied Math and Statistics |
| 1999-2005 | <i>Graduate Student Researcher</i> Distributed Information Retrieval Group, School of Computer Science, Carnegie Mellon University |
| 2004 | <i>Invited Lecturer</i> Software Institute, Nanjing University |

Selected Awards and Grants

Grants:

US Department of Education, Explicit Scaffolding for Word Learning in Context through Multimedia Multilingual Word Annotation, (co-PI, 2008-2011)
Air Force Research Young Investigator Award: Adaptive Information Filtering (PI, 2008-2011)
National Science Foundation Grant (CISE-IIS): III-COR: Proactive Personalized Information Integration and Retrieval (PI, 2007-2010)
Google Faculty Research Award (PI, 2005-2006)
Microsoft Beyond Search Semantic Computing and Internet Economics Award (PI, 2008)
Research Gift from Yahoo (PI, 2006)
Research Gift from Nokia Lab (PI, 2007)
Research Gift from NEC Lab (PI, 2007)
Research Gift from Bosch Lab (PI, 2007-2008)
Research grant from Los Alamos National Lab (Single PI)
UC Micro grant (Twice PI, once co-PI, 2006, 2007)

Awards

2002 Best Paper Award at 2002 Annual International ACM SIGIR Conference (SIGIR is the best conference in Information Retrieval. I am the first author of the only Best Paper out of several hundred submissions in 2002. More information is available at <http://www.sigir.org/awards/awards.html>)

IBM Fellowship (2003-2004, 2004-2005)
First Class Scholarships at Tsinghua University(1994-1999)

Industry Experience

2009 Consultant for Exlibris on recommender systems.

2008 Consultant for HP Lab, Palo Alto, CA on large scale web text mining tasks.

2005 - 2007 Technical Advisory Board Member for Provilla Inc., Mountain View, CA
Information Leaking: Provide technical advice on the company's product for preventing sensitive information from leaving organizational boundaries. The company is acquired by Trend Micro Inc. in 2007.

2004 - 2005 Consultant for Ziff Brothers Investments, New York: *News Recommendation*: Help improving a news filtering system used by financial analysts.

2001 - 2002 Consultant for Mars-System Inc., University of Pittsburgh Medical Center:
Medical Information Extraction: Develop an intelligent information extraction algorithm to convert unstructured text in hospital diagnosis records to structured data.

2003 - 2004 Co-op at IBM Almaden Research Center: *Semi Structured Information Retrieval*: Design and evaluate a fast text retrieval system on top of the DB2 relational database. Conduct research on statistical learning for On-Line Analytical Processing(OLAP).

2002 Intern at IBM Almaden Research Center: *Social Networks*: Research on mining social networks from emails to automatically recommend experts on different topics.

Publications in Refereed Journals, Conferences and Workshops

Yun Chi, Shenghuo Zhu, Koji Hino, Yihong Gong, and Yi Zhang, iOLAP: A Framework for Analyzing the Internet, Social Networks, and Other Networked Data, To appear in IEEE Transactions on Multimedia

Yun Chi, Shenghuo Zhu, Yihong Gong, Yi Zhang, Probabilistic polyadic factorization and its application to personalized recommendation. ACM 17th Conference on Information and Knowledge Management (CIKM), 2008

Yi Zhang, Complex Adaptive Filtering User Profile Using Graphical Models, Journal of Information Processing and Management 44 , pp. 1886-1900 (2008)

Yi Zhang, Adaptive Information Filtering, To appear as a book chapter in Text Mining: Theory, Application, and Visualization, Ashok Srivastava (Editor), Mehran Sahami (Editor), Chapman Hall/Crc Data Mining and Knowledge Discovery Series

Sarah K. Tyler, Yi Zhang: Open Domain Recommendation: Social Networks and Collaborative Filtering. ADMA 2008: 330-341

Jonathan Koren, Y. Zhang, and X. Liu. Personalized Faceted Search. In Proceedings of the 17th International Conference on World Wide Web, Beijing, China (WWW 2008). (11% accepted)

Yi Zhang, Wei Xu, Fast Exact Maximum Likelihood Estimation for Mixture of Language Models, In Proceedings of the 30st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, Netherland (SIGIR poster 2007). A longer version is accepted for publication, IPM Information Processing and Management Journal, 2008.

