RULE VISUALIZATION OF PROTEIN MOTIF SEQUENCE DATA FOR SECONDARY STRUCTURE PREDICTION

Dr. Leong Lee

Abstract:

Protein secondary structure prediction has been a well-studied research problem in bioinformatics for years. This talk will describe some of the work that we have done to address this problem, in particular a rule-based data mining method called RT-RICO (Relaxed Threshold Rule Induction from Coverings). Our method surpassed the accuracy, or Q3 score, that had been reported for other computational methods for protein secondary structure prediction using the standard datasets, RS126 and CB396. The success of our rule-based method supported the belief that there are meaningful statistical relationships between any secondary structure position and its neighboring amino acids. However, because of the vast amount of rules generated by RT-RICO, potentially useful information within a rule set was difficult to identify. Herein we discuss the results of examining those RT-RICO rules using an existing association rule visualization tool, modified to account for the non-Boolean characterization of protein secondary structure.

Bio:

Leong Lee received his Ph.D. in Computer Science degree from the University of Missouri (Missouri University of Science and Technology) in July 2010. He is now a Visiting Assistant Professor in the Department of Computer Science at the University of North Carolina at Greensboro. He received his B.S. degree from the National University of Singapore, his M.S. (Information Management) degree from the National University of Ireland, Dublin and his M.S. (Computer Science) degree from the University of Missouri. His research focuses on data mining and ontologies with respect to bioinformatics. His current research interests include protein secondary structure prediction, ontology theory and web application development for bioinformatics.