PATTERN SEARCH EXPLORATION GUIDED BY MINIMUM DESCRIPTION LENGTH PRINCIPLE

Tatiana Makhalova, Sergei O. Kuznetsov, Amedeo Napoli

PROBLEM STATEMENT

Data: transactional dataset over *M* attributes
Patterns: regular itemsets.
Objective: to discover a small set of non-redundant and interesting patterns that describe together a large portion of data and that can be easily interpreted.
Challenges:

- the search space of the size $2^{2^{|M|}}$
- ``interestingness'' is subjective measure.

APPROACHES TO PM

STATIC "TELL ME WHAT I ASK FOR"

- Idea: mining under non-changeable assumptions about interestingness (interesting measures) [Geng and Hamilton, 2006]
- Example: in frequency-based Pattern Mining (PM), one assumes that all the patterns with a frequency greater than a minimum threshold are interesting.
- Drawbacks: all patterns are very similar and the choice of interestingness measure not always can be justified.

DYNAMIC ``TELL ME WHAT I NEED TO KNOW"

- Idea: setting initial knowledge on dataset and gradually extend a pattern set by selecting the most "unexpected" patterns w.r.t. the current pattern set and knowledge. The knowledge is progressively updating together with the pattern set.
- Example: **Krimp** - an **MDL-based** greedy covering by pre-computed patterns.

MDL in PATTERN MINING. BASICS

"THE BEST PATTERN SET SHOULD COMPRESS DATA AT BEST"

- Pattern set is represented as a **code table** *CT*.
- **Cover function** *cover(X,CT)* produces a set of disjoint itemsets from *CT* that fully cover all attributes from *X*.
- **Probability distribution** on pattern set.
  Usage returns the number of times X is used in covering of D, i.e.,
  $$usage(X) = |\{g \in G \mid X \in cover(\{g\}', CT)\}|,$$
  with usage(X) ≤ frequency(X).
  Usage-based probability estimates:
  $$P(X) = \frac{usage(X)}{\sum_{X^* \in CT} usage(X^*)}.$$
- **Optimal code** (the Shannon code) L(code(X)) = −log(P(X)), i.e. shortest lengths are assigned to most commonly used patterns.
- **Objective:** minimise the description length
  $$L(D, CT) = L(CT \mid D) + L(D \mid CT),$$
  where the length of the dataset D encoded by this CT is
  $L(D \mid CT) = \sum_{X \in CT} usage(X)L(code(X))$.
  The length of CT is $L(CT \mid D) = \sum_{X \in CT} L(code(X))$.

CLOSED ITEMSETS

Closed itemsets are presented in the framework of FCA [Ganter and Wille, 1999].

A **formal context** is a triple D = (G, M, I), where $G = \{g_1, g_2, \dots, g_n\}$ is a set of objects, $M = \{m_1, m_2, \dots, m_k\}$ is a set of attributes and $I \subseteq G \times M$ is an incidence relation, i.e. (g, m) ∈ I if the object g has the attribute m. The **derivation operators** (·)′ are defined for Y ⊆ G and X ⊆M as follows:

$$Y' = \{ m \in M \mid \forall g \in Y : gIm\}, \quad X' = \{ g \in G \mid \forall m \in X : gIm\}.$$

Y′ is the set of attributes common to all objects of Y and X′ is the set of objects sharing all attributes of X. An object g is said to contain a pattern (set of items) X ⊆ M if X ⊆ {g}′. The double application of (·)′ is a closure operator. A ***closed set*** X is such that X = X′′ = (X′)′. There does not exist another closed set Z such that X ⊆ Z ⊆ X′′. A ***(formal) concept*** is a pair (Y, X), where Y ⊆ G, X ⊆ M and Y′ = X, X′ = Y (then X = X′′ and Y = Y′′).

STATE-OF-THE-ART

DYNAMIC PM. KRIMP in few words [Vreeken *et al.*, 2011]

- Patterns are chosen form a **candidate set**, e.g., a set of frequent patterns.
- **Order of candidates**: length, frequency, lexicographical.
- Patterns are being added in CT gradually using a greedy strategy.
- A pattern is accepted to the code table if it minimise the total length L(D, CT).

WHAT'S WRONG?

