Xuerui Wang

CONTACT INFORMATION

1201 Sycamore Ter Spc 202
Sunnyvale, CA 94086
USA

Voice: (408) 338-5681
E-mail: xuerui_wang@yahoo.com
WWW: http://www.cs.umass.edu/~xuerui

RESEARCH INTERESTS

Statistical and computational machine learning (ML), data mining (DM) for large data sets, online advertising, information retrieval (IR), topic models of text, and social network analysis (SNA).

EDUCATION

University of Massachusetts, Amherst, Massachusetts, USA

Doctor of Philosophy in Computer Science — Apr. 2009
- Advisor: Andrew McCallum

Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Master of Science in Knowledge Discovery and Data Mining — May 2003
- Advisor: Tom Mitchell

Tsinghua University, Beijing, P. R. China

Master of Engineering in Control Theory and its Applications — Jul. 2001
- Advisor: Wenhuang Liu

Bachelor of Engineering in Automation — Jul. 1999

EXPERIENCE

Yahoo! Labs, Santa Clara, California — Feb. 2009—Present

Scientist, Contextual and Display Advertising

Conducting research on non-guaranteed delivery (NGD) traffic forecasting, evaluating model performance and developing system APIs.

Studied performance based user click feedback methods to improve ad placement in Yahoo!'s Keystone contextual advertising system, developed evaluation metrics and adopted different exploration and exploitation strategies.

Designed an empirical Bayes framework to smooth click-through rate (CTR) estimation by leveraging the data hierarchy and temporal continuation in data.

Conducted research on forecasting ad performance in the Keystone contextual advertising system via searching page matching in history.

University of Massachusetts, Amherst, Massachusetts — Jun. 2004—Jan. 2009

Research Assistant, Structured Topic Models: Jointly Modeling Words and Their Accompanying Modalities

Invented novel undirected topic models with both fast inference and clear interpretability. Incorporated information from multiple, heterogeneous modalities, and is much more convenient than their directed counterparts.

Designed new probabilistic, generative models to improve role discovery and group discovery in social networks by taking (textual) attributes of interactions into account. Applications include email messages, political voting records and academic literatures.

Studied dynamic topic trends in large text collections in a probabilistic, generative manner with timestamps as observed random variables. Interesting difference from

traditional Markov transition based dynamic models.

Designed new topical n-gram models that discover topical phrases in context, significantly increased interpretability compared to the bag-of-words models. Better performance in IR experiments on large TREC collections.

Invented efficient multi-conditional learning methods combining generative and discriminative models, and applying them to various classification, clustering, and information retrieval tasks.

Synthesis Project: Inferring Gene Annotations in Gene Ontology

Developed new generative models to predict Gene Ontology annotations from massive gene expression data.

Yahoo! Research, Santa Clara, California   May. 2008—Aug. 2008

Research Intern, Robust Cross-Language Query Classification with external Web Evidence

Developed new methodologies to classify non-English queries by first collecting Web evidence in the native language of the original queries, machine translating the evidence into English, and inferring the queries' class labels from the translated evidence.

Experimented a new framework in online advertising to populate products of interest via query rewriting techniques.

Yahoo! Research, Santa Clara, California   Jun. 2007—Sep. 2007

Research Intern, Search based Forecasting of Ad Volume in Online Advertising

Invented a two-level search based method for realtime forecasting the future performance of internet ads based on replaying billions of historical data.

UtopiaCompression Corp., Los Angeles, California   Jun. 2003—Jun. 2004

Research Scientist, Pattern-Driven Image/Video/Text Compression/Mining

Designed and developed an intelligence-based, pattern-driven image compressor that comprehends an image as a unified and interrelated entity, instead of unrelated blocks of data.

Developed efficient statistical machine learning algorithms to extract features from images and to deal with missing values in image data.

Conducted research on XML compression using machine learning methods, and co-wrote accepted proposals for research funding from governmental organizations such as DoD, NASA and NSF.

Carnegie Mellon University, Pittsburgh, Pennsylvania   Aug. 2001—May 2003

Master Thesis: Scientific Data Mining to Understand Human Brain Function

Developed machine learning methods that can be used to discover the spatial-temporal fMRI patterns that support probabilistic predictions about the cognitive states of human subjects.

Discovered representations that are intermediate between high level cognitive states and the raw fMRI voxel activities, and designed classifiers that could be efficiently trained across subjects and across contexts.

Graduate Assistant, Multi-Agent Learning

Conducted research on multi-agent reinforcement learning using profit sharing plan that allows agents to learn a behavior progressively without any instruction and only with delayed rewards.

Co-Designer, Fly Through The Universe

Designed an R-tree-based algorithm to index terabyte astronomical data sets and digitally simulated a craft exploring the universe made up of millions of galaxies.

