Abstract:

Informatics and computers have not yet become as pervasive in chemistry as they have in physics and biology. Drawing analogies from bioinformatics, key ingredients for progress in chemoinformatics are the availability of large, annotated databases of compounds and reactions, data structures and algorithms to efficiently search these databases, and computational methods to predict the physical, chemical, and biological properties of new compounds and reactions. We will describe the development of: (1) a large public database of compounds and reactions (ChemDB); (2) machine learning kernel methods to predict molecular properties; and (3) the applications of these methods to drug screening/design problems and the identification of new drug leads against a major disease. More broadly, we will discuss some of the challenges and opportunities for computer science, AI, and machine learning in chemistry.

Abstract:

This talk will present some general problem areas and solutions in two fields of applications of machine learning: bioinformatics and Geographic Information Systems (GIS). The bioinformatics arena is very broad and encompasses many problems such as gene finding in sequences, molecular pathway construction, protein structure prediction etc. We will outline our research on finding important keywords from the biomedical literature by statistical analysis and some natural language analysis. We have also incorporated ontologies such as UMLS (Unified Medical Language System) to determine relationships among biological and medical concepts. The primary goal of this work has been to interpret the long lists of genes that are derived in microarray experiments used to understand and treat diseases. We are able to cluster genes based on their functional similarity. We have also used lists of keywords as feature vectors to drive SVM models for a classification of literature. In particular, we have dealt with the classification of relevant literature for Public health at the CDC (Centers of Disease Control). We will briefly explain the discovery of biomarkers for cancer using a technique that combines SVM and gene ontology.