MINING LONG SHARABLE PATTERNS IN TRAJECTORIES OF MOVING OBJECTS

BY

GYŐZŐ GIDOFALVI AND TORBEN BACH PEDERSEN

Presented by Stig Jørgensen

Publication information

- Journal: GeoInformatica
- Published by Springer Netherlands
- Published online – December 2007
- Two-versions

Outline

- What is a Long, Sharable pattern?
- Identification of trips
- Frequent Itemset Mining
- Approahces
- INFATI dataset and discoveries
- Performance
- Related work
- Relation to our project
- Comments

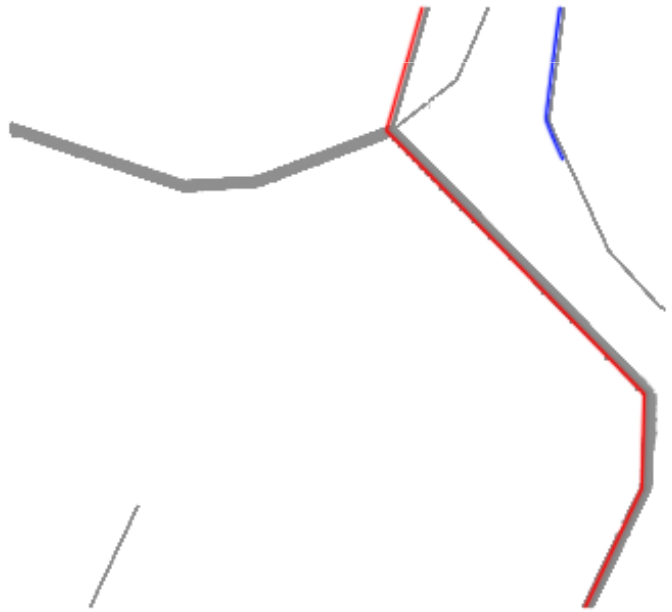
Long Sharable Patterns

- What is a long, sharable pattern?
  - Similarities
    - Time
    - Location
  - Example
    - Carpooling

Identification of patterns

Identify trips

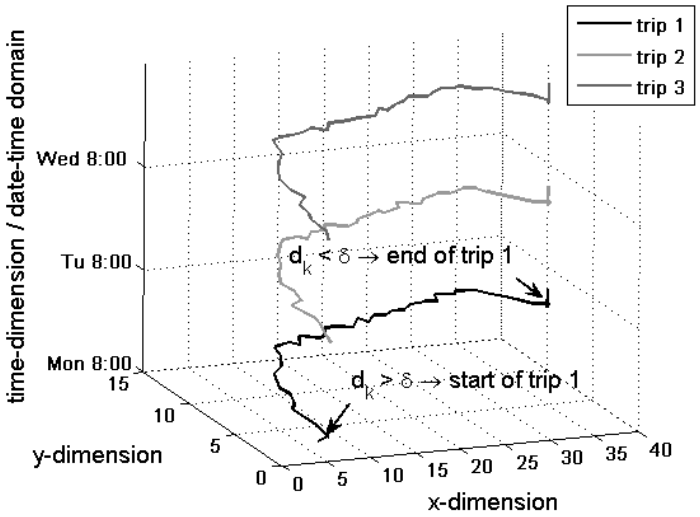
- Filtering out non-trips
- A trip must be longer than a displacement



Identification of patterns

Projection of temporal dimension

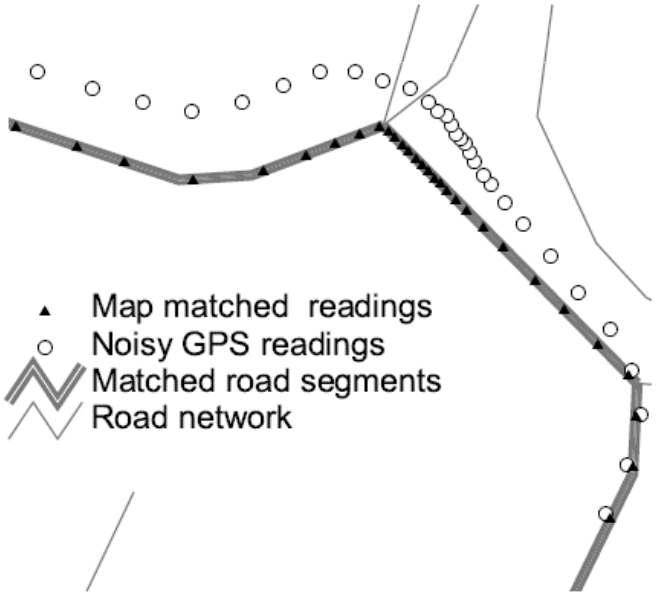
- Map date time to recurring events:
  - Time-of-day
  - Day-of-month
  - Day-of-week

