FP-TREE MINING

KDD - Project Presentation

Maria Cutumisu

New approach

- FP-tree vs. Apriori
- Basic concepts of FP-tree mining:
  - – FP-tree
  - – FP-growth
  - – Divide & conquer

Strategy

- First scan - item counts, header list.
- Second scan - sort transactions in descending order, create FP-tree
- Tree traversal-extract conditional pattern base for each element of the header list, beginning with the last one
- Generate frequent patterns out of the conditional patterns

Command Line Parameters

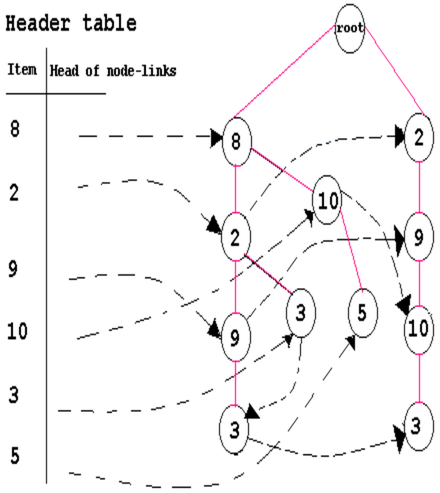
- The following arguments must be provided in the command line:
  - – a synthetically generated transaction database: every transaction consists of a TID and a list of items
  - – a threshold
  - – a file where the frequent patterns will be stored

Implementation

- All algorithms are implemented in C
- Data structures used:
  - – struct node: for the tree
  - – struct list: for the header table
  - – struct merge: for the conditional patterns
  - – struct tran: for transactions

Example

TID	Items	Ordered
1	2, 6, 9, 7, 8	8, 2, 9
2	5, 10, 8	8, 10, 5
3	4, 5, 10, 8	8, 10, 5
4	2, 3, 8, 1	8, 2, 3
5	2, 3, 9, 10	2, 9, 10, 3
6	2, 3, 8, 9	8, 2, 9, 3



Conclusions

- FP-tree based mining method
  - – flexible
  - – can be further improved
- Tested on various data sets
  - – between 100 and 100K transactions