

Algorithms for Frequent Sets on Streams  
Ed Robertson, edrbtsn@cs.indiana.edu

- Motivation – Association Rules and Frequent Sets
- Motivation – Data Streams
- Frequent Sets in Streams – Current Status
- Proposed project
- Other projects

## Motivation - Association Rules and Frequent Sets

- Example: market basket analysis
  - ◇ data from supermarket scanners
  - ◇ want to know when sales of products are associated
  - ◇ men who buy diapers commonly buy beer
- Example - Amazon recommender
  - ◇ people who have purchased [the same set of books you have] also buy . . .
- Requires discovering *frequent sets*
  - ◇ set of items occurs in at least  $x\%$  of input baskets
- Problem is combinatorics
  - ◇ many possible sets need to be considered
- Off-line case well considered: *apriori* algorithm prunes sets

## Motivation - Data Streams

- Stream model
  - ◇ input data flow does not terminate
  - ◇ must answer queries about input thus far received
- Example - monitoring network traffic
- Problem is efficiency
  - ◇ must keep up with flow
  - ◇ no or minimal secondary storage use
  - ◇ guaranteed approximations, “time windows” typical work-arounds

## Frequent Sets in Streams

### Current Status

- Would like frequent sets for past intervals as well as everything up to now
- In pure form, combinatoric problem even worse!
- Existing algorithm:
  - ◇ guaranteed approximation
  - ◇ past intervals within scaled windows

example: last 300 seconds, preceding 10 minutes, 3 preceding quarter hours, preceding days

## Frequent Sets in Streams Proposed Project

- Many more experiments
  - ◇ what are reasonable data set characteristics?
- Reimplementation for performance
- Improve algorithms
  - ◇ amortized updates

## Other Projects

- Extending EER tool for metadata and constraints
  - ◇ can graphical metadata model be translated unambiguously to relational schema?
  - ◇ can constraints be derived from graphical representations?
  - ◇ are graphical models easy for users to understand?
  - ◇ can graphical models be implemented with usable operations
- Enterprise-level modeling
  - ◇ understanding how modelers build high-level (meta-meta-) models
  - ◇ formalizing the same
  - ◇ participating in development of international standards for model-based enterprise interoperability