

Xin Zhang

Ph.D in Computer Science and Engineering
Arizona State University
Tempe, AZ 85287-8809

Tel: (623) 215 – 5157
Email: xin@asu.edu
<http://www.public.asu.edu/~xzhang24>

Research Interests:

- **Data Mining and Machine Learning:** Bayesian learning, kernel selection, text mining, classification, clustering, information theory
- **Bioinformatics:** inferring protein-protein interactions and gene regulatory networks, learning causal connection between genes, predicting DNA-protein interactions, knowledge extraction from biological texts, representation and reasoning about signal pathways.
- **Pharmacogenomics:** optimization of assay design on siRNA high-throughput screening (HTS), hit selection, bio-marker identification and validation, statistical analysis on HTS data.
- **Artificial intelligence:** causal theory, knowledge representation and reasoning (with applications to common-sense reasoning, and reasoning with incomplete information), logic programming, planning, multi-agent systems, and AI in general.

Education:

Ph.D, Computer Science	Arizona State University, Tempe, AZ, August 2008
Dissertation: Learning biological interactions from multiple data sources.	
Research Advisor: Dr. Chitta Baral	
M.S., Computer Engineering	University of Arkansas, Fayetteville, Arkansas, 2002
M.S., Computer Science	Harbin Institute of Technology, Harbin, China, 1998
BS, Computer Science	Harbin Institute of Technology, Harbin, China, 1996

Honors and Awards

- **Dissertation Fellowship**, Arizona State University, 2007-2008 (the only recipient in Ira A. Fulton School of Engineering, Arizona State University).
- Grace Hopper Women in Computing Scholarship, 2007.
- **Outstanding Teaching Assistant Award**, Department of Computer Science & Engineering, Arizona State University, 2005-2006 (2 out of 46 TAs in CSE department)
- Nominated for the GPSA Teaching Excellence Award, Arizona State University, 2006-2007.
- Recipient of Certificate for outstanding work as teaching assistant, Department of Computer Science and Engineering, Arizona State University, April, 2006.
- GPSA Travel Grant Award, Arizona State University, Oct 2006.
- GCAT scholarship award, Arizona State University, Spring 2004.
- AAAI Student Travel Scholarship Award, AAAI conference 2004.

- Japanese Excellent Exchange Student Scholarship (Japanese Education Department) 1997-1998.
- GuangHua Outstanding Graduate Student Scholarship, Harbin Institute of Technology, China 1997.
- National People's Scholarship, Harbin Institute of Technology, 1992 – 1996.

Research Experience:

1. Identification of key factors in assay optimization for siRNA high-throughput screening

Large-scale RNAi screening is a powerful tool for functional genomic study. Compared to other types of plate-based screening, outlier appearance for RNAi screen is more affected by variations introduced during experiments, including cell transfection, incubation time, endpoint assay sensitivity, plate-to-plate and well-to-well changes. In the study, I identified cell transfection to be a critical step for RNAi screening and developed a modified z factor that provides indications on three aspects of an assay: signal dynamic window, transfection efficiency window, and siRNA toxicity window. To better prioritize the unique parameters for RNAi screen, I further propose a modified data normalization process from an established B-score method.

2. Learning causal relationship between genes and its application on melanoma data

Inferring causal relationships between genes is a very important issue in system biology. I propose a joint learning scheme that integrates multiple data sources. I presented two algorithms, the modified IC (mIC) algorithm and the mIC_CoD algorithm, that combine steady state data with partial prior knowledge of gene topological ordering information, for jointly learning causal relationships among genes. I further applied the algorithms on a melanoma microarray dataset, and identify several important causal relationships within a network of 10 genes. Among the discovered connections, the causal relationships associated with WNT5A, a gene playing an important role in melanoma, are supported by literature.

3. Joint learning of logic relationships for studying protein function

Identifying logic relationships between proteins is essential for understanding their function within cells. In the research, I presented a new approach for joint learning of protein logic relationships from both protein phylogenetic profile and Rosetta Stone data. I used a Bayesian model to incorporate phylogenetic profile data via a likelihood and Rosetta Stone data via a prior. By extending pairwise and triplet logic analysis, I proposed a general method for identifying high order protein logic relationships, such that the presence or absence of one protein can be predicted by the profiles of two or more other proteins. I used the notion of proper function to reflect the actual effect of the predictors on the target protein. With the generalized definitions and framework, the model can be easily extended to infer protein logic relationships with larger number of proteins.

