Preliminary: Discriminative Subgraph Patterns

Given two sets of graphs, namely a positive set and a negative set, discriminative subgraph patterns are subgraph patterns that appear more frequently in one graph set than in the other. A user-specified function of frequencies is used to measure the discrimination power of patterns.

Challenges
- Subgraph pattern search space is huge
- No Apriori property: discrimination power is not monotonic to pattern size

Greedy Search fast-probe
We observe that discriminative patterns are more likely to be extended into more discriminative patterns.

We allow a pattern to be extended from multiple smaller patterns to reduce the risk of missing highly discriminative patterns.

Learning from Search History
- Search record: a sequence of discrimination scores of ancestor patterns
- Search history: all search records found in the preliminary greedy search, fast-probe

A prefix tree and a summary table are generated based on search history. The tree and table are used to estimate score upper-bounds in the branch-and-bound search.

Intuition and Workflow

Mining graphs
- Have a lot of graphs
- Extract useful patterns
- prediction

Mining search history of graph mining?
- search history composed of many subgraphs
- Can we do something here?
- Want to predict the upper-bound of a pattern

Greedy search fast-probe to collect history
- Generate prediction tree and table based on history
- Branch-and-bound search with history-based upper-bound estimation

Search record of pattern C-A-B-D: 0.5, 0.8, 1.0
Comparison with LEAP: Discrimination Scores of Top Patterns
- 16 SCOP (Structural Classification Of Proteins) protein families with 250 randomly selected outsiders as negative set
- Each protein family has 26-65 members
- On average, each graph has 250 nodes and 2700 edges

Comparison with GAIA and COM: Classification Accuracy
- 5-fold cross validation
- Normalized accuracy = (sensitivity + specificity) / 2

Current Project Members
Wei Wang, faculty member
Ning Jin, graduate research assistant

For More Information
Wei Wang
Professor
919-962-1744
weiwang@cs.unc.edu