Yi Zhang, Jonathan Koren, Efficient Bayesian Hierarchical User Modeling for Recommendation Systems, In Proceedings of the 30st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, Netherland (SIGIR 2007). (17.5% accepted)

Yi Zhang, Wei Xu, Fast Exact Maximum Likelihood Estimation for Mixture of Language Models, In Proceedings of the 30st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, Netherland (SIGIR poster 2007). A longer version is accepted for publication, IPM Information Processing and Management Journal, 2008.

Jonathan Koren, Yi Zhang, Sasha Ames, Carlos Maltzahn, Ethan Miller. Searching and Navigating Petabyte Scale File Systems Based on Facets. In Proceeding of 2007 ACM Petascale Data Storage Workshop (PDSW 07), International Conference for High Performance Computing, Networking, Storage and Analysis (SC07).

Sung Kim, James Whitehead, Yi Zhang, Classifying Software Changes: Clean or Buggy? To appear in IEEE Transaction on Software Engineering.

Zuobing Xu, Ram Akella, Yi Zhang, Incorporating Diversity and Density in Active Learning for Relevance Feedback. In Proceedings of the European Conference on Information Retrieval, Lecture Notes on Computer Science, 2007. (22% accepted)

Ergin Elmacioglu, Min-Yen Kan, Dongwon Lee, Yi Zhang, Web Based Linkage. In Proceedings of Workshop on Web Information and Data Management, Lisboa, Portugal (WIDM 2007). (25% accepted)

Philip Zigoris, Yi Zhang Bayesian Adaptive User Profiling with Explicit & Implicit Feedback, Conference on Information and Knowledge Management (CIKM 2006). (15% accepted)

Phillip Zigoris, Damian Eads and Yi Zhang, Unsupervised Learning of Tree Alignment Models for Information Extraction, Workshop on Mining Structured Data, International Conference on Data Mining, 2006.

Haward Jie, Yi Zhang, Personalized Faceted Query Expansion, ACM SIGIR Faceted Search workshop 2006.

Ashok Srivastava, Ram Akella, Vesselin Diev, Sakthi Preethi, Dawn McIntosh, Emmanuel Pontikakis, Zuobing Xu and Yi Zhang, Enabling the Discovery of Recurring Anomalies in Aerospace Problem Reports using High-Dimensional Clustering Techniques IEEE Aerospace Conference 2006.

Yi Zhang, Jamie Callan, Combine Multiple Forms of Evidence while Filtering, In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing , Vancouver, Canada, 2005. (31.6% accepted)

Yi Zhang, Using Bayesian Priors to Combine Classifiers for Adaptive Filtering, In Proceedings of the 27st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, United Kingdom, 2004. (22% accepted)

Yi Zhang, Wei Xu, Jamie Callan, Exploration and Exploitation in Adaptive Filtering Based on Bayesian Active Learning, In Proceedings of the Twentieth International Conference on Machine Learning (ICML 2003). (32% accepted)

Byron Dom, Iris Eiron, Alex Cozzi, Yi Zhang, Graph-Based Ranking Algorithms for E-mail Expertise Analysis, In Data Mining and Knowledge Discovery Workshop (DMKD2003) in ACM SIGMOD Conference (SIGMOD2003).

Yi Zhang, Wei Xu and Jamie Callan, Exact Maximum Likelihood Estimation for Word Mixtures, In Text Learning Workshop in International Conference on Machine Learning (ICML 2002).

Yi Zhang, Jamie Callan and Thomas Minka, Novelty and Redundancy Detection in Adaptive Filtering, In Proceedings of the 25st Annual International ACM SIGIR Conference on Research and

Development in Information Retrieval (SIGIR 2002). (20% accepted)

Yi Zhang and Jamie Callan, Maximum Likelihood Estimation for Filtering Thresholds, In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001). (23% accepted)

Other publications

Yi Zhang and Jamie Callan, An Unbiased Generative Model for Setting Dissemination Thresholds, Book chapter in W. B. Croft and J. Lafferty, editors, Language Modeling for Information Retrieval (pp. 189-218), Kluwer Academic Publishers, 2003.