4. Kernel-based method for predicting DNA-binding proteins

DNA-binding proteins play an important role in cellular activities, including DNA replication, DNA repair, packaging and regulation of transcription through discrete DNA-binding domains. In the research, I present a kernel-based learning method for predicting DNA-binding proteins by integrating multiple data sources, including protein structural domain and sequence information from Protein Family (Pfam) database, functional expression data from microarray experiments, protein-DNA binding data from ChIP-chip experiments, and the protein functional annotations from the Gene Ontology. The method applies a kernel function on each type of data source and combines the kernels for classification using Support Vector Machines (SVMs).

5. Inferring disease-related protein-protein interactions from multiple data sources

Protein-protein interactions are vital to the functions of cells. Experimental studies indicate that many diseases are related to subtle molecular events such as protein interactions. In the research, I described a Bayesian modeling framework for jointly learning protein-protein interactions from multiple data sources, including genomic data obtained from non-homology-based methods, the BIND

experimental database, the protein functional annotations from the Gene Ontology and the disease-related genes from NCBI OMIM database. The method is able to identify some known disease-related interactions such as mad cow disease-related prions interactions, and also predicts some potential interactions involving the homologs of human disease-related proteins.

6. Knowledge representation, reasoning and planning with incomplete information

A state-based regression function for planning domains with incomplete information and sensing actions was presented. I improved binary domains, and employed the 0-approximation to define the regression function. I proved the soundness and completeness of our regression formulation with respect to the definition of progression and develop a conditional planner that utilizes our regression function.

Teaching Experience:

Spring 2007: Guest Lecturer, Arizona State University

Lectured for graduate course “CSE571 Artificial Intelligence” on topics of causal theories and learning causal relationship between genes.

Spring 2006: Guest Lecturer, Arizona State University

Lectured for graduate course “CSE598 Application of AI to molecular biology” on topics of Bayesian network and joint learning of protein-protein interactions.

Aug 2003– May 2007: Teaching Assistant, Dept. of Computer Science, Arizona State University

- **Courses**

- CSE205 Concept of Computer Science and Data Structure

Description: object-oriented design and data structures, principles of programming languages and programming in Java, overview of computer science topics

- CSE360 Software Engineering

Description: software life cycle models, project management, team development environments and methodologies, software architectures, quality assurance and standards, legal, ethical issues

- **Duties**

- Leading and lecturing lab sessions
- Developing lab plans
- Grading homework, projects and exams
- Maintaining class web pages (see current class website at: <http://courses.eas.asu.edu/cse205>)
- Giving lectures in the case of instructor absence

Professional Services and Activities

- Vice Chair of IEEE Computer Society Student Chapter, Arizona State University, 2006-2007.
- Outreach Officer in WCS (Women in Computer Science), Arizona State University, 2006-2007.
- Technical program committee member for International Conference on Machine Learning and Applications (ICMLA), 2006 & 2007.
- Judge for 7th Annual MGE@MSA Student Research Conference, Tempe, AZ, May 2007.
- Mentor of FURI (Fulton Undergraduate Research Initiative) program, 2005-2007.
- Student Judge for Science and Technologies, University of Arkansas, 2001 & 2002.

Publications:

Journal publications:

1. Xin Zhang, Seungchan Kim, Tie Wang, and Chitta Baral.. "Joint learning of logic relationships for studying protein function using phylogenetic profiles and the Rosetta Stone method". *IEEE Transactions on Signal Processing*, Vol 54, No. 6, June, 2006.
2. Tran Cao Son, Phan Huy Tu, and Xin Zhang. "Reasoning about Sensing Actions in Domains with Multi-Valued Fluents". *Studia Logica*, 2004 .
3. Qiang Que*, Xin Zhang*, Hongwei Yin, Jeff Kiefer, Donald Chow, Chitta Baral, and Spyro Mousses. "Identification of Key Factors in Assay Optimization for Improving Quality of RNAi Screening", minor revision, *Journal of Biomolecular Screening*. (*equally contributed to this work), resubmitted on August 18, 2008.

