①

$I=\{i_1, i_2, \cdots i_d\}$ 是 d 个 items in set in a dataset

$T=\{t_1, t_2 \cdots t_N\}$ 是 N 个 transactions in set in a dataset.

t_j 是 I in subset

if an itemset contains k items then called k-itemset

②

support count $\sigma(X)$：包含X in transactions in 数量

Support：$\frac{\sigma(X)}{|T|}$ $|T|=N$ 即 transaction 数量

Association rule：$X \to Y$ where $X \subset I$ $Y \subset I$ and $X \cap Y = \emptyset$

e.g. {Diaper, Milk} → {Beer}

③

$$\text{Support}(X) = \frac{\sigma(X)}{|T|} = \frac{\sigma(X)}{N} = P(X)$$

即 X 出现 in 概率

$$\text{Support}(X \to Y) = \frac{\sigma(X \cup Y)}{|T|} = P(X \cup Y)$$

$$\text{Confidence}(X \to Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} = P(Y|X)$$

即出现 X 时 出现 Y in 概率

表示为 transaction 中出现 X 和 Y in 概率（次数）

除以 transaction 中出现 X in 概率（次数）

例：

Example: X= { Beer, Milk, Diaper }; $\sigma(X)$= ?
X= { Milk, Diaper }; $\sigma(X)$= ?

Consider the rule {Milk, Diaper} → {Beer}

Support({Milk, Diaper} → {Beer}) = 2/5 = 0.40
Confidence({Milk, Diaper} → {Beer}) = 2/3 = 0.67

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

32

$X = \{Beer, milk, diaper\}$

$\Rightarrow \sigma(X) = 2$

$X = \{Milk, Diaper\}$

$\Rightarrow \sigma(X) = 3$

$\Rightarrow \{Milk, diaper\} \rightarrow \{Beer\}$

Support: $\frac{2}{5}$ (占全部 transaction P)

confidence: $\frac{2}{3}$ (占买了 Milk 和 diaper 的 P)

ARM (association rule mining) 的 rules

的可能情况当有 d items : $3^d - 2^{d+1} + 1$