Chemical Genetics and Recommender Systems: Different Problems, Same Solutions

Abstract: In this talk we will present our research in two different areas: Chemical genetics and Recommender systems. The goal of Chemical genetics is to identify small organic molecules that can be used to alter the function of proteins and has emerged as an important experimental technique for studying and understanding complex biological systems. The goal of Recommender systems is to filter vast amounts of information in order to identify the distinct pieces of information that is of relevant to a user. Recommender systems have emerged as a key enabling technology to e-commerce by functioning as virtual experts that are keenly aware of the user’s preferences and tastes. Though Chemical genetics and Recommender systems are entirely different application areas, the characteristics of their underlying data, their dependencies, and the problems arising in them lend themselves to very similar algorithmic solutions. In this talk we explore these similarities and present various methods for improving the accuracy of target-specific structure-activity-relationship and structure-selectivity-relationship models that leverage activity and structural information from similar targets and ligands. And various methods for building highly accurate sparse models for top-N recommendations that incorporate both historical information across different users as well as intrinsic information about the users and the items involved. Finally, we will conclude by outlining a number of different problems that are common to these areas than we believe can benefit from a cross-fertilization of ideas and solution methodologies.

1 George Karypis is a professor at the Department of Computer Science & Engineering at the University of Minnesota, Twin Cities. His research interests spans the areas of data mining, bioinformatics, cheminformatics, high performance computing, information retrieval, collaborative filtering, and scientific computing. His research has resulted in the development of software libraries for serial and parallel graph partitioning (METIS and ParMETIS), hypergraph partitioning (hMETIS), for parallel Cholesky factorization (PSPASES), for collaborative filtering based recommendation algorithms (SUGGEST), clustering high dimensional datasets (CLUTO), finding frequent patterns in diverse datasets (PAFI), and for protein secondary structure prediction (YASSPP). He has coauthored over 200 papers on these topics and a book title “Introduction to Parallel Computing” (Publ. Addison Wesley, 2003, 2nd edition). In addition, he is serving on the program committees of many conferences and workshops on these topics, and on the editorial boards of the IEEE Transactions on Knowledge and Data Engineering, Social Network Analysis and Data Mining Journal, International Journal of Data Mining and Bioinformatics, the journal on Current Proteomics, Advances in Bioinformatics, and Biomedicine and Biotechnology.