Mining Association Rules: Anti-Skew Algorithms

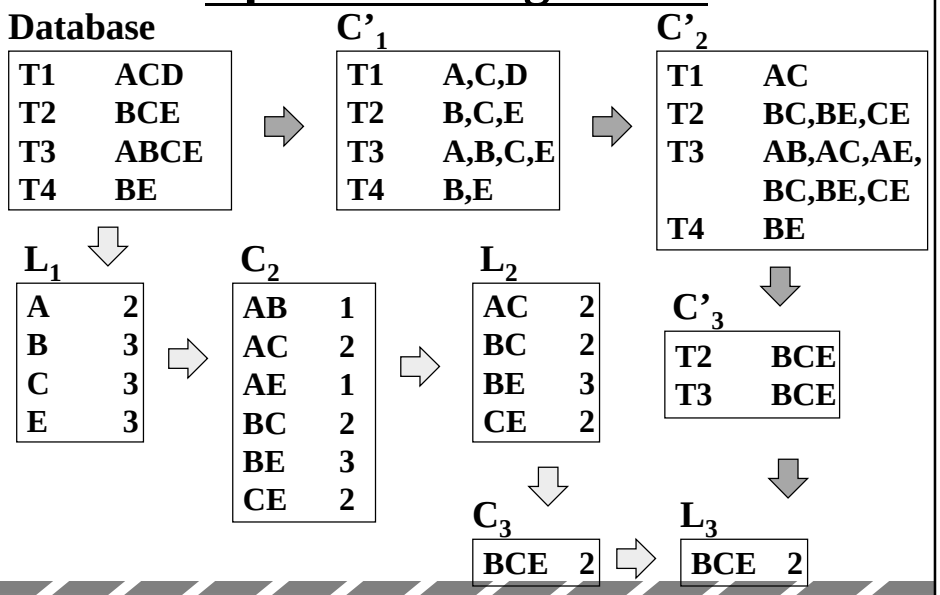
Jun-Lin Lin and Margaret H. Dunham
Southern Methodist University, USA

International Conference on Data Engineering, pp. 486-493, 1998.

Preliminary

- **R. Agraval, T. Imielinski, and A. Swami**
  Mining Association Rules between Sets of Items in Large Databases
  ACM-SIGMOD Conference, pp.207-216, 1993.
  - **brute force**
- **R. Agraval and R. Srikant**
  Fast Algorithms for Mining Association Rules
  VLDB Conference, pp.487-499, 1994.
  - **level-wise**
    - *a lot of variations*

AprioriTid Algorithm



Preliminary

- **A. Savasere, E. Omiecinski and S. Navathe**

  An Efficient Algorithm for Mining Association Rules in Large Databases

  VLDB Conference, pp.432-444, 1995.

  - **partitioning**
    - *a lot of variations*

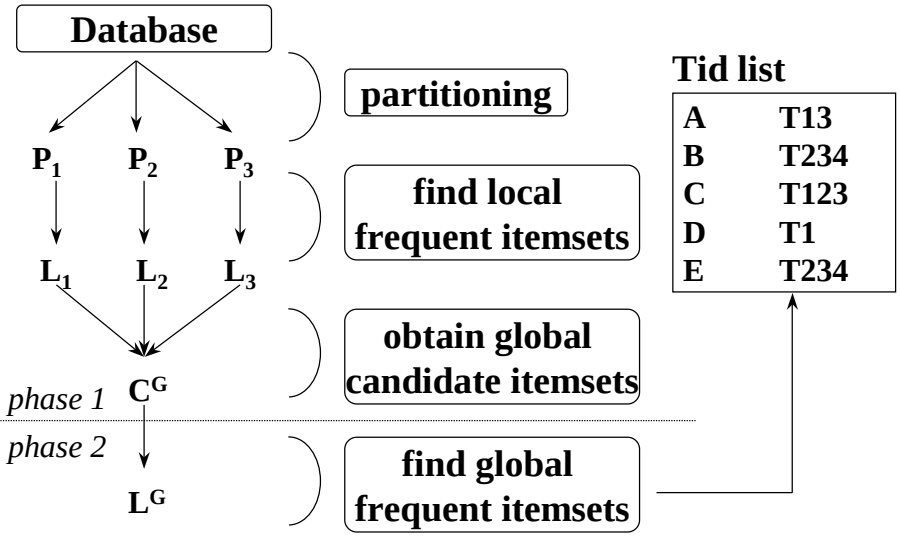
- **H. Toivonen**

  Sampling Large Databases for Association Rules

  VLDB Conference, pp.134-145, 1996.

