به نام خدا

تحلیل داده های حجیم

تمرین سری دوم

استاد: دکترغلامپور

دانشجو:سجاد هاشم بیکی

پاییز1401

سوال1)

الف:

$$conv(A \to B) = \frac{1 - \mathrm{S}(B)}{1 - conf(A \to B)} = \frac{p(A) * p(B')}{p(A \cup B')}$$

$p(B')$ is the probability that B does not appear in a basket.

Conviction compares the probability that A appears without B if they were dependent with the actual frequency of the appearance of X without B. Unlike confidence, conviction factors in both $p(A)$ and $p(B)$ and always has a value of 1 when the relevant items are completely unrelated. In contrast to lift, conviction is a directed measure because it also uses the information of the absence of the consequent. Hence, conviction is monotone in confidence and lift.

ب:

Lift is not sensitive to rule direction:

$$lift(A \to B) = \frac{conf(A \to B)}{\sup(B)} = \frac{\sup(A \cup B)}{\sup(B)\sup(A)} = \frac{conf(B \to A)}{\sup(A)} = lift(B \to A)$$

ج:

ضعف پارامتر confidence بدین صورت است که ممکن است یک rule مقدار confidence بالایی داشته باشد ولی اصلا interest نباشد.مثلا B را ایتم شیر در نظر بگیریم،احتمالا confidence بالایی به ازای A های مختلف داشته باشد چرا که،ایتم شیر پرفروش است(در اکثرسبدها است) و احتمالا مستقل از A خریداری میشود،که در این صورت confidence بالاست اما interest نیست.

اما در دیگر پارمترها مانند interest , lift , conviction این موضوع درنظرگرفته شده است.بطوریه احتمال B (ساپورت B تقسیم بر تعداد سبدها) به رابطه انها اضافه شده است.