Ya (Anne) Zhang

737 E El Camino Real #437
Sunnyvale, CA 94087

Phone: (814) 883-1323
E-mail: yzz100@psu.edu
Web: http://www.personal.psu.edu/yzz100

RESEARCH INTERESTS

- Bioinformatics and Computational biology: *microarray data analysis, proteomics, genomics, protein interaction network, biological network*
- System biology: *biological pathways and networks, genetic regulatory network, signaling pathway*
- Machine learning and data mining: *data clustering, graph mining, text mining*

TEACHING INTERESTS

- Bioinformatics
- Computational biology
- Data mining
- Machine learning
- Database
- Pattern recognition

EDUCATION

The Pennsylvania State University, University Park, PA August 2000-present
Ph.D. in Information Sciences and Technology (expected in May 2005)
Dissertation: ***Inferring Biological Informative Protein-Protein Interactions: A Machine Learning Approach***
Advisors: Dr. Hongyuan Zha and Dr. Chao-Hsien Chu

Tsinghua University, Beijing, China September 1996- July 2000
B. S. in Biological Sciences and Biotechnology

PROFESSIONAL EXPERIENCES

Research Assistant

School of Information Sciences and Technology, Penn State University May 2001- present

- Predicted protein-protein interactions with a domain-based approach
- Discovered partially co-regulated genes from time-series microarray data with biclustering
- Discovered motifs from yeast promoters based on instance density
- Predicted splice sites in DNA sequences with SVM-based techniques
- Developed progressively display system for very high resolution biomedical images

Research intern

Lawrence Berkeley National Laboratory Summer 2004

- Compared sequence-based and structure-based protein domain definitions
- Identified protein functional modules from high-throughput protein complex data with hyperclique pattern discovery techniques

Lab Lecturer

IST210 *Organization of Data* Fall 2003 and Fall 2001
School of Information Sciences and Technology, Penn State University

- Provide lectures on html, java, java script, SQL, and PHP
- Supervise student's semester-long project of building database applications for a real-world scenario such as ecommerce applications

- Mentor students on a personalized basis for all work

IST220 *Networking and Telecommunications* Spring 2003
School of Information Sciences and Technology, Penn State University

- Provide lectures on html, java, java script, Perl, cgi, and MySQL database
- Supervise student's semester-long project to build real-world network applications
- Mentor students on a personalized basis for all work

Teaching Assistant

IST511: *Information Management* Fall 2004
School of Information Sciences and Technology, Penn State University

IST420: *Fundamentals of Systems and Enterprise Integration* Spring 2004
School of Information Sciences and Technology, Penn State University

IST230 *Language, Logic, and Discrete Mathematics* Fall 2002
School of Information Sciences and Technology, Penn State University

GRANT WRITING EXPERIENCE

Proposal title: "*Development of Quantitative Analysis Tools for Mapping Cellular Signaling Pathways*"

Funded by Department of Health's Tobacco Formula Funded Health Research Program, 2004.

PI: Cheng Dong, Ph.D. Co-PIs: Hongyuan Zha, Ph.D and David Antonetti, Ph.D.

I drafted large portions of the proposal and helped design specific aims.

PUBLICATIONS

Refereed Journal Papers:

1. **Y. Zhang,** X. Ji, C. H. Chu, and H. Zha, "Correlating Summarization of Multi-source News with K-Way Graph Biclustering", *ACM SIGKDD explorations* (*special issue on Web Content Mining*), 2005, accepted.
2. J. Z. Wang, K. Grieb, **Y. Zhang**, C. Chen, Y. Chen, and J. Li, "Machine Annotation and Retrieval for Digital Imagery of Historical Materials", International Journal on Digital Libraries, 2005, accepted.