- Too many patterns.
- The model is affected by heuristics (disjoint-cover constraints and usage-based probability estimates)

KRIMP. AN EXAMPLE

Candidate set
ABC
BDE
BD
DE
CD

ST	P(X)
A	4/21
B	5/21
C	4/21
D	4/21
E	4/21

Step 1 ✓

CT	P(X)
ABC	3/15
D	4/15
E	4/15
B	2/15
A	1/15
C	1/15

Step 2 ✓

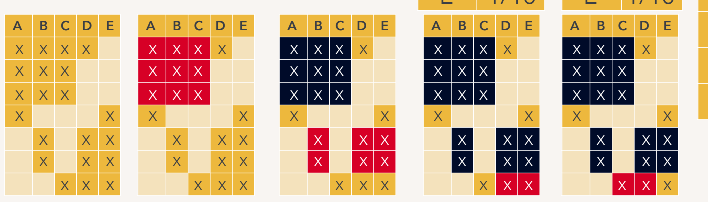
CT	P(X)
ABC	3/11
BDE	2/11
D	2/11
E	2/11
A	1/11
C	1/11

Step 3 ✗

CT	P(X)
AB	3/10
BDE	2/10
DE	1/10
A	1/10
C	1/10
D	1/10
E	1/10

Step 4 ✗

CT	P(X)
AB	3/10
BDE	2/10
CD	1/10
A	1/10
C	1/10
D	1/10
E	1/10



Frequent closed patterns ordered by length, frequency, lexicographically.

Step 1 ✓ Step 2 ✓ Step 3 ✗ Step 4 ✗

CT	P(X)
ABC	3/11
BDE	2/11
D	2/11
E	2/11
A	1/11
C	1/11

ABC, BDE are MDL-optimal patterns.

PROPOSED APPROACH

PATTERN SPACE EXPLORATION

- From frequent to 2-closed itemsets.

Pros: form $O(2^{|M|})$ to $O(|M|^2)$; parameter-free; additional compression.

- From usage-based to frequency-based estimates.

Pros: less dependent on heuristics, capture structure underlying the data rather than side effects from heuristics.

- Explore patterns space efficiently, i.e., using projection, closed itemsets, breadth-first search guided by MDL objective, "partial forgetting" information about mined structure.

SUMMARY

- Substantial shrinkage of the number of attribute to consider (projection size) after the 1st iteration.
- Fast convergence.
- Meaningful interpretation.
- Simple enumeration techniques, a quadratic-sized space to explore.
- Pruning pattern space with projections and an MDL-based criterion.

PROPOSED APPROACH. AN EXAMPLE

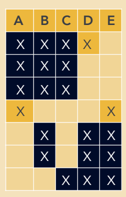
PATTERNS TO EXPLORE

2-closed itemsets	Derived from
ABC	AB, AC, BC
ABCD	AD
AE	AE
BD	BD
BDE	BE
CD	CD
DE	DE
CDE	CE

Induced by a pair of attributes, ordered by length, frequency, lexicographically

MDL-OPTIMAL

2-closed itemsets
ABC
ABCD
AE

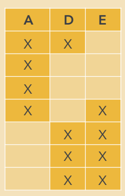


Step 1

"KeepItSimple" projection onto original dataset

projection using uncovered attributes

"ExplainInDetail" projection onto dataset of uncovered relations



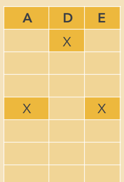
PATTERN TO EXPLORE

2-closed itemsets	Derived from
DE	DE
AD	AD
AE	AE

MDL-OPTIMAL

2-closed itemsets

PATTERN SPACE
ABC
BDE
CDE



2-closed itemsets	Derived from
AE	AE

2-closed itemsets
AE

PATTERN SPACE
ABC
BDE
CDE
AE

Step2

That is an analogue of a ``candidate set" in Krimp

REFERENCES

[Ganter and Wille, 1999] B Ganter and R Wille. Formal concept analysis: Logical foundations. Springer Verlag Berlin, RFA, 1999.

[Geng and Hamilton, 2006] Liqiang Geng and Howard J Hamilton. Interestingness measures for data mining: A survey. ACM Computing Surveys (CSUR), 38(3):9, 2006.

[Vreeken et al., 2011] Jilles Vreeken, Matthijs Van Leeuwen, and Arno Siebes. Krimp: mining itemsets that compress. DM and KD, 23(1):169–214, 2011.

Future work: numerical and graph pattern mining.