A Graph-based Approach to Mining Inter-transaction Association Rules

Chung-Wen Cho, Yi-Hung Wu, Jane Liu, and
Arbee L. P. Chen

Department of Computer Science
National Tsing Hung University

Outline

- Motivation
- A classification of related problems and their transformations
- The graph-based approach to mining inter-transaction association rules
- Conclusion

Motivation

- Recent advances in data mining have been broadly studied. Many problems and solutions have been proposed individually, but the relationships among them are not clear
- mining inter-transaction association rules is new and challenging

A classification of Related Problems

Table 1: A summary of the problem classification

Type	Data Format	Pattern	Count-by	Results	Applications
1	Transactions	Intra-transaction	Transactions	✓ Classical association rules	✓ Marketing
2	Sequences	Inter-transaction	Sequences	✓ Sequential patterns ✓ Traversal patterns	✓ Marketing ✓ Web browsing ✓ Music analysis
3	A sequence	Inter-transaction	Segments	✓ Repeating patterns ✓ Frequent episodes ✓ Partial periodical patterns ✓ Inter-transaction association rules ✓ Cyclic/Calendric association rules	✓ Marketing ✓ Music analysis ✓ Stock analysis ✓ Event search

We classify the related problems into three types according to three attributes: Data format, Pattern, and Count-by.

Problem Transformations

- The solutions to one problem can be applied to the others if the problem transformations are well defined.
  - intra-type transformation
    - transformations between the problems of the same types
  - inter-type transformation
    - transformations between the problems of different types

Intra-Type Transformation

- Type 2
  - Mining traversal patterns → Mining sequential patterns

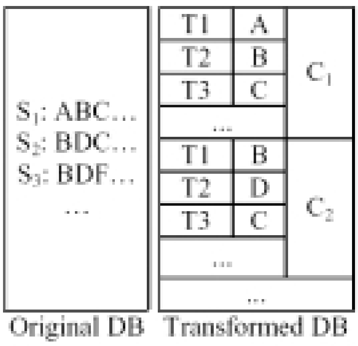


Figure 1: Navigation paths → Customer sequences

Inter-Type Transformation

- Type 3 → Type 2
  - Mining inter-transaction association rules → Mining sequential patterns

SID	Window
100	a(abc)(bc)
200	(abc)(bc)a
300	(bc)a
400	a
…	

Original DB

CID	Sequence
100	$A_0(A_1B_1C_1)(B_2C_2)$
200	$(A_0B_0C_0)(B_1C_1)A_2$
300	$(B_0C_0)A_1$
400	A_0
…	

Transformed DB

Figure 2: Sliding windows → Customer sequences

Mining Inter-transaction Association Rules (1/2)

TID	Transaction
T1	A, B, D
T2	A, C, E
T3	B, C, E
T4	B, D, E
T5	A
T6	B, C

Figure 5: An example database

Window	TID	Transaction
1	T1 T2 T3	A, B, D A, C, E B, C, E
2	T2 T3 T4	A, C, E B, C, E B, D, E
3	T3 T4 T5	B, C, E B, D, E A
4	T4 T5 T6	B, D, E A B, C
5	T5 T6	A B, C
6	T6	B, C

Sliding DB

SID	TID	Extended itemset
1	T1 T2 T3	A(0), B(0), D(0) A(1), C(1), E(1) B(2), C(2), E(2)
2	T2 T3 T4	A(0), C(0), E(0) B(1), C(1), E(1) B(2), D(2), E(2)
3	T3 T4 T5	B(0), C(0), E(0) B(1), D(1), E(1) A(2)
4	T4 T5 T6	B(0), D(0), E(0) A(1) B(2), C(2)
5	T5 T6	A(0) B(1), C(1)
6	T6	B(0), C(0)

Extended DB

Figure 6: Sliding windows → Segments

- Segment (Window)
- Extended itemset (Eitemset)
- K-eitemset

Mining Inter-transaction Association Rules (2/2)

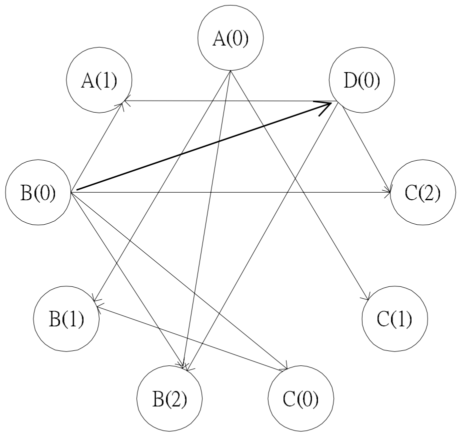
- The Support of a Eitemset
  - A(0)C(1)) = 3/6 = 100%
- The Confidence of a rule
  - X → Y
  - Conf(A(0)→C(1))
    = Sup(A(0)C(1)) / Sup(A(0))
    = 0.5 / 0.5 = 100%
- Large Eitemset &
  Inter-transaction association
  rule
  - minsup & minconf

SID	TID	Extended itemset
1	T1	A(0), B(0), D(0)
	T2	A(1), C(1), E(1)
	T3	B(2), C(2), E(2)
2	T2	A(0), C(0), E(0)
	T3	B(1), C(1), E(1)
	T4	B(2), D(2), E(2)
3	T3	B(0), C(0), E(0)
	T4	B(1), D(1), E(1)
	T5	A(2)
4	T4	B(0), D(0), E(0)
	T5	A(1)
	T6	B(2), C(2)
5	T5	A(0)
	T6	B(1), C(1)
6	T6	B(0), C(0)

Extended DB

The Graph-based Approach (1/3)

- Graph Index Structure



The Graph-based Approach (2/3)

- E-Matrix

	A(0):3	B(0):5	C(0):1	D(0):3
A(1)	∅	{1,4}	∅	{1,4}
B(1)	{2,5}	∅	{2,3}	∅
B(2)	{1,2}	{1,4}	∅	{1,4}
C(0)	∅	{3,6}	∅	∅
C(1)	{1,2,5}	∅	∅	∅
C(2)	∅	{1,4}	∅	{1,4}
D(0)	∅	{1,4}	∅	∅

The Graph-based Approach (3/3)

- Discovering large k-eitemsets
  - Candidate generation
    - there are three corresponding edges starting from the node A(0) in the graph-index.
    - A(0)B(1), A(0)B(2), and A(0)C(1).
    - Candidate 4-eitemset <A(0)B(1)B(2)C(1)>, <A(0)B(2)B(1)C(1)>,…, are generated
  - Support computation
    - The support of <A(0)B(1)B(2)C(1)> is 1/6 (it is only contained in segment 2). It is not a large 4-eitemset.

Conclusion

- Contribution
  - We classify the problems into three types
    - intra-type transformation, inter-type transformation
  - We propose a graph-based approach for mining length k inter-transaction association rules, which only needs to scan the database twice.
- Future work
  - A more comprehensive framework that considers all the related problems.
  - develop a mining algorithm that can be applied to any of the related problems

Appendix