On Computing Minimal Generators in Multi-Relational Data Mining with respect to θ-Subsumption

Noriaki Nishio
Atsuko Mutoh
Nobuhiro Inuzuka

Nagoya Institute of Technology, JAPAN

Tue 18 Sept, 2012

INTRODUCTION

In pattern mining process

a lot of patterns including redundant ones are enumerated.

[McGarry, K., Knowledge Eng. Review 20, 2005]
[Silberschatz, A., IEEE Trans. Knowl. Data Eng. 8, 1996]

To generate only important patterns

- Closed patterns [Pasquier, N., ICDT, 1999]
- Minimal Generators [Bastide, Y., Computational Logic, 2000]

Multi-Relational Data Mining (MRDM)

- Frequent patterns which involve multiple tables

Not much studies on minimal generators in MRDM

- We consider the minimal generators in MRDM.
- Reduce those redundancy.

CONTENS OF MY PRESENTATION

1. Introduction of backgrounds
   - Formal Concept Analysis
2. Previous studies on pattern reduction
   - Closed patterns
   - Minimal Generators
3. Proposal of a reducing method in MRDM patterns
4. Experimental results and conclusions

Formal Concept Analysis (FCA)

Propositional FCA

- Analyzes object-attribute data.

Formal Context: (G, M, I)

- G: Objects
- M: Attributes
- I: Relation between G and M

	fruit	red	green
apple	x	x	
banana	x		
carrot		x	

Relational FCA

- Analyzes multi-relational tables including a key table.

Formal Relational Context: (G, M, I)

- G: Instances of key relation
- M: Relational patterns
  e.g.) F1 = food(A) ← fruit(A).
- I: An instance satisfies a formula

	F1	F2	F3
food(apple)	x	x	
food(banana)	x		
food(carrot)		x	

CLOSED PATTERNS AND MINIMAL GENERATORS (1)

- **Closed pattern:** maximal pattern in an equivalence class in the sense of producing same extent.
- **Minimal generator (mingen):** minimal pattern in an equivalence class.

	1	2	3	4	5	6	7	8	9
t1	x	x			x	x	x		x
t2		x	x	x	x				
t3	x	x					x	x	x
t4	x						x		x
t5		x					x		x
t6		x							

CLOSED PATTERNS AND MINIMAL GENERATORS (1)

- **Closed pattern:** maximal pattern in an equivalence class in the sense of producing same extent.
- **Minimal generator (mingen):** minimal pattern in an equivalence class.

	1	2	3	4	5	6	7	8	9
t1	x	x			x	x	x		x
t2		x	x	x	x				
t3	x	x					x	x	x
t4	x						x		x
t5		x					x		x
t6		x							

CLOSED PATTERNS AND MINIMAL GENERAOTRS (2)

Closed patterns are important by two properties.

Completeness: [Pasquier, N., ICDT, 1999]

- The set C of frequent closed patterns have the complete information on the set F of all frequent patterns and their frequencies.

Compactness: [Mannila, H., KDD, 1996]

- Frequent closed patterns are possibly exponentially fewer than |F|.

Mingens are important by their nature.

- Mingens are a **core** patterns in many contexts, e.g., database design (key sets) and pattern mining (minimal LHS).

SSMG: [Dong, G., 2005]

- The mingens which redundant ones are removed.

PROPOSE <NON-REDUNDANT MINGENS> IN MRDM

Propsoe **Logical Minimal Generator (LMG)**

which is non-redundant mingen in the logical sense.

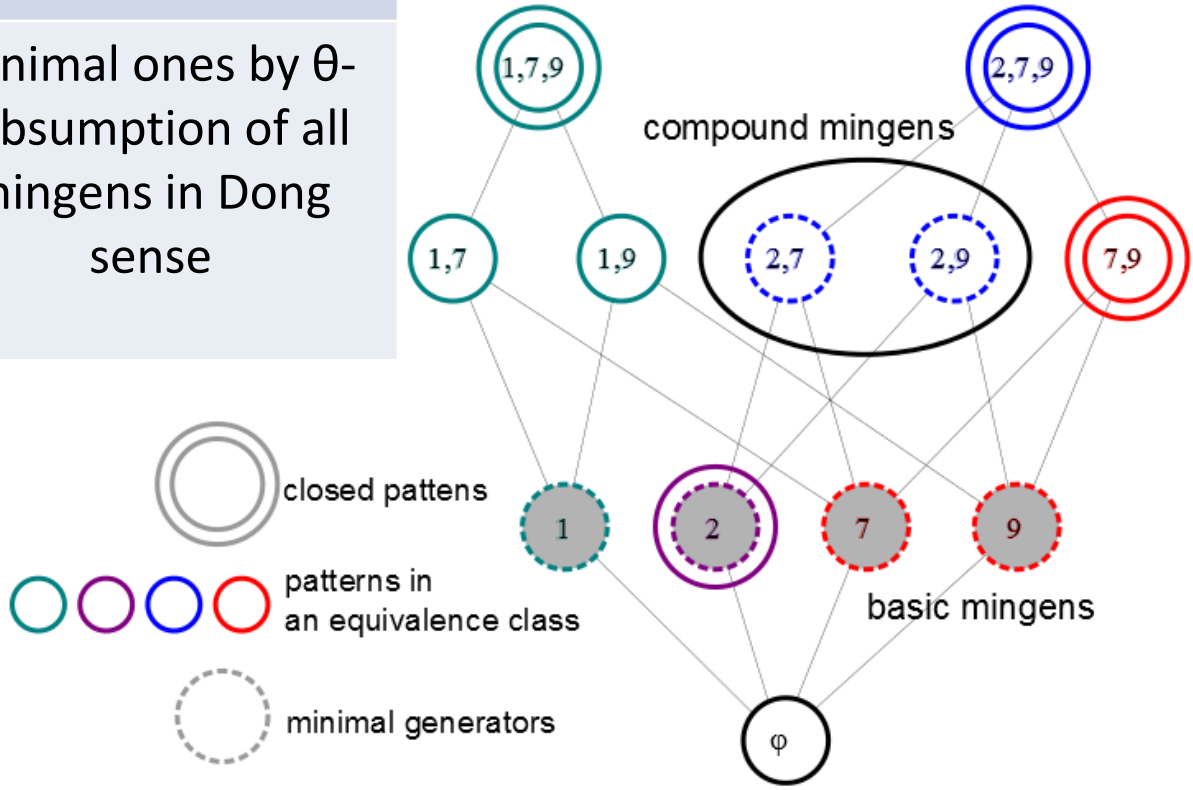
Keys of our proposal

- Application SSMG to relational patterns
  - Propose a formal context for relational patterns.
  - Then enanble to apply SSMG to relational patterns.
- Selection mingens by θ-subsumption order
  - Relational patterns have an independence relation.
  - Pattern reduction based on θ-subsumption order.

SSMG [DONG, 2005] AND LMG [OUR PROPOSAL]

SSMG and LMG select representative mingens of all mingens.

	SSMG	LMG
Basic mingens	selects by users	minimal ones by θ-subsumption order
compound mingens	selects only mingens determined by different SSMG	minimal ones by θ-subsumption of all mingens in Dong sense



Experimental Results and Conclusions

In experiments

- LMG reduce redundancy of relational patterns.
- But the current implementation is not scalable.

Conclusions

- Reviewed the importance of mingens from aspect of MRDM.
- Proposed a concept of LMG and a procedure to compute them.
- LMG are important class of mingens which is minimal in the logical sense.
- We need further study on scalability.