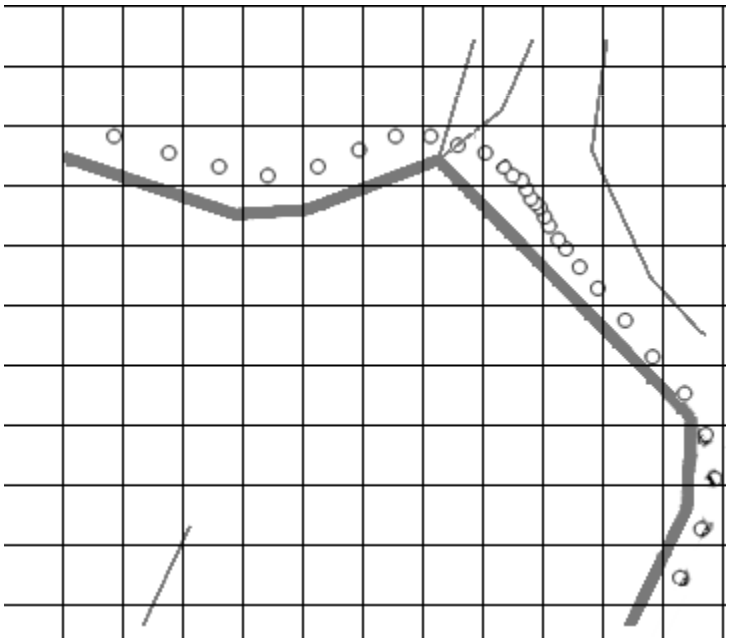


Identification of patterns

Substitute noisy GPS readings

- Road network based generalization
- Region-based generalization





Frequent Itemset Mining

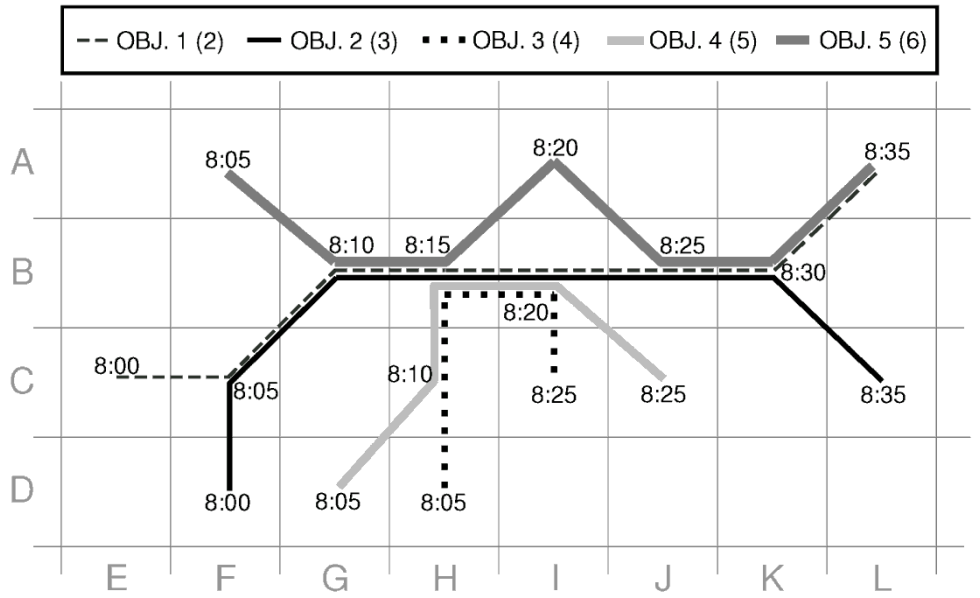
- Has to be modified to support Carpooling:
  - A frequent item set has to be long
  - A frequent item set has to shared by an amount of travelers.
- Our data can be converted a format <oid, tid, s>, that is required for Frequent Itemset Mining