Tsinghua University, Beijing, P. R. China   Jun. 1998—Jul. 2001

Master Thesis: Research and Design for Knowledge Management System

Formulated a new architecture of knowledge management systems and designed a web-based knowledge management system facing multi-source data.

Research Assistant, HY-CIMS Project

Conducted research on new decision support technologies with data mining / data warehousing. Studied the experimental infrastructure for distributed database system based on MySQL. Developed the undergraduate major-choosing expert system.

Research Assistant, Web-based Decision Support Systems

Designed and implemented a web-based decision support system using Oracle, Lotus Notes and MS Visual Basic.

Jiangsu Huaiyin Factory, Huaiyin, Jiangsu, P. R. China Jun. 1999—Aug.1999

Chief Designer, REMONTANT Project

Designed and implemented the model-based product development system, the key part of the REMONTANT project, using Oracle and MS Visual Basic.

The Clover Co. Ltd., Chongqing, P. R. China   Jun. 1997—Aug.1997

Summer Intern, Electronic Form King

Developed the expression evaluation module of Electronic Form King with MS Visual C++.

RECENT HONORS The Graduate School Fellowship, University of Massachusetts, May, 2007

Finalist, the Microsoft Research/Live Labs Graduate Fellowship, Dec., 2006

Passed the Ph.D. Portfolio with *distinction* at University of Massachusetts, May 2006

The Best Foundational Paper Award, American Medical Informatics Association, Nov. 2003

The Graduate Fellowship, School of Computer Science, Carnegie Mellon University, Aug. 2001, Aug. 2002

Rockwell Automation Scholarship, Rockwell International Corporation, Dec. 2000

Graduated with *honor* from Tsinghua University, Jul. 1999

Sequent-Chen Daren Outstanding Student Scholarship, Hong Kong Sequent Ltd., Dec. 1998

Student Social Work Scholarship, Excellent Student Cadre, Tsinghua University,

May 1998

Outstanding Student Scholarship, Tsinghua University, Nov. 1996, Nov. 1997

Social Practice Scholarship (Golden Prize), Tsinghua University, Oct. 1996

PATENTS

1. Vanja Josifovski, Evgeniy Gabrilovich, Andrei Broder, Bo Pang, and Xuerui Wang, Cross-Lingual Query Classification, Pending, Application filed in Oct. 2008

2. Vanja Josifovski, Xuerui Wang, Marcus Fontoura, and Andrei Broder, System and Method for Estimating an Amount of Traffic Associated with a Digital Advertisement, Pending, Application filed in Nov. 2007

SERVICES

Program Committee, The 2nd ACM SIGKDD Workshop on Social Network Mining and Analysis (SNA-KDD)

Reviewer, The ACM Transactions on Information Systems (TOIS)

Reviewer, Journal of Machine Learning Research (JMLR)

Reviewer, Information Processing & Management (IPM)

Reviewer, Association for Computational Linguistics (ACL)

Reviewer, International Conference on Machine Learning (ICML)

Reviewer, Uncertainty in Artificial Intelligence (UAI)

Reviewer, American Association for Artificial Intelligence (AAAI)

Reviewer, Neural Information Processing Systems (NIPS)

Reviewer, ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD)

Reviewer, ACM Conference on Information and Knowledge Management (CIKM)

Graduate Representative, Department of Computer Science, University of Massachusetts

Librarian of Graduate Student Library, Department of Computer Science, University of Massachusetts

SKILLS

Programming languages: C/C++, Java, Hadoop, Pig Latin, Matlab, Perl, Python, Mathematica, Splus, R, SQL, Assembly Language (Intel X86 series), Fortran 77, Pascal, Visual Basic, CLISP, Prolog, HTML, etc.

Systems: Windows 95/98/NT/ME/2000/XP/Vista, Macintosh, Unix (Linux especially).

Languages: Chinese (native), English (fluent), Japanese (fair) and German (basic).

GRADUATE COURSEWORK

Computer Science	Machine Learning, Statistical Approaches to Learning and Discovery, Multimedia Databases and Data Mining, Graduate Algorithms, Computational Analyses of Brain Imaging, Information Retrieval, Advanced Software Engineering, Theory of Computation, Bioinformatics, etc.
Statistics	Probability and Statistics, Intermediate Statistics, Statistical Computing, Time Series Analysis, Nonparametric Methods, etc.

REFEREED
PUBLICATIONS

1. Xuerui Wang, Wei Li, Ying Cui, Bruce Zhang, and Jianchang Mao, Click-Through Rate Estimation for Rare Events in Online Advertising, *Online Multimedia Advertising: Techniques and Technologies*, IGI Global, 2010.