Refereed Conference Papers:

3. **Y. Zhang**, H. Zha, and C. H. Chu, "Revealing Partially Co-regulated Genes through Time-course Biclustering", in Proc. of *IEEE International Conference on Information and Technology: Coding and Computing*, (*ITCC* 2005), 2005, to appear.
4. **Y. Zhang**, C. H. Chu, H. Zha, Y. Chen, and X. Ji, "Discovering Motifs from Biosequences Based on Instance Density," in Proc. of *IEEE International Conference on Information and Technology: Coding and Computing*, (*ITCC* 2005), 2005, to appear.
5. H. Xiong, X. He, C. Ding, **Y. Zhang**, V. Kumar, S. R. Holbrook, "Identification of Functional Modules in Protein Complexes via Hyperclique Pattern Discovery", in Proc. of *the Pacific Symposium on Biocomputing*, (*PSB* 2005), 2005.
6. **Y. Zhang** and J. Z. Wang, "Progressive Display of Very High Resolution Images using Wavelets," Journal of *American Medical Informatics Association*, Symposium Supplement, vol. 2002 suppl., pp. 944-948, November 2002. [**Nominated for the Best Paper Award**]

Other Publications:

7. **Y. Zhang**, H. Zha, J. Z. Wang and C. Chu, "Gene Co-regulation vs. Co-expression," *the Eighth Annual International Conference on Research in Computational Molecular Biology* (RECOMB), San

Diego (March 27-31, 2004)., Currents in Computational Molecular Biology A. Gramada & P. Bourne (eds.) ACM Press, March 2004., pp 232-233. (Poster)

8. **Y. Zhang**, H. Zha, J. Z. Wang and C. Chu, "Clustering of Time-course Gene Expression Data," *the Eighth Annual International Conference on Research in Computational Molecular Biology* (RECOMB), San Diego (March 27-31, 2004)., Currents in Computational Molecular Biology A. Gramada & P. Bourne (eds.) ACM Press, March 2004., pp.240-241. (Poster)
9. H. Wang & **Y. Zhang**, "Book review: Shaping Web Usability: Interaction Design in Context," *Information Processing and Management*, 39(4), 665 – 666, 2003.

WORKING PAPERS

1. **Y. Zhang**, J.-M. Chandonia1, C. Ding and S. R. Holbrook, "Comparative Mapping of Sequence-based and Structure-based Protein Domains", Journal of *BMC Bioinformatics*, under review.
2. **Y. Zhang**, H. Zha, and C. Chu, "Inferring Interacting Domains: Challenges and Solutions", submitted to *the 13th Annual International conference on Intelligent Systems for Molecular Biology* (*ISMB* 2005).
3. **Y. Zhang**, C. H. Chu, H. Zha, and Y. Chen, "Splice Site Prediction with Linear Kernel and Bayesian Mapping", to be submitted to a *Expert Systems with Applications: Special Issue on Intelligent Bioinformatics Systems.*

SELECTED PROJECTS

- **Inferring protein-protein interactions**

  Discovering interacting proteins is essential for solving the functional genomics puzzle. This project aims at inferring unobserved protein-protein interactions from high throughput interaction data, which only cover a small portion of interactions. As proteins are assumed to interact through their interaction domains, a domain-based approach is employed for the inference. Existing studies tend to oversimplify the problem by introducing two biologically unfounded assumptions: domain interactions are between two individual domains and are independent of each other. To overcome the limitations, the problem of interaction inference is modeled as a constraint satisfiability problem and is solved with linear programming. A summary of the preliminary work is to be submitted to *the 13th Annual International conference on Intelligent Systems for Molecular Biology* (*ISMB* 2005).

- **Comparative mapping of Sequence-based and Structure-based Protein Domains**

  This project provides some insight on domain definition through a comparative mapping of a sequence domain database and a structure domain database. A direct matching approach and an indirect matching approach based on bipartite graph are proposed for the comparative mapping. Focusing on the disagreement, we examine the functions and evolutionary histories of the domains to suggest which domain definition is biologically more informative. A full version of the work has been submitted to *BMC Bioinformatics*.

