High Utility Itemsets (HUI)

Marzena Kryszkiewicz
Warsaw University of Technology

Utility of Itemsets, High Utility Itemsets

Example Transaction Dataset D with utilites

Id	*Transaction*
T_1	a(1) c(10) d(1)
T_2	a(2) c(6)
T_3	a(2) b(2) d(6)
T_4	b(4) c(13) d(3)
T_5	b(2) c(4)
T_6	a(1) b(1) c(1) d(1)

- *Utility of itemset* X is the sum of utilities of X's items in transactions containing X.
- *ut*({abcd}) = 4, *ut*({cd}) = 29, *ut*({d}) = 11.
- X is called a *high utility itemset* (*HUI*) if its utility ut(X) > threshold *minUtil*.
- Let *minUtil* = 28:
  - {cd} is a high utility itemset
  - {d} and {abcd} are not high utility itemsets.
- **Property.** There is no guarantee that the utility of itemset X is greater than or equal to the utility of its proper superset.

Transaction Utility of Itemset

Example Transaction Dataset D with utilities

Id	*Transaction*	*Utility of transaction*
T_1	a(1) c(10) d(1)	12
T_2	a(2) c(6)	8
T_3	a(2) b(2) d(6)	10
T_4	b(4) c(13) d(3)	20
T_5	b(2) c(4)	6
T_6	a(1) b(1) c(1) d(1)	4

- *Utility of a transaction* is the sum of utilities of all items in this transaction.
- *Transaction utility of itemset* X is the sum of utilities of transactions containing X.
- *ut*({abcd}) = 4, *ut*({cd}) = 29, *ut*({d}) = 11.
- *tut*({abcd}) = 4, *tut*({cd}) = 36, *tut*({d}) = 46.
- **Property.**
  - tut(X) ≥ ut(X).
  - The transaction utility of itemset X is greater then or equal to the transaction utility of its proper superset.

Two-Phase Algorithm: Phase I

Example Transaction Dataset D with utilities

Id	*Transaction*	*Utility of transaction*
T_1	a(1) c(10) d(1)	12
T_2	a(2) c(6)	8
T_3	a(2) b(2) d(6)	10
T_4	b(4) c(13) d(3)	20
T_5	b(2) c(4)	6
T_6	a(1) b(1) c(1) d(1)	4

- minUtil = 28.

Candidates (with transaction utilities):

- $C_1 \rightarrow$Promising C_1: a(34) b(40) c(50) d(46)
- $C_2 \rightarrow$ Promising C_2: ab(14) ac(24) ad(26) bc(30) bd(34) cd(36)
- $C_3 \rightarrow$ Promising C_3: bcd(24)

Two-Phase Algorithm: Phase II

Example Transaction Dataset D with utilities

Id	*Transaction*	*Utility of transaction*
T_1	a(1) c(10) d(1)	12
T_2	a(2) c(6)	8
T_3	a(2) b(2) d(6)	10
T_4	b(4) c(13) d(3)	20
T_5	b(2) c(4)	6
T_6	a(1) b(1) c(1) d(1)	4

- minUtil = 28.

Result of phase I (with transaction utilities):

- Promising C_1: a(34) b(40) c(50) d(46)
- Promising C_2: bc(30) bd(34) cd(36)

Phase II (with utilities):

- Promising $C_1 \rightarrow HUI_1$: a(6) b(9) c(34) d(11)
- Promising $C_2 \rightarrow HUI_2$: bc(25) bd(17) cd(29)

References

- Ying Liu, Wei-keng Liao, Alok Choudhary: A Two-Phase Algorithm for Fast Discovery of High Utility Itemsets. PAKDD 2005: 689-695
- http://www.philippe-fournier-viger.com/spmf/