2. Wei Li, Xuerui Wang, Ruofei Zhang, Ying Cui, Jianchang Mao, and Rong Jin, Exploitation and Exploration in a Performance based Contextual Advertising System, *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010.

3. Xing Wei, Fuchun Peng, Huishin Tseng, Yumao Lu, Xuerui Wang and Benoit Dumoulin, Search with Synonyms: Problems and Solutions, *Proceedings of the 23rd International Conference on Computational Linguistics*, 2010.

4. Haizheng Zhang, Ke Ke, Wei Li and Xuerui Wang, Graphical Models based Hierarchical Probabilistic Community Discovery in Large-Scale Social Networks, *International Journal of Data Mining, Modelling and Management*, Vol. 2, No. 2, pp. 95-116, 2010.

5. Xuerui Wang and Andrew McCallum, Structured Topic Models: Jointly Modeling Text and Its Accompanying Modalities, *VDM Verlag, ISBN: 978-3-639-20557-2*, 2009.

6. Xuerui Wang, Andrei Broder, Marcus Fontoura and Vanja Josifovski, A Search-based Method for Forecasting Ad Impression in Contextual Advertising, *Proceedings of the 18th International World Wide Web Conference*, pp. 491-500, 2009.

7. Xuerui Wang, Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski and Bo Pang, Cross-Language Query Classification using Web Search for Exogenous Knowledge, *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining*, pp. 74-83, 2009.

8. Xuerui Wang, Andrei Broder, Marcus Fontoura and Vanja Josifovski, A Note on Search based Forecasting of Ad Volume in Contextual Advertising, *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pp. 1343-1344, 2008.

9. Xuerui Wang, Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski and Bo Pang, Cross-lingual Query Classification: a Preliminary Study, *Proceedings of the 17th ACM Conference on Information and Knowledge Management Workshop on Improving Non-English Web Searching*, pp. 101-104, 2008.

10. Xuerui Wang, Chris Pal and Andrew McCallum, Generalized Component Analysis for Text with Heterogeneous Attributes, *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 794-803, 2007.

11. Xuerui Wang, Andrew McCallum and Xing Wei, Topical N-grams: Phrase and Topic Discovery, with an Application to Information Retrieval, *Proceedings of the 7th IEEE International Conference on Data Mining*, pp. 697-702, 2007.

12. Haizheng Zhang, Wei Li, Xuerui Wang, C. Lee Giles, Henry C. Foley and John Yen, HSN-PAM: Finding Hierarchical Probabilistic Groups from Large-Scale Networks, *Proceedings of the 7th IEEE International Conference on Data Mining Workshop on Data Mining in Web 2.0 Environments*, pp. 27-32, 2007.

13. Andrew McCallum, Xuerui Wang and Andres Corrada-Emmanuel, Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email, *Journal of Artificial Intelligence Research*, Vol. 30, pp. 249-272, 2007.

14. Xing Wei, Jimeng Sun and Xuerui Wang. Dynamic Mixture Model for Multiple Time Series, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pp. 2909-2914, 2007.

15. Andrew McCallum, Xuerui Wang and Natasha Mohanty, Joint Group and Topic Discovery from Relations and Text, *Statistical Network Analysis: Models, Issues and New Directions, Lecture Notes in Computer Science 4503*, pp. 28-44, 2007.

16. Chris Pal, Michael Kelm, Xuerui Wang, Greg Druck and Andrew McCallum, On Discriminative and Semi-Supervised Dimensionality Reduction, *The 20th Annual Conference on Neural Information Processing Systems Workshop on Novel Applications of Dimensionality Reduction*, 2006.

17. Xuerui Wang and Andrew McCallum, Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends, *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 424-433, 2006.

18. Andrew McCallum, Chris Pal, Greg Druck and Xuerui Wang, Multi-Conditional Learning: Generative/Discriminative Training for Clustering and Classification, *Proceedings of the 21st National Conference on Artificial Intelligence*, pp. 433-439, 2006.

19. Wei Li, Xuerui Wang and Andrew McCallum, A Continuous-Time Model of Topic Co-occurrence Trends, *Proceedings of the 21st National Conference on Artificial Intelligence Workshop on Event Extraction and Synthesis*, pp. 48-53, 2006.

20. Xuerui Wang, Natasha Mohanty and Andrew McCallum, Group and Topic Discovery from Relations and Their Attributes, *Advances in Neural Information Processing Systems 18*, pp. 1449-1456, 2006.

21. Chris Pal, Xuerui Wang, Michael Kelm and Andrew McCallum, Multi-Conditional Learning for Joint Probability Models with Latent Variables, *The 19th Annual Conference on Neural Information Processing Systems Workshop on Advances in Structured Learning for Text and Speech Processing*, 2005.