- **Identification of Functional Modules from Protein Complexes**

  Proteins in a cell usually act cooperatively through interacting to perform certain functions. The proteins required for a common elementary function form a functional module. Protein complexes embody information about functional modules and therefore are used for functional module mining. We applied a hyperclique pattern discovery technique to extract functional modules from protein complex data. Annotation of discovered patterns with Gene Ontology suggests that hyperclique patterns are highly likely to represent functional modules. A summary of the preliminary work has been accepted for publication by the *Pacific Symposium on Biocomputing* 2005 (*PSB* 2005).

- **Correlating Summarization of Multi-source News with K-Way Graph Biclustering**

  We constructed a machine learning method to highlight similarities of multi-source news with shared (sub)topics. Weighted bipartite graphs are employed to model pairs of news articles. Considering two news articles may share several subtopics, spectral K-way clustering is applied to aligns the (sub)topics. A mutual reinforcement principle is then applied to extract topic sentences within each subtopic group. A full version of the work has been accepted for publication in *SIGKDD explorations special issue on Web Content Mining*, 2005.

- **Wavelet-based Image Retrieval System**

  Web DEMO: http://wang14.ist.psu.edu/~zhang/wavezoom/
  We developed a progressive image display system based on wavelet transform. The system was designed to transmit and display high-resolution images (i.e. several gigabytes per image), especially medical images, with great fidelity. Particular regions of the image could be retrieved from the database and displayed in various resolutions. The system includes a file server, a web server and a web interface. A summary of the preliminary work has been published in Journal of *American Medical Informatics Association* (Symposium Supplement) and presented at *AMIA 2002 Symposium*. The paper was nominated for the best paper award.

AWARDS & HONORS

- Travel Fellowship, the Eighth Annual International Conference on Research in Computational Molecular Biology (RECOMB), 2004
- Best Paper Award Nomination, AMIA Annual Symposium, 2002
- Third Prize, “Challenge Cup Science & Technology Contest”, Tsinghua University, China, 1998.
- Excellent Students Scholarship, Tsinghua University, China, 1997.

SELECTED COURSES

Bioinformatics and Data Mining

- Biostatistical methods
- Bioinformatics
- Data Mining
- Probabilistic Algorithms

Computer and Information Sciences

- Data Structures
- Digital Image Processing
- Multimedia Indexing and Retrieval
- Information Management
- Computer Networks
- Introduction to Computer Architecture
- Human-Computer Interaction
- Human Information Behavior
- Qualitative Research Methodology
- Integrative Theories and Methods of the Information Sciences and Technology

Biological Science

- Concepts in Biomolecular Science
- Cell Growth and Differentiation
- Biological regulation
- Molecular Biology
- Biochemistry
- Genetics

PERSONAL INFORMATION

- Gender: Female
- Citizenship: P.R. China
- Visa: F-1

REFERENCES

- Dr. Hongyuan Zha, Professor of Computer Science and Engineering; Affiliate Professor of Statistics; Affiliate Professor of Information Science and Technology

  343F IST Building
  Department of Computer Science and Engineering
  The Pennsylvania State University
  University Park, PA 16802
  Phone: (814) 863-0608 (O)
  Fax: (814) 865-3176
  Email: zha@cse.psu.edu

- Dr. Chao-Hsien Chu, Associate Professor of Information Sciences and Technology; Affiliate Associate Professor of Management Science (Smeal College of Business)

  301K IST Building
  School of of Information Sciences and Technology
  The Pennsylvania State University,
  University Park, PA 16802
  Phone: (814)865-4446(O)
  Fax: (814)865-6426
  Email: chu@ist.psu.edu

- Dr. Stephen R. Holbrook, Staff Scientist, Structural Biology Department, Physical Biosciences Division, Lawrence Berkeley National Laboratory

  Department of Structural Biology and Computational and Theoretical Biology
  Physical Biosciences Division
  Lawrence Berkeley National Laboratory, MS 64-123
  1 Cyclotron Road
  Berkeley, California 94720
  Phone: (510)486-4304 (O)
  Fax: (510)486-6798
  E-mail: srholbrook@lbl.gov