

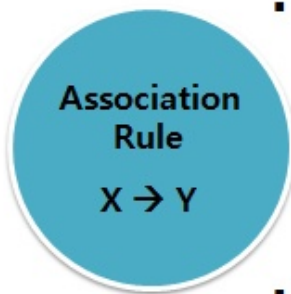
연관규칙 평가 척도 : 지지도(Support), 신뢰도(Confidence), 향상도(Lift)

X와 Y를 서로 공통원소가 없는 항목들의 집합이고 (즉, $X \cap Y = \emptyset$),

$X \rightarrow Y$: 연관규칙 if X then Y ← 기본 base

N : 전체 transaction 수 (즉, 전체 row 개수)

$n(X)$ = : 전체 transaction에서 항목집합 X를 포함하는 transaction 수라고 하면,



▪ 지지도(Support) : $s(X \rightarrow Y) = \frac{n(X \cup Y)}{N}$

← X와 Y를 동시에 포함하는 거래수.

▪ 신뢰도(Confidence) : $c(X \rightarrow Y) = \frac{n(X \cup Y)}{n(X)}$

▪ 향상도(Lift) : $Lift(X, Y) = \frac{c(X \rightarrow Y)}{s(Y)}$

↑ 핵심!!

[R 분석과 프로그래밍] <http://rfriend.tistory.com>

$s(Y) = \frac{n(Y)}{N}$

지지도(Support), 신뢰도(Confidence), 향상도(Lift) 예시

Customer ID	Transaction ID	Items
1131	1번	계란, 우유
2094	2번	계란, 기저귀, 맥주, 사과
4122	3번	우유, 기저귀, 맥주, 콜라
4811	4번	계란, 우유, 맥주, 기저귀
8091	5번	계란, 우유, 맥주, 콜라

연관규칙 {계란, 맥주} → {기저귀} 에 대해

X Y

지지도(Support)

$$\begin{aligned} s(X \rightarrow Y) &= n(X \cup Y) / N \\ &= n(2번, 4번) / N \\ &= 2 / 5 = 0.4 \end{aligned}$$

신뢰도(Confidence)

$$\begin{aligned} c(X \rightarrow Y) &= n(X \cup Y) / n(X) \\ &= n(2번, 4번) / n(2번, 4번, 5번) \\ &= 2 / 3 = 0.667 \end{aligned}$$

향상도(Lift)

$$\begin{aligned} Lift(X \rightarrow Y) &= c(X \rightarrow Y) / s(Y) \\ &= 0.667 / 0.6 = 1.111 \end{aligned}$$

N = 5 (전체 transaction 개 수)

$$\begin{aligned} s(Y) &= n(Y) / N \\ &= n(2번, 3번, 4번) / N = 3 / 5 = 0.6 \end{aligned}$$

[R 분석과 프로그래밍] <http://rfriend.tistory.com>

위의 3개의 rule 평가 척도를 소개하였는데요, 각각의 평가 관점이 다르기 때문에 어느 하나만을 가지고 rule을 평가하는 것이 아니라 보통은 3개의 척도를 모두 사용합니다.

① 보통 (1) 특정 지지도(Support)와 신뢰도(Confidence) 이하의 rule은 screening out 시키게끔 해놓고,

(minimum support, minimum confidence)

② 향상도(Lift) 내림차순(양의 관계를 찾을 때)으로 sorting을 해서 rule을 평가하는 식으로 이용하곤 합니다.

그리고 관심이 있는 상품이나 item이 있으면 목적에 맞게 해당 item이 left-hand side 나 right-hand side 에 있는 rule만을 subset으로 선별해서 보기도 하구요.