Demostraciones sheet 7

miércoles, 1 de mayo de 2024 16:52

1) a) $\text{lift}(X \to \neg Y) = \dfrac{\text{conf}(X \to \neg Y)}{\text{rsup}(\neg Y)} > K$

$\text{conf}(X \to \neg Y) > K \cdot \text{rsup}(\neg Y) \Rightarrow \dfrac{\text{conf}(X \to \neg Y)}{K} < \text{rsup}(\neg Y)$

Rsup de Y esta definida por la siguiente formula $\text{rsup}(X \to Y) = P(XY)$, entonces $\text{rsup}(\neg Y) = P(\neg Y)$

$\dfrac{\text{conf}(X \to \neg Y)}{K} > P(\neg Y)$  $\text{conf}(X \to Y) = \dfrac{P(XY)}{P(X)}$, entonces

$\dfrac{\frac{P(X \neg Y)}{P(X)}}{K} > P(\neg Y) \to \dfrac{P(X \neg Y)}{K \cdot P(X)} > P(\neg Y) \to \dfrac{P(X \neg Y)}{K} > P(X)P(\neg Y)$

$\dfrac{1}{K} > \dfrac{P(X)\,P(\neg Y)}{P(X \neg Y)} \Rightarrow \dfrac{1}{K} > \text{conv}(X \to Y)$

Para esta demostración la relación entre el lift y la conv, está basada en que la conv (X -> Y) < 1/k si y solo si lift (X -> ⌐Y) > k, en este caso por medio de despeje de inecuaciones de primer grado se puede despejar la ecuación que por definición representa el lift entre una premisa y una conclusión, ese despeje nos da como resultado una interpretación de la definición de conv entre una premisa y una conclusión, esto también puede ser visto a través de la definición de conv (X --> Y) esta es igual a 1/lift (X --> ⌐Y), el cual es el inverso del lift de una premisa y una conclusión negada

b) $\text{lift}(X \to \neg Y) > 1 \to \dfrac{\text{conf}(X \to \neg Y)}{\text{rsup}(\neg Y)} > 1$

$\text{rsup}(\neg Y) = 1 - \text{rsup}(Y)$

$\text{conf}(X \to \neg Y) = 1 - \text{conf}(X \to Y)$

$\dfrac{1 - \text{conf}(X \to Y)}{1 - \text{rsup}\,Y} > 1 \to 1 - \text{conf}\ X \to Y > 1 - \text{rsup}(Y) \to -\text{conf}(X \to Y) > -\text{rsup}(Y)$

$\to \boxed{\text{conf}(X \to Y) < \text{rsup}(Y)}$

Como podemos ver se obtuvo la premisa marcada a través de inecuaciones, la cual nos dice que la confianza entre X y Y es menor al rsup de la conclusión, si aplicamos esta premisas para el lift de los dos Itemsets positivos, podemos ver que el numerador será menor que el denominador y por definición de

$\text{lift}(X \to Y) = \dfrac{\text{conf}(X \to Y)}{\ldots}$

$$\text{list}(x \to y) = \frac{\text{conf}(x \to \neg y)}{\text{rsup}(Y)}$$

$$\text{list}(x \to Y) < 1 \iff \text{lift}(x \to \neg Y) > 1$$

premisas para el lift de los dos Itemsets positivos, podemos ver que el numerador será menor que el denominador y por definición de fracciones, si el denominador es mayor que el numerador el cociente será un valor entre 0 y 1 por lo que es menor que 1, haciendo verdaderas las dos premisas de la demostración si y solo si

2) Para esta demostración se va a volver a contemplar el ejemplo presentado en la demostración 1.3 de la sheet 6, el ejemplo será anexado a continuación

$X = \{apple, banana\}$ $W = \{apple\}$ $Y = \{grape\}$

$W \subset X$

$W \cup Y = \{apple, grape\}$

$X \cup Y = \{apple, banana, grape\}$

Transactions

1. {apple, banana, grape}
2. {apple, grape}
3. {apple, banana}
4. {apple, banana}
5. {apple}

$sup(X) = 3$ $sup(W) = 5$

$sup(W \cup Y) = 2$ $sup(X \cup Y) = 1$

$$conf(X \to Y) = \frac{sup(X \cup Y)}{sup(X)} = \frac{1}{3} \qquad conf(W \to Y) = \frac{sup(W \cup Y)}{sup(W)} = \frac{2}{5}$$

$$\boxed{conf(X \to Y) < conf(W \to Y)}$$

$$Imp(X \to Y) = conf(X \to Y) - \max conf(W \to Y)$$

siendo $W \subset X$

Teniendo en cuenta el resultado del ejemplo de la demostración 1.3 y la definición del improvement, podemos ver que se puede dar el caso de que la máxima confianza entre W y Y sea mayor a la confianza de X y Y, por lo tanto si es posible que el improvement sea negativo.

Esto significa que la "mejora" que estamos dándole a la regla de asociación agregándole más elementos, es decir utilizar a X que es el conjunto completo envés de w que solo incluye una parte de x, no mejora la confianza de la regla, en otras palabras hace que sea menos confiable sugiriendo que los elementos adicionales están induciendo ruido o complejidad innecesaria en la regla de asociación.

