Data-mining algorithms for cheminformatics

Problem Statement

Our Chemical Compound Classification Framework

Feature generation and classification

FSG : topological features

AI-TREES: Generating All trees upto size k

Combination Features

Classifying Feature Vectors (Model learning)

Feature Generation

- How to assess clustering results?
  - diversity
  - tightness
  - domain expert
  - Better measures of similarity?

Docking

Challenges:
- Speed: number of conformations/min
- Accuracy: geometric and energetic
- Good scoring function
- Conformational changes during binding

Future Work

Future Work

Chemical compounds

Virtual Screening

Library design via clustering

Chemical Compound Classifier

Class = Active / Inactive

Test Chemical Compound(s)

Training Chemical Compounds

Chemical Compound Classification Framework

Our Chemical Compound Classification Framework

Build Classification Model

Transform Chemical Compounds

Feature Vectors

Results : Comparison of daylight and FSG

Results : Comparison of QSAR and Subdue

Results : Comparison of all methods (wilcoxon signed rank test used over 16 datasets)

People Involved:
Kevin DeRonne (deronne@cs.umn.edu)
Nikhil Wale (nwale@cs.umn.edu)
Ying Zhao (yzhao@cs.umn.edu)
Michihiro Kuramochi (kuram@cs.umn.edu)
Huzefa Rangwala (rangwala@cs.umn.edu)
Christopher Kaufman (kauffman@cs.umn.edu)
Professor George Karypis (karypis@cs.umn.edu)