Approahces

- Naïve approach
  - Finding sub-trajectories through k-way self-joins.
  - Iterative manner
  - Running time
    - Worst case: Exponential

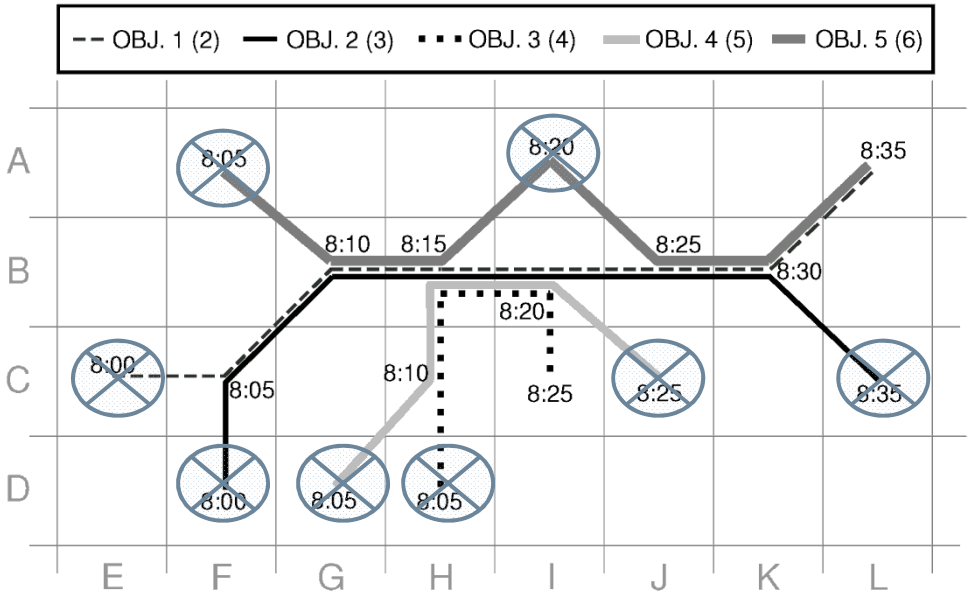
Projection-based LSP mining

- Temporal dimension: 5 Minutes
- Spatial dimension: Square cells
- 5 step iterative approach



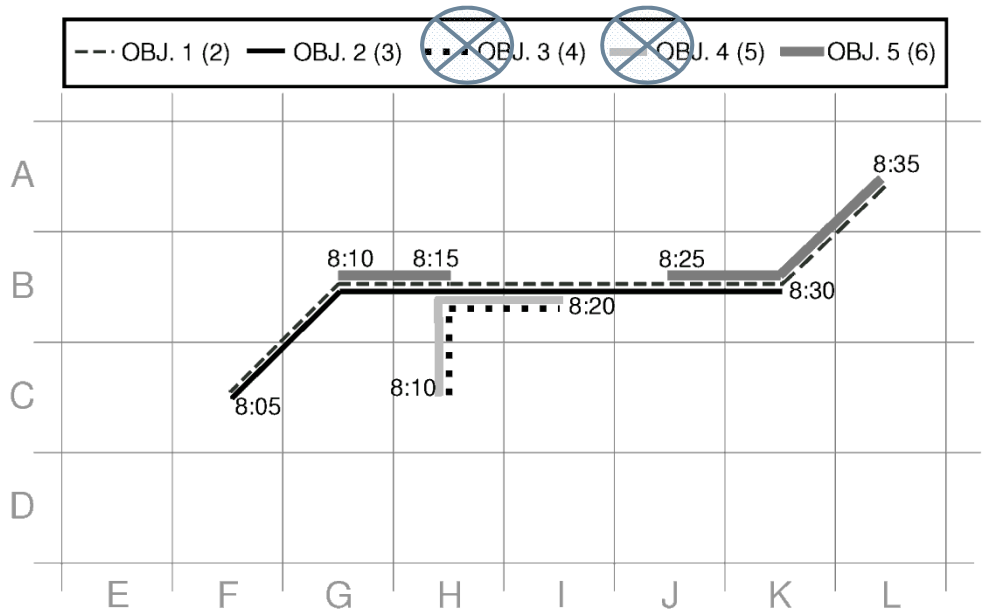
Step 1 - Filter infrequent items

- An item is frequent if:
  - The amount of transactions that contain an item is >= 4
  - The amount of unique objects associated with those transactions is >= 2



