

TITLE: Basic Concepts: Frequent Patterns @ Association Rules

Group B2	
Chong Kai Zhe	A20EC0186
Low Junyi	A20EC0071
Madina Suraya	A20EC0203
Vincent Boo Ee Khai	A20EC0231

Defination

- A pattern that occurs frequently in a data set
- Frequent Pattern Mining (AKA Association Rule Mining) is an analytical process that finds frequent patterns, associations, or causal structures from data sets found in various kinds of databases such as relational databases, transactional databases, and other data repositories. Given a set of transactions, this process aims to find the rules that enable us to predict the occurrence of a specific item based on the occurrence of other items in the transaction. (Patel, 2018)

Support

How often a given rule appears in the database being mined. (Perera, 2022)

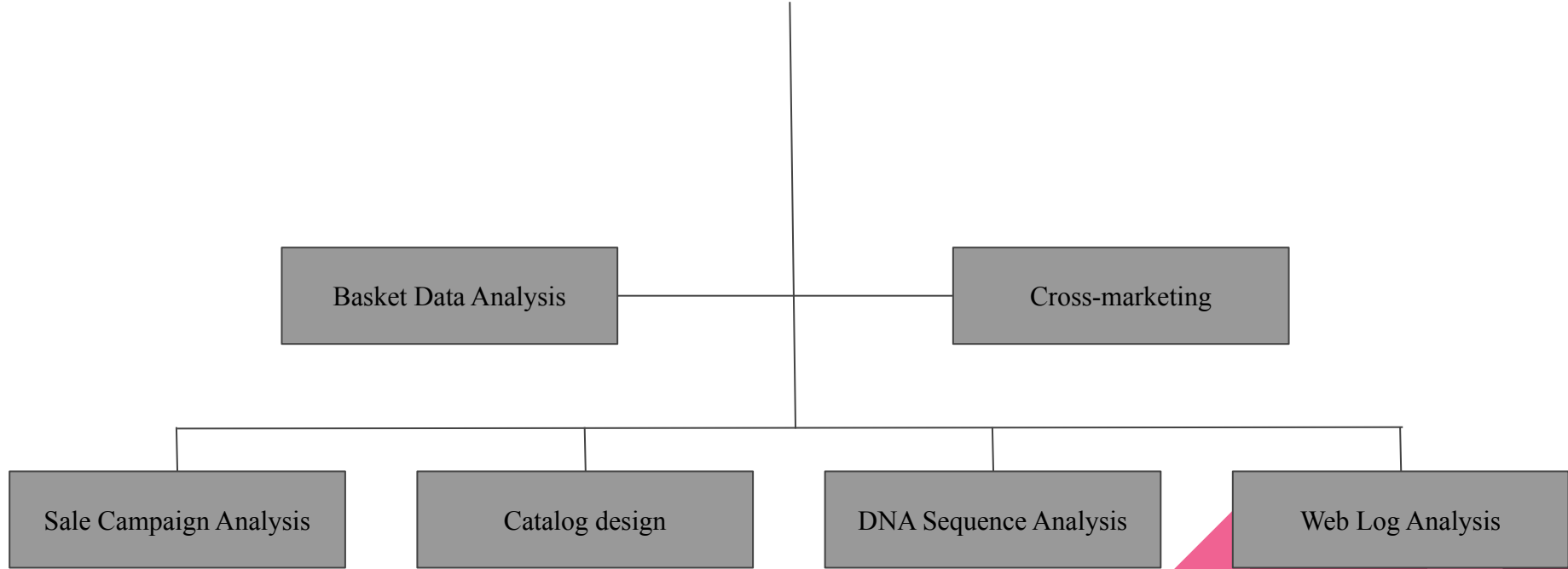
$$\text{Support} = \text{Frequency (A, B)} / N$$