Jordi Mondria, Thomas Wu, Yi Zhang, The Determinants of International Investment and Attention Allocation: Using internet search query data (working paper).

Yi Zhang, Bayesian Graphical Models for Adaptive Information Filtering, Ph.D. Dissertation, 2005.

Ethan Zhang, Yi Zhang , UCSC at TREC 2006 Blog Opinion Mining, In Proceedings of the 2006 Text REtrieval Conference. National Institute of Standards and Technology, special publication, 2006.

Kevyn Collins-Thompson, Paul Ogilvie, Yi Zhang, and Jamie Callan, Information filtering, novelty detection, and named-page finding, In Proceedings of the 2002 Text REtrieval Conference, National Institute of Standards and Technology, special publication, 2002.

Yi Zhang and Jamie Callan, The bias problem and language models in adaptive filtering, In Proceedings of the 2001 Text REtrieval Conference, National Institute of Standards and Technology, special publication, 2001.

Yi Zhang and Jamie Callan, YFilter at TREC-9, In Proceedings of the Ninth Text REtrieval Conference, National Institute of Standards and Technology, special publication, 2000.

Selected Public Lectures and Activities

Invited talk at Jason Study (an independent group of scientists that advises the U.S. government on matters of science and technology http://en.wikipedia.org/wiki/JASON_Defense_Advisory_Group), 2007

Invited talk at Universities: Stanford, Carnegie Mellon, et. al.

Invited keynote speech at Information Retrieval and Applications of Graphical Models Workshop at ACM SIGIR, 2007

Tutorial on “Supervised and Semi-supervised Learning for Information Retrieval” at ACM SIGIR, Amsterdam, 2007

Tutorial on “Machine Learning for Information Retrieval” at ACM SIGIR, Seattle, 2006

Invited talk at HP Labs, CA 2007

Invited talk at Google Faculty Summit, CA, 2006

Invited participant at Microsoft Faculty Summit, Seattle, USA 2006

Invited talk at SRI Artificial Intelligence Lab, CA, USA 2006

Invited talk at NEC Cupertino Research Lab, CA, USA 2005

Invited talk at Microsoft Research Asia, Beijing, China 2004

Invited talk at IBM Almaden Research Center, CA, USA 2004

Invited talk at Microsoft Research Cambridge, United Kingdom 2004

Invited talk at Chinese Academy of Science, Chengdu Campus, China 2004

Activities in Professional Associations

2008, Program co-Chair for Information Retrieval Area, ACM Conference on Information and Knowledge Management (CIKM)

2005, 2006, 2007, 2008, 2009 Program Committee Member of Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR)

2008 Program Committee Member of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

2008 Program Committee Member of International Conference on Machine Learning (ICML)

2007 Program Committee Member for SIGIR workshop on Learning to Rank

2007 Program Committee Member for SIGIR workshop on Information Retrieval and applications of Graphical Models

2007 Program Committee Member for 18th European Conference on Machine Learning (ECML) and the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)

2007 Program Committee Member for Ninth International Workshop on Web Information and Data Management in CIKM 2007

2007 Program Committee of the 2nd Context-based Information Retrieval (CIR-07) workshop

2007 Program Committee of International Workshop on Dynamic Taxonomies and Faceted Search

2006 Program Committee Member of Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (COLING-ACL);

2006, 2007 Program Committee Member of Human Language Technology conference - North American Chapter of the Association for Computational Linguistics Annual Meeting (HLT/NAACL)

2006 Program Committee Member of IEEE Granular Computing

2006 Program Committee Member of ACM Symposium on Applied Computing Special Track on Information Access and Retrieval (ACM SAC-IAR)

2006 Program Committee Member of Symposium on String Processing and Information Retrieval (SPIRE)

2004-2006 Reviewer for Journal of Information Retrieval, ACM Transactions on Information Systems, Journal of American Society for Information Science, Journal of Knowledge and Information Systems(KAIS), The special issue of Journal of Database Management on Database Technology for Enhancing National Security, IEEE Transactions on Multimedia, Machine Learning Journal, Neural Information Processing Systems (NIPS), Conference on Information and Knowledge Management (CIKM), ACM Annual Special Interest Group on Computer-Human Interaction Conference (SIGCHI 2006)