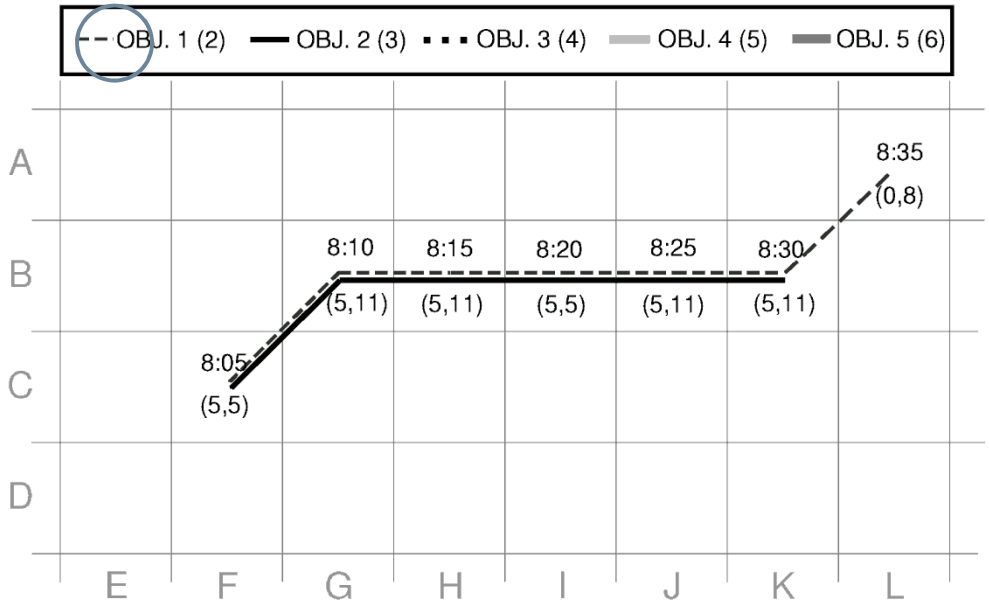
Step 2 – Filter short transactions

- A trajectories is short:
  - If the length of the trajectories >= 4



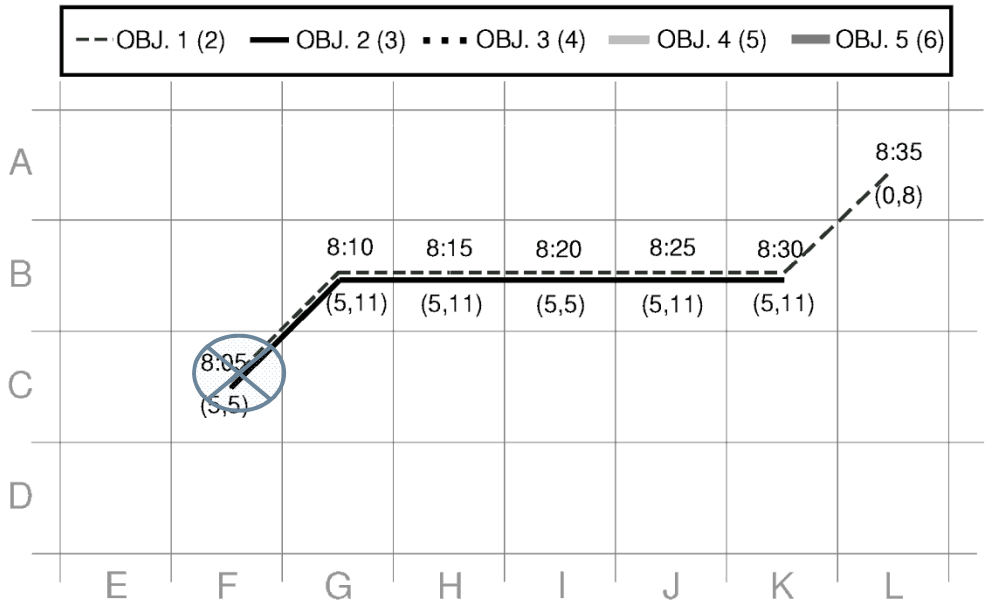
Step 3 & 4 – Project and select

- Takes one out one element and projects it to another DB.
- Selects the most frequent itemset



