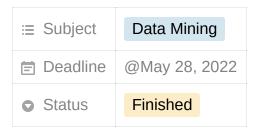
Association Rule: SETM



Thoery

STEM algorithm is a proposed solution to calculate large itemsets and each member of the large itemsets is in the form of <TID, itemset> where TID is unique identifier.

- Candidate itemsets are generated on-the-fly as the database is scanned, but count at the end of the pass
- 1. New candidate itemsets are generated the same way as in AIS algorithm, but the TID of the generating transaction is saved with the candidate items set in sequential structure
- 2. At the end of the pass, the support count of candidate itemset is determined by aggregating this sequential structure

Advantages

- Fair performance when dealing with a small amount of dataset
- Easy to implement

Disadvantages

- There are too many candidate itemsets and unnecessarily generating itemsets
- For each candidate items, there are as many entries as its support value
- Require large amounts of space to store if there are many TID

Association Rule: SETM 1

Example of Applying SETM concept

Database

Transaction ID	Items
1	CAKE, CANDLE, GIFT
2	CANDLE, GIFT

L1

Itemset	Support			
{CAKE}	1			
{CANDLE}	2			
{GIFT}	1			

C2

Itemset	TID			
{CAKE, CANDLE}	1			
{CANDLE, GIFT}	1			
{CANDLE, GIFT}	2			

C3

Itemset	TID
{CAKE, CANDLE, GIFT}	1

Code Implementation

Association Rule: SETM 2

Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	1/12/10 8:26	2,55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	1/12/10 8:26	3,39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	1/12/10 8:26	2,75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	1/12/10 8:26	3,39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	1/12/10 8:26	3,39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	1/12/10 8:26	7,65	17850	United Kingdom
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	1/12/10 8:26	4,25	17850	United Kingdom
536366	22633	HAND WARMER UNION JACK	6	1/12/10 8:28	1,85	17850	United Kingdom
536366	22632	HAND WARMER RED POLKA DOT	6	1/12/10 8:28	1,85	17850	United Kingdom
536368	22960	JAM MAKING SET WITH JARS	6	1/12/10 8:34	4,25	13047	United Kingdom
536368	22913	RED COAT RACK PARIS FASHION	3	1/12/10 8:34	4,95	13047	United Kingdom

- Read the data from CSV file which is convert from Excel file
- Preprocessing and implement the concept of SETM

References

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/69727e5d-9262-4 0a2-ba4e-91d4454b2b31/A_SURVEY_OF_ASSOCIATION_RULES.pdf

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/a587ce14-3054-4 0f7-b3e6-6a9c97a3da6b/Chapter6_Association_Rule.pdf

Association Rule: SETM 3