



## **Protein Structure Prediction and Its Understanding Based on Machine Learning Methods\***

Yi Pan  
Chair and Professor  
Computer Science Department  
Georgia State University  
Atlanta, GA 30303, USA  
Email: pan@cs.gsu.edu

### **Abstract**

Understanding protein structures is vital to determining the function of a protein and its interaction with DNA, RNA and enzyme. The information about its conformation can provide essential information for drug design and protein engineering. While there are over a million known protein sequences, only a limited number of protein structures are experimentally determined. Hence, prediction of protein structures from protein sequences using computer programs is an important step to unveil proteins' three dimensional conformation and functions. As a result, prediction of protein structures has profound theoretical and practical influence over biological study. The explanation of how a decision is made during prediction is also important for improving protein structure prediction and guiding the "wet experiments". In this talk, we will show how to use machine learning methods to improve the accuracy of protein structure prediction and to interpret prediction results. We will report our research on using neural networks, Support Vector Machines combined with Decision Tree and Association Rule for protein structure prediction, rule extraction and prediction interpretation. Evaluation and comparisons of various prediction and rule extraction systems will be presented and future research direction in this area will also be identified.

\* This work was jointly with Hae-jin Hu, Jieyue He, Rob Harrison, and PC Tai.

**Biography:** Yi Pan is the chair and a professor in the Department of Computer Science and a professor in the Department of Computer Information Systems at Georgia State University. Dr. Pan received his B.Eng. and M.Eng. degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and his Ph.D. degree in computer science from the University of Pittsburgh, USA, in 1991.

Dr. Pan's research interests include parallel and distributed computing, networking, and bioinformatics. Dr. Pan has published more than 100 journal papers with over 30 papers published in various IEEE journals. In addition, he has published over 100 papers in refereed conferences and co-edited 33 books (including proceedings). He has received many awards from agencies such as NSF, AFOSR, JSPS, IISF and Mellon Foundation. Dr. Pan has served as an editor-in-chief or editorial board member for 15 journals including IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on NanoBioscience, and IEEE Transactions on Systems, Men, and Cybernetics. He has also served as a guest editor for many journals including IEEE/ACM Transactions on Computational Biology and Bioinformatics and IEEE Transactions on NanoBioscience. He has organized numerous international conferences and workshops.

Dr. Pan has delivered over 10 keynote speeches at international conferences. Dr. Pan is an IEEE Distinguished Speaker (2000-2002), a Yamacraw Distinguished Speaker (2002), and a Shell Oil Colloquium Speaker (2002). He is listed in Men of Achievement, Who's Who in Midwest, Who's Who in America, Who's Who in American Education, Who's Who in Computational Science and Engineering, and Who's Who of Asian Americans.