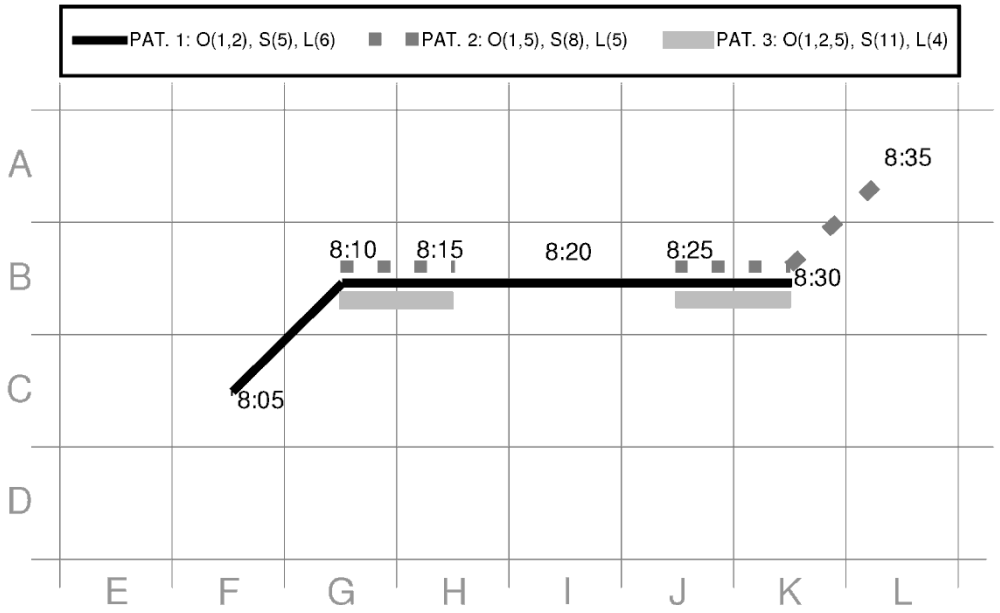
Step 5 - Deletion

- Deletes unnecessary items from predecessor DB.



Pattern discovery and deletion phase

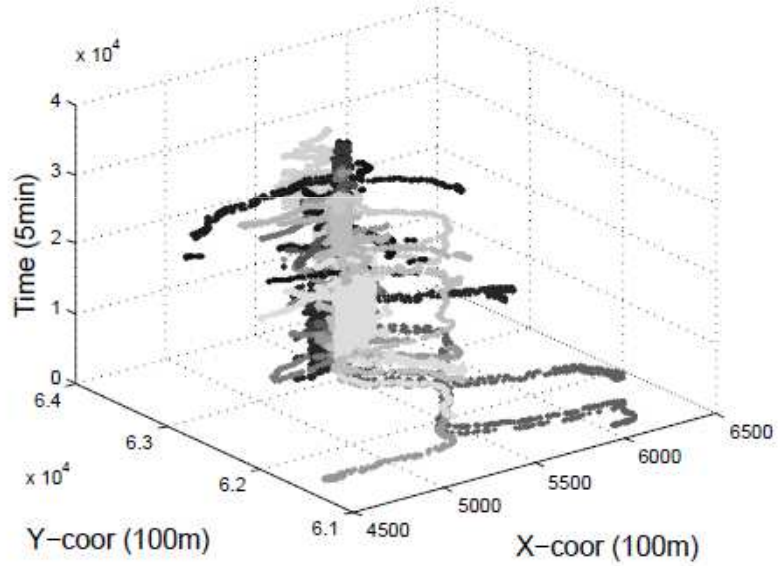
- Iterative process over step 3 to 5.

