Basic Concepts for Association Discovery

An association rule is written A => B where A is the *antecedent* and B is the *consequent*. Both sides of an association rule can contain more than one item. Techniques used in Association discovery are borrowed from probability and statistics. ***Support, confidence and Lift*** are three important evaluation criteria of association discovery.

Support

The level of support is how frequently the combination occurs in the market basket (database). Support is the percentage of baskets (or transactions) that contain both A and B of the association, i.e. % of baskets where the rule is true

> Support(A => B) = P(A ∩ B)

Expected confidence

This is the probability of the consequent if it was independent of the antecedent. Expected confidence is thus the percentage of occurrences containing B

> Expected confidence (A => B) = P(B)

Confidence

The strength of an association is defined by its confidence factor, which is the percentage of cases in which a consequent appears given that the antecedent has occurred. Confidence is the percentage of baskets having A that also contain B, i.e. % of baskets containing B among those containing A. Note: Confidence(A => B) ≠ Confidence(B => A).

> Confidence(A => B) = P(B | A)

> Lift

$$Lift(A \Rightarrow B) = \frac{P(B \mid A)}{P(B)} = \frac{P(A \cap B)}{P(A)P(B)}$$

Lift is equal to the confidence factor divided by the expected confidence. Lift is a factor by which the likelihood of consequent increases given an antecedent. Expected confidence is equal to the number of consequent transactions divided by the total number of transactions. Lift is the ratio of the likelihood of finding B in a basket known to contain A, to the likelihood of finding B in any random basket.

Association Rules

Transaction ID \ items	iphones	ipad	Apple TV	Apple watch	Airpod
1	1	1	0	1	0
2	1	1	1	0	0
3	1	1	1	0	1
4	0	0	1	1	1
5	1	0	1	1	0
6	1	1	1	1	1
7	1	0	0	0	0
8	1	1	1	1	0
9	0	0	0	1	1
10	1	1	1	0	0

Rule 1 : Buying iphones ⇒ buying ipad

(iphones) ⇒ (ipads)

Rule : 2 (ipads) ⇒ (Apple TV)

Rules 3 (iphones, ipads) ⇒ (Apple watch)

Rule 4 (ipads, Apple TV) ⇒ (Apple watch)

⋮

How "significant" is a rule $(A \Rightarrow B)$?

① Support of a rule:

$$\text{support}(A \Rightarrow B) = P(A \cap B) = \frac{\#\ A \text{ and } B}{\#\ \text{total items}}$$

$$\text{support}(\text{iphones} \Rightarrow \text{ipads}) = \frac{6}{10} = 60\%$$

② Confidence

$$\text{confidence}(A \Rightarrow B) = P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{\#\ A \text{ and } B}{\#\ A}$$

$$\text{confidence}(\text{iphones} \Rightarrow \text{ipads}) = \frac{6}{8} = 75\%$$

③ Expected confidence: $P(B) = \frac{\#\ B}{\#\ \text{total items}}$

$=$

Exp. confidence (iphones $\Rightarrow$ ipads) $= \frac{\#\ \text{ipads}}{\#\ \text{total}} = \frac{6}{10}$

$= 60\%$

(4) lift $= \frac{\text{confidence}}{\text{Expected confidence}} = \frac{75\%}{60\%} = 1.25$