  - **sampling**

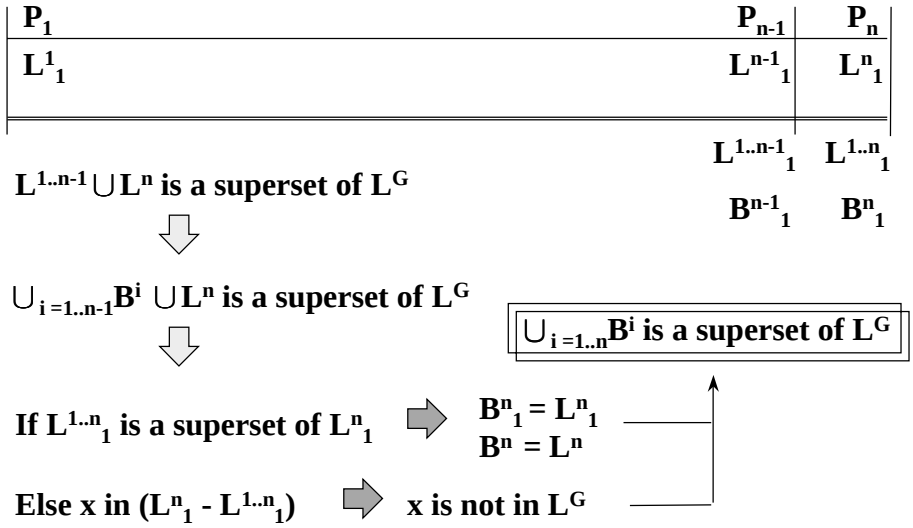
SON Algorithm



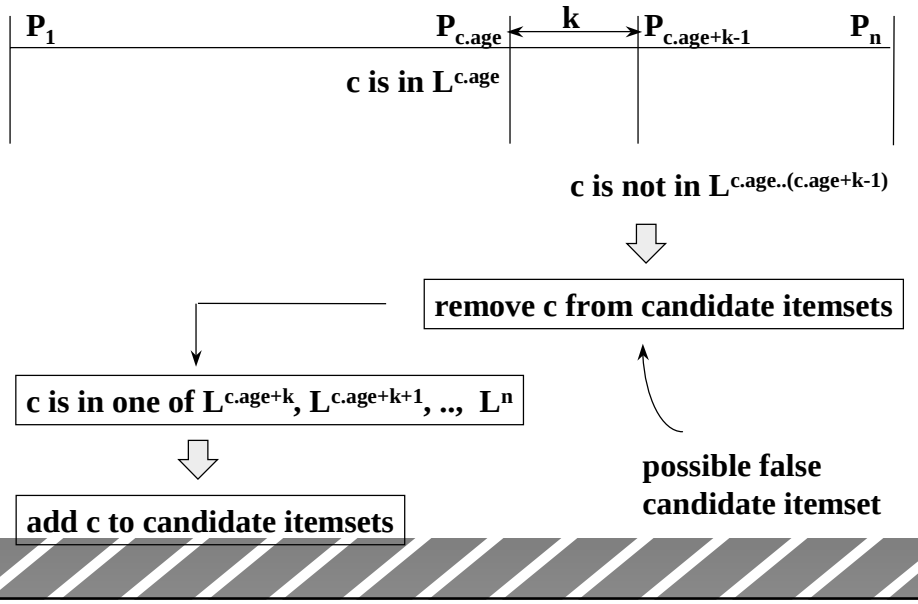
Introduction

- **Goal**
  - **use cumulative count of each candidate itemset to archieve the illusion of a large partition**
  - **employ prior knowledge collected during mining process to reduce the number of candidate itemsets and identify false candidate itemsets at an earlier stage**
- **Idea**
  - **better local frequent itemsets $B^i_1 = L^i_1 \cap L^{1..i}_1$**
    - $C^G = \cup_{i=1..n} B^i \rightarrow C^G$ *is a superset of* L^G
  - **each candidate itemset c: (c.age, c.count)**

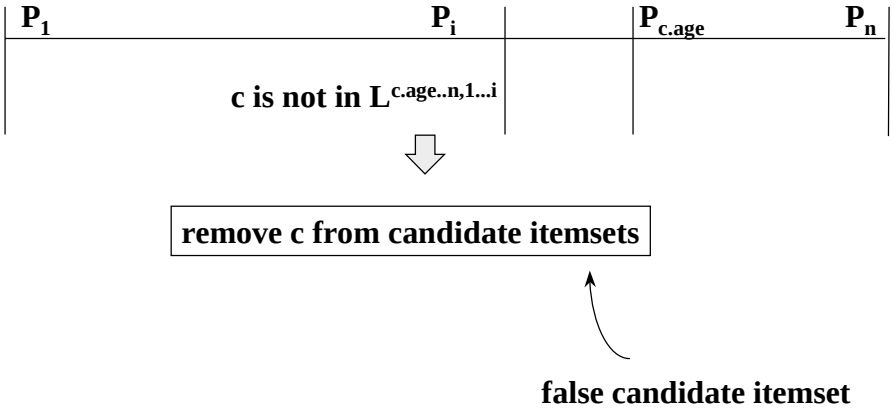
Early Local Pruning



First Global Anti-Skew



Second Global Anti-Skew



Conclusion

- Improvement on partitioning approach
  - early local pruning
  - first global anti-skew
  - second global anti-skew
- Merge partitioning and sampling approaches
  - random sampling on first few partitions
- Issue
  - How to properly integrate partitioning and sampling approaches
    - *sampling on each larger partition*