Conference publications:

4. Xin Zhang and Chitta Baral. "Kernel method for predicting DNA-binding proteins in yeast from heterogeneous data sources". New Investigator Technical Paper in *Grace Hopper Women in Computing (GHC)*, Colorado, Oct 1-4, 2008.
5. Xin Zhang and Chitta Baral. "Jointly Learning Biological Interactions from Multiple Data Sources: Application to Pharmacogenomics". New Investigator Technical Paper in *Grace Hopper Women in Computing (GHC)*, Orlando, Oct 17-20, 2007.
6. Xin Zhang, Chitta Baral and Seungchan Kim. "An Algorithm to Learn Causal Relations between Genes from Steady State Data: Simulation and Its Application to Melanoma Dataset." *10th Conference on Artificial Intelligence in Medicine (AIME 2005)*.
7. C. Baral, Le-Chi Tuan, Xin Zhang, Tran Son. "Regression with Respect to Sensing Actions and Partial States", *19th National Conference on Artificial Intelligence (AAAI'04)*, 2004.
8. Jim Estipona, Luis Tari, and Xin Zhang. "Sporadic breast cancer gene expression signature". student paper in *WNAR/IMS Meeting*, Albuquerque, NM, June 27-30, 2004.
9. Tran Cao Son, Phan Huy Tu, and Xin Zhang. "Reasoning about Sensing Actions in Domains with Multi-valued Fluents". *Proc. of the IJCAI-03 Workshop on Nonmonotonic Reasoning, Action, and Change (NRAC'03)*, Acapulco, Mexico
10. Xin Zhang and Henry Hexmoor. "Balancing Individual Capabilities and Social Peer Pressure for Role Adoption", *In International / Central and Eastern European Conference on Multi-Agent Systems (CEEMAS-03)*, V. Marik, J. Muller, M. Pechoucek (Eds), pages 100-110, Prague, Springer LNAI2691., 2003.
11. Henry Hexmoor, and Xin. Zhang. "Socially Intelligent Air Combat Simulator", *In Proceedings of The Seventh Pacific Rim International Conference on Artificial Intelligence*, (PRICAI-02), pages 365-374, Tokyo, Japan, 2002.
12. Xin Zhang and Henry Hexmoor. "Algorithms for Utility-based Role Exchange", The 7th International Conference on Intelligent Autonomous Systems (*IAS-7*), Marina Del Ray, CA, 2002.
13. Xin Zhang and Henry Hexmoor, "Utility based Role Exchange", *In the Proceedings of International / Central and Eastern European Conference on Multi-Agent Systems (CEEMAS)*, Pages 413-420, Poland, 2001.
14. Henry Hexmoor and Xin Zhang, "Norms, Roles, and Simulated RoboCup", *In 2nd workshop on norms and institutions in multiagent systems*, (Agents 2001), pages 35-42, Montreal, CA, ACM press, 2001.

15. Kelvin Wang and Xin Zhang, "Experimentation with a Host-Based Parallel Algorithm for Image Processing", *International Conference on Traffic and Transportation Studies (ICTTS 2000)*, 736-743, 2000.
16. H. Hara, X. Zhang, T. Kato, K. Sugawara, "ADIPS Framework and Its Application", *12th International Conference on Information Networking*, Tokyo, Japan, *IEEE Computer Society*, 1998
17. Susumu Konno, Xin Zhang, Atsuhiko Sugiyama, Seikoo Takahashi, "On a View Model of Agent in the Cyber Office", *12th International Conference on Information Networking*, Tokyo, Japan, *IEEE Computer Society*, Jan.21-23, 1998

Submitted:

- Joint learning of protein-protein interactions from genomic data, experimental databases and the Gene Ontology (collaborate with Luis Tari and Chitta Baral, submitted to *BMC Bioinformatics* in April 2008).

References:

Prof. Chitta Baral
Department of Computer Science and Engineering
Arizona State University
Tempe, AZ 85287-8809
chitta@asu.edu
480-727-6047

Dr. Quick Que
Head of Computational Pharmacogenomics
Translational Genomics Research Institute (TGen)
13208 East Shea Blvd
Scottsdale, AZ 85259
qqe@tgen.org
602-358-8353

Prof. Jieping Ye
Department of Computer Science and Engineering
Arizona State University
Tempe, AZ 85287-8809
jieping.ye@asu.edu
480-727-7451

Dr. Mutsumi Nakamura
Department of Computer Science and Engineering
Arizona State University
Tempe, AZ 85287-8809
Mutsumi@asu.edu
480-965-1757