Auto Query Steering for Interactive Data Exploration Applications

Kyriaki Dimitriadou  
Brandeis University  
kiki@cs.brandeis.edu

Olga Papaemmanouil  
Brandeis University  
olga@cs.brandeis.edu

Yanlei Diao  
UMass Amherst  
yanlei@cs.umass.edu

Interactive Data Exploration

• Human-in-the-loop applications that search enormous datasets to discover interesting information.

Medical Applications

Scientific Applications

Financial Applications

Research Challenges

• Effective Data Exploration:
  - Given a classification model identify new training samples to improve the accuracy of the system.

• Efficient Sample Acquisition:
  - Reduce the total cost of the query steering process:
    → sampling overhead.
    → number of steering iterations.

Preliminary Results

User effort decreased by 58-68% depending on the initial sampling technique.

Classification performance is higher than 96%.

Our Data Exploration Approach

• Use decision trees to build classification model.
• Exploit node split rules to identify exploration areas.

Objective: Automatically capture the user’s interest and steer him towards interesting data areas.

Automatic Query Steering Framework

User relevance feedback

User Profile Learning

Query Formulation

Space Exploration

Sample Extraction

Samples

Database

Ongoing Work – Sample Acquisition

• Optimizations on exploratory workloads:
  - Avoiding sparse regions using histograms.
  - Across-iteration optimization.

• Profile driven optimizations:
  - Leverage histories of user interactions (exploration profiles) :
    → predict exploration trajectories.
    → pre-fetch samples.
    → identify interesting attributes.