Confidence

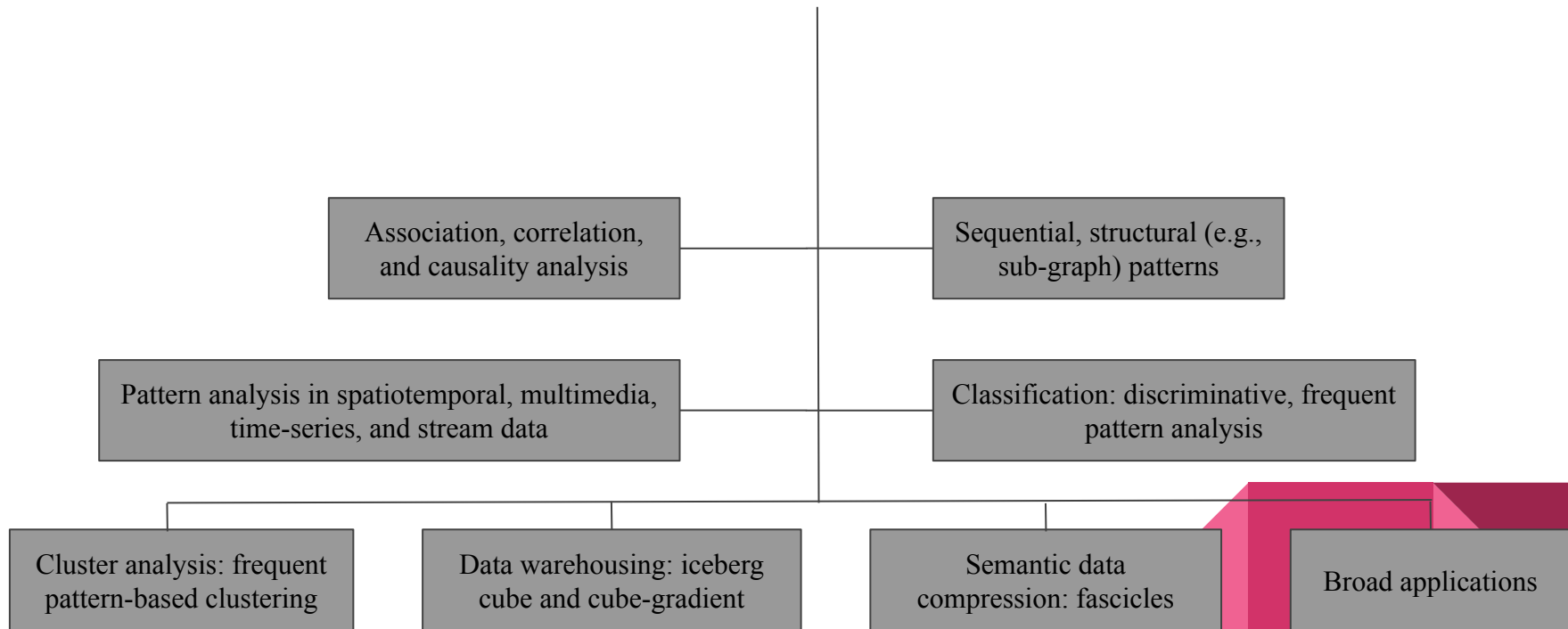
The number of times a given rule turns out to be true in practice. (Perera, 2022)

$$\text{Confidence} = \text{Frequency (A, B)} / \text{Frequency (A)}$$


Application

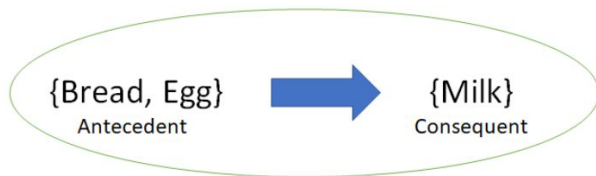


Importance



Basic Concept

- **itemset**: A set of one or more items
- **k-itemset**: $X = \{x_1, \dots, x_k\}$
- **(absolute) support** or **support count of X**: Frequency or occurrence of itemset X.
- **(relative) support, s**: fraction or probability of transactions that contains X.
- An itemset X is **frequent** if X's support is **not less** than **minimum support, *minsup*** threshold.