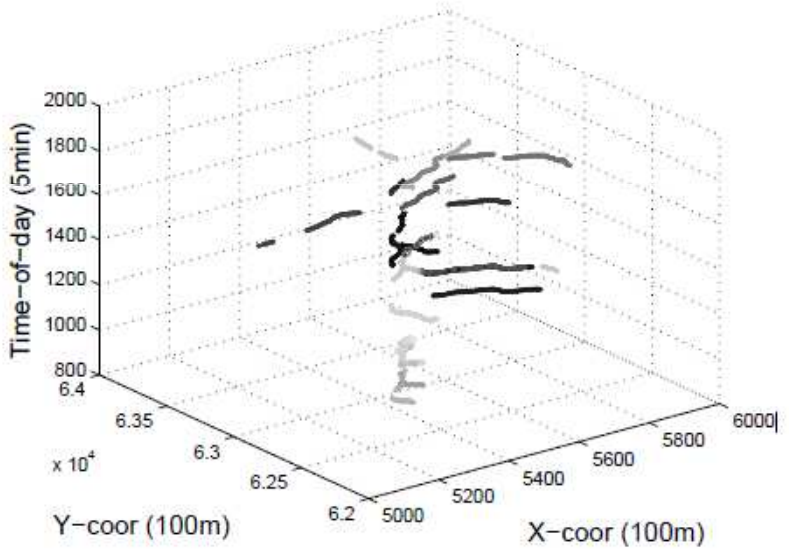


INFATI dataset

- Real dataset from Aalborg
- 20 unique test cars
- Transformation from noisy readings into 100 *100 m 5 minute spatio-temporal regions.
- ~ 200.000 unique items in 3.699 transactions

LSP Discovery in INFATI



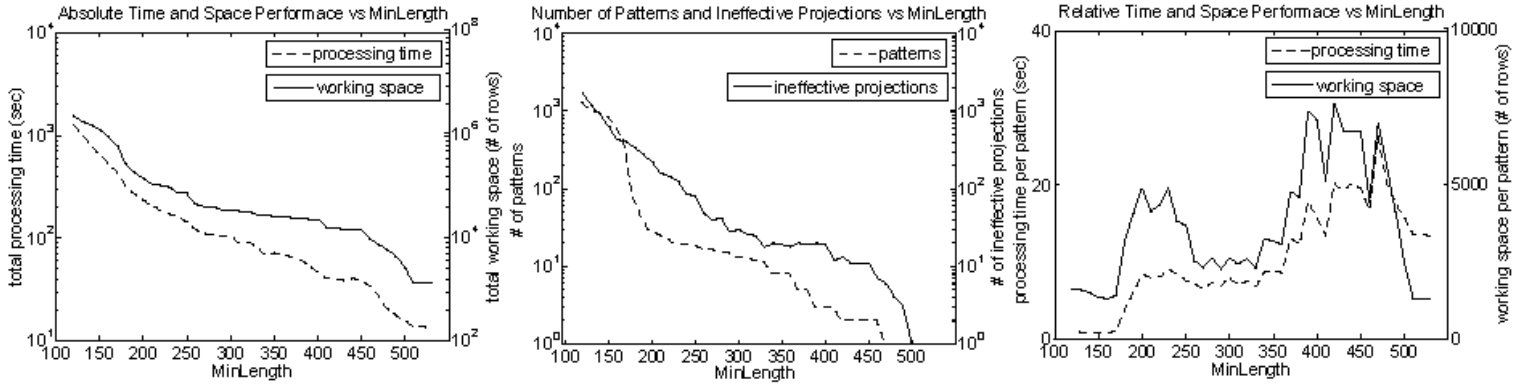


Alternative modeling

- Macro patterns
  - Works on origin and destination.
  - Requires modification to the Distance concept.
- Hybrid model using both Macro and Micro-Patterns
  - Scales better
  - Does not find all local LSPs

Performance – minLength

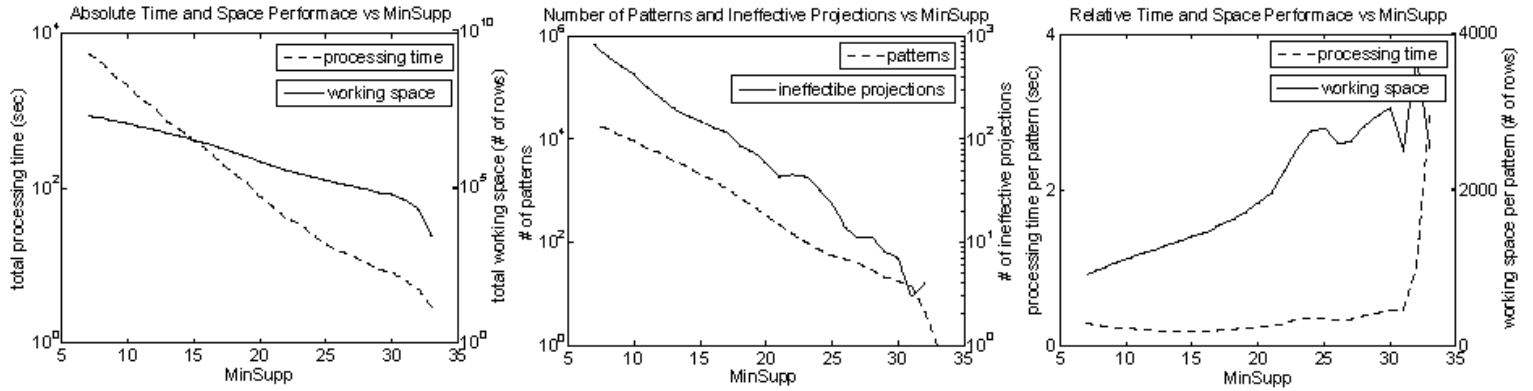
- Running time and space increase exponentially as minLength decrease.
- Average running time decrease lineary as minLength decrease.