22. Xuerui Wang, Natasha Mohanty and Andrew McCallum, Group and Topic Discovery from Relations and Text, *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Workshop on Link Discovery: Issues, Approaches and Applications*, pp. 28-35, 2005.

23. Andrew McCallum, Andres Corrada-Emmanuel and Xuerui Wang, Topic and Role Discovery in Social Networks, *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pp. 786-791, 2005.

24. Andrew McCallum, Andres Corrada-Emmanuel and Xuerui Wang, The Author-Recipient-Topic Model for Topic and Role Discovery in Social Networks: Experiments with Enron and Academic Email, *The 18th Annual Conference on Neural Information Processing Systems Workshop on Structured Data and Representations in Probabilistic Models for Categorization*, also appeared as UMass Technical Report UM-CS-2004-096, 2004.

25. Xuerui Wang, Rebecca Hutchinson and Tom Mitchell, Training fMRI Classifiers to Discriminate Cognitive States across Multiple Subjects, *Advances in Neural Information Processing Systems 16*, pp. 709-716, 2004.

26. Tom Mitchell, Rebecca Hutchinson, Radu Niculescu, Francisco Pereira, Xuerui Wang, Marcel Just and Sharlene Newman, Learning to Decode Cognitive States from Brain Images, *Machine Learning*, Vol. 57, Issue 1-2, pp. 145-175, 2004.

27. Tom Mitchell, Rebecca Hutchinson, Marcel Just, Radu Niculescu, Francisco Pereira and Xuerui Wang, Classifying Instantaneous Cognitive States from fMRI data, *Proceedings of the American Medical Informatics Association 2003 Annual Symposium*, pp. 465-469, 2003.

28. Tom Mitchell, Rebecca Hutchinson, Marcel Just, Sharlene Newman, Radu Stefan Niculescu, Francisco Pereira and Xuerui Wang, Machine Learning of fMRI Virtual Sensors of Cognitive States, *The 16th Annual Conference on Neural Information Processing Systems Workshop on Computational Neuroimaging: Foundations, Concepts and Methods*, 2002.

29. Xuerui Wang and Wenhuang Liu, Research on CBR in Knowledge Management Systems, *Computer Engineering and Applications*, Vol. 38, No. 2, pp. 181-184, 2002.

30. Zefeng Zheng, Xuerui Wang and Wenhuang Liu, Research on a Solution of Business Questions Modeling in Decision Support Systems, *Proceedings of the 2001 International Conference on Artificial Intelligence*, 2001.

31. Xiu Li, Shouju Ren, Wenhuang Liu and Xuerui Wang, A Solution of Job-Shop Scheduling Problems based on Genetic Algorithms, *Proceedings of the 2001 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 3, pp. 1823-1828, 2001.

32. Xiu Li, Xuerui Wang, Wenhuang Liu, and Lin Liao, Research on Web-based Data Warehouse using XML, *Proceeding of the 3rd International Conferences on Information Technology and Information Networks*, Vol. 5, pp. 42-47, 2001.

33. Xuerui Wang, Wenhuang Liu, Lin Lei and Shouju Ren, Research for Internet-based Integration and Self-organization Supply Chain Management, *Proceedings of the 4th World Multiconference on Systemics, Cybernetics and Informatics, Volume I in Information Systems*, 2000.

34. Lin Lei, Wenhuang Liu, Shouju Ren and Xuerui Wang, ERP-Based Business Process Reengineering, *Proceedings of the 16th IFIP World Computer Congress, Information Technology for Business Management*, 2000.

UNREFEREED
PUBLICATIONS

35. Xuerui Wang, Structured Topic Models: Jointly Modeling Words and Their Accompanying Modalities, *Ph.D. Thesis*, UM-CS-PhD-2009-003, 2009.

36. Xuerui Wang, David Kulp and Andrew McCallum, Inferring Gene Annotations in Gene Ontology from Gene Expression Data, *UMass CS Synthesis Project Report*, 2006.

37. Xuerui Wang and Andrew McCallum, A Note on Topical N-grams, *UMass Technical Report* UM-CS-2005-071, 2005.

38. Andrew McCallum, Xuerui Wang and Chris Pal, Predictive Random Fields: Latent Variable Models Fit by Multiway Conditional Probability with Applications to Document Analysis, *UMass Technical Report* UM-CS-2005-053, 2005.

39. Xuerui Wang, Tom Mitchell and Rebecca Hutchinson, Using Machine Learning to Detect Cognitive States across Multiple Subjects, *CMU CALD KDD Project Report*, 2003.

40. Xuerui Wang and Tom Mitchell, Detecting Cognitive States Using Machine Learning, *CMU CALD Technical Report for Summer Work*, 2002.