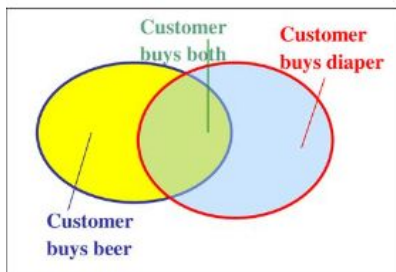


Itemset = {Bread, Egg, Milk}

Basic Concept

- **support, s**: probability that a transaction contains $X \cup Y$
- **confidence, c**: conditional probability that transaction having X also contains Y .
- **Itemset, $X = \{x_1, \dots, x_k\}$**

Transaction-id	Items bought
10	A, B, D
20	A, C, D
30	A, D, E
40	B, E, F
50	B, C, D, E, F



Let $sup_{min} = 50\%$, $conf_{min} = 50\%$
Freq. Pat.: $\{A:3, B:3, D:4, E:3, AD:3\}$

Association rules:

$A \rightarrow D$ (60%, 100%)

$D \rightarrow A$ (60%, 75%)

LIFT

- Lift is a proposal for selecting and classifying patterns according to their potential interest to the user (Alashqur et al, 2015).
- It allows the time and space costs of the mining process to be reduced.

$$Lift(\{X\} \rightarrow \{Y\}) = \frac{(Transactions\ containing\ both\ X\ and\ Y)/(Transactions\ containing\ X)}{Fraction\ of\ transactions\ containing\ Y}$$

- Lift controls the support (frequency) of consequent while calculating probability of occurrence of $\{Y\}$, with given $\{X\}$.
- $\{X\}$ provides confidence of having $\{Y\}$.
- Value of lift greater than 1 shows a high association between $\{Y\}$ and $\{X\}$.
- The more the value of lift, greater chances of $\{Y\}$ if there is $\{X\}$.

Calculation 1

Transaction	Item Occurrence
T1	A, B, C
T2	A, C, D
T3	A, B, C, D
T4	A, D, E
T5	B, C

Support = Frequency (A, D) / N

Confidence = Frequency (A, D) / Frequency A

Possible Association Rule: $A \Rightarrow D$

Total no. of Transactions(N) = 5

Frequency (A,D) = > Total no. of instances together A with D = 3

Frequency (A) = > Total no. of occurrence in A = 4

Support = $\frac{3}{5}$, 60% of all the transactions show that A and D are together

Confidence = $\frac{3}{4}$, 75% of transaction in A have D

Calculation 2

Occurrence frequency / frequency / support count / absolute support: number of transactions that contain the itemset.

Relative support: percentage of the occurrence.

Support ($X \rightarrow Y$) = $P(X \cup Y)$

Confidence ($X \rightarrow Y$) = $P(Y|X)$ = Support Count ($X \cup Y$) / Support Count (X)

Tid	Items bought
10	Tea, Nuts, Diaper
20	Tea, Coffee, Diaper
30	Tea, Diaper, Eggs
40	Nuts, Eggs, Milk
50	Nuts, Coffee, Diaper, Eggs, Milk

Diaper $\rightarrow$ Tea (60%, 75%)

Support = Frequency (Diaper, Tea) / N, $P(\text{Diaper} \cup \text{Tea}) = \frac{3}{5} = 60\%$

Confidence = Frequency (Diaper, Tea) / Frequency (Diaper) = $P(\text{Tea} | \text{Diaper}) = \frac{3}{4} = 75\%$

Support: 60% of all the transactions show that Diaper and Tea are purchased together.

Confidence: 75% of customers who purchased a Diaper also bought the Tea. $\frac{3}{4} = 75\%$

Reference

Patel, K. (2018, November 14). *What is Frequent Pattern Mining (Association) and How*

Does it Support Business Analysis? DATAVERSITY.

<https://www.dataversity.net/frequent-pattern-mining-association-support-business-analysis/#>

Perera, P. (2022, January 6). *Frequent pattern mining, Association, and Correlations.*

Medium. <https://towardsdatascience.com/frequent-pattern-mining-association-and-correlations-8fa9f80c22ef>

Hussein, N., Alashqur, A., & Sowan, B. (2015). Using the interestingness measure lift to generate association rules.

Journal of Advanced Computer Science & Technology, 4(1), 156.

