

Data Mining

Find meaningful data patterns from a large data set
Computationally very expensive process.

Frequency of an itemset

Freq, {bread, egg} = No. of shopping carts that included bread as well as egg

Support of an itemset = $\frac{\text{No. of shopping carts with the itemset}}{\text{Total \# of shopping carts}}$
range: 0 to 1

Association rule: {LHS itemset} $\Rightarrow$ {RHS itemset}
ie if a customer purchase LHS itemset, s/he likely to purchase the RHS itemset

$$\begin{aligned}\text{Confidence of a rule} &= \frac{\text{Support \{LHS \cup RHS\}}}{\text{Support \{LHS\}}} \\ \{milk\} \Rightarrow \{cheese\} &= \frac{\text{Support \{milk, cheese\}}}{\text{Support \{milk\}}} \\ &= \frac{\frac{75}{1000}}{\frac{300}{1000}} = \frac{75}{300} = 0.25\end{aligned}$$

If support (RHS) is close to the confidence value
then ~~presence~~ distribution of RHS
among all shopping carts is uniform.

$$\text{Support \{cheese\}} = \frac{250}{1000} = 0.25 \text{ among all shopping carts.}$$

cheese is uniformly distributed among all shopping carts.

Interest of a rule:

Deviation of the support {RHS} from the confidence of the rule.

$$\begin{aligned} & \{beer\} \Rightarrow \{chips\} \\ \text{Confidence} &= \frac{\text{Support}\{beer, chips\}}{\text{Support}\{beer\}} = \frac{\frac{150}{1000}}{\frac{450}{1000}} \\ &= \frac{150}{450} = \frac{1}{3} = 0.33 \end{aligned}$$

$$\text{Support}\{RHS\} = \text{Support}\{chips\} = \frac{600}{1000} = 0.60$$

$$\begin{aligned} \text{interest}_{\{beer\} \Rightarrow \{chips\}} &= \left| \text{Confidence} - \text{Support}\{RHS\} \right| \\ &= \left| 0.33 - 0.60 \right| = 0.27 \end{aligned}$$

Generation of Frequent Itemsets

Frequency of an itemset must be $\geq$ minimum support value
(or support for the itemset)

e.g. Minimum Support = 0.4

? Level-1 ^{Frequent} itemsets

↓
Level-2 "

↓
Level-n "

Algorithms

A-Priori

Frequent-Pattern

Brute
Force

Generate all candidate itemsets for each level.
Check for minm frequency for each candidate itemset

Frequent-Pattern

Compute Level-1 Frequent itemset

Build Frequent pattern tree (FP Tree)

Prune FP Tree to generate higher order frequent itemsets.