(a) Absolute time and space (b) Number of patterns (c) Relative time and space

Performance - minSupp

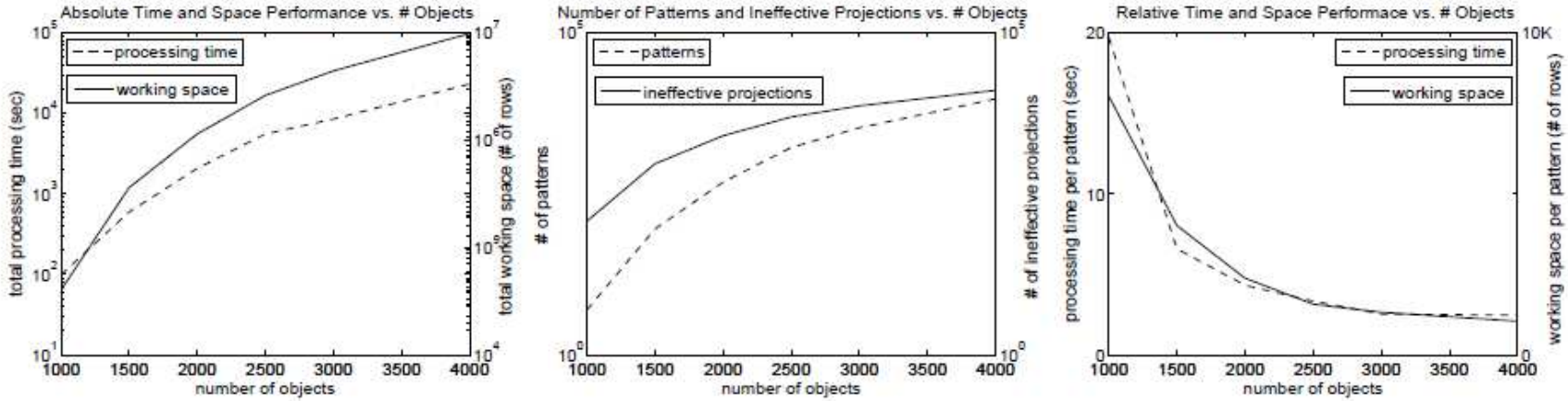
- Running time and space increase exponentially as minLength decrease.
- Average running time decrease lineary as minLength decrease.



(a) Absolute time and space (b) Number of patterns (c) Relative time and space

Performance - scalability

- As the patterns increase linearly the amount of patterns increase sub–exponentially.
- Amount of time/space required per pattern decrease to a constant.



(a) Absolute time and space (b) Number of patterns (c) Relative time and space

Summary

- LSP mining method is effective and robust
- Scale up test
  - Running time and space required scales exponentially with input size.
- Macro modeling
  - Effective, yet insensitive to user–defined parameter settings.
- Hybrid model
  - Able to find most local LSPs effectively.

Related work

- Frequent Itemset Mining
- All frequent item sets are too large
  - Closed Frequent Itemsets (CFI)
  - Compression

Relation to our project

- Article picked on interest.
- Product status
  - Airport Case
- Flow analysis
  - Convert our data into FIM accepted data
    - <Signal strength, time>
    - <oid, tid, s>

Comments about article

- Reads good
  - Covers a lot of areas with various detail level.
- Ordering of article
  - Introduction of micro LSP mining method before introducing macro and hybrid model.
- Scope of the article
  - More focus on hybrid model.

Thank you for listening

- Questions?
- Comments?