3) $Y \subseteq W \subseteq YZ$

$sup(XY) = sup(XYZ)$

De esta afirmación podemos indicar 2 cosas, la primera es que los soportes de estos pueden ser igual siempre y cuando se dé el caso de que sean iguales los conjuntos, lo segundo es que podemos decir que toda transacción

Ahora por esta segunda afirmación se puede afirmar que como el soporte de XY siempre es igual a el soporte de XYZ, entonces la adición de Z al itemset XY no afecta si soporte por lo que es que los soportes de estos pueden ser igual siempre y cuando se dé el caso de que sean iguales los conjuntos, lo segundo es que podemos decir que toda transacción que tenga a W también tiene a Y, además de que W esta contenido completamente en YZ

$$sup(Y) = sup(W) = sup(YZ)$$

afirmar que como el soporte de XY siempre es igual a el soporte de XYZ, entonces la adición de Z al itemset XY no afecta si soporte por lo que no afecta la frecuencia de XY

a) $$conf(X \to Y) = \frac{sup(XY)}{sup(X)} \qquad conf(X \to W) = \frac{sup(XW)}{sup(X)}$$

$$sup(XY) = sup(XYZ) \longrightarrow sup(XW) = sup(XYW)$$

Es se puede decir ya que Y es un subset de W por lo que añadir W a XY (formando XW) no cambia las transacciones donde ocurre XY dado que cualquier transacción que tenga XY debe si o si contener a W por ser un subset, así que añadir W no afectar las transacciones donde ocurre XYW

$$\boxed{\frac{sup(XY)}{sup(X)} \quad \frac{sup(XYZ)}{sup(X)} = \frac{sup(WX)}{sup(X)} = \frac{sup(XYW)}{sup(X)}}$$

$$conf(X \to \neg Y) = \frac{sup(X\neg Y)}{sup\ X} \qquad conf(X \to \neg W) = \frac{sup(X \neg W)}{sup(X)}$$

$$sup(X\neg Y) \overset{1}{=} sup(X \neg W) \Rightarrow X\neg Y \overset{2}{\subseteq} X \neg W$$

Esta igual dad de soportes surge de la afirmación de que el complemento de Y (la negación de Y) está contenido en el complemento de W (negación de W) ya que el complemento de W representa a su vez todo lo que no está en Y como vimos anteriormente por lo tanto se da la afirmación 2 que es la premisa para la igualdad de los soportes

$$\boxed{\frac{sup(X\neg Y)}{sup\ X} = \frac{sup(X\neg W)}{sup(X)}}$$

b) $$lift(X \to W) \geq lift(X \to Y)$$

$$\frac{conf(X \to W)}{rsup(W)} \geq \frac{conf(X \to Y)}{rsup(Y)} \qquad \frac{\frac{rsup(XW)}{rsup(X)}}{rsup(W)} \geq \frac{\frac{rsup(XY)}{rsup(X)}}{rsup(Y)}$$

$$\frac{rsup(XW)}{\cancel{rsup(X)}\ rsup(W)} \geq \frac{rsup(XY)}{\cancel{rsup(X)}\ rsup(Y)}$$

$$rsup(WX) = rsup(XY)$$

$$rsup(W) \leq rsup(Y)$$

$$\frac{rsup(xw)}{rsup(w)} \geq \frac{rsup(xY)}{rsup(Y)}$$

Con estas dos afirmaciones planteadas ahora hacemos el análisis de la expresión simplificada, como podemos ver el soporte de W al ser menor que el de Y el cociente del resultado será mayor por la definición de cociente, por lo tanto el lift (X --> W) será mayor que el lift (X --> Y)

Gracias a ala anterior demostración sabemos que el soporte de wx y el soporte de xy puede ser igual, por lo tanto para este análisis será tomado como que es igual, en cuanto a rsup(w) y el rsup(Y) sabemos por las condiciones del ejercicio que cada elemento de Y esta en W pero W puede tener más elementos, por lo tanto el soporte de Y será mayor, al tener la posibilidad de ser más frecuente al tener menos ítems

c) $conv(x \to w) \geq conv(x \to y)$ $\quad \frac{1 - P(w)}{1 - conf(x \to w)} \geq \frac{1 - P(Y)}{1 - conf(x \to w)}$

$$\frac{1 - P(w)}{1 - \frac{P(xw)}{P(x)}} \geq \frac{1 - P(Y)}{1 - \frac{P(xY)}{P(x)}} \qquad \frac{1 - P(w)}{\frac{P(x) - P(xw)}{P(x)}} \geq \frac{1 - P(Y)}{\frac{P(x) - P(xY)}{P(x)}}$$

$$\frac{\cancel{P(x)} \cdot (1 - P(w))}{P(x) - P(xw)} \geq \frac{\cancel{P(x)}(1 - P(Y))}{P(x) - P(xY)} \qquad \frac{1 - P(w)}{P(x) - P(xw)} \geq \frac{1 - P(Y)}{P(x) - P(xY)}$$

$$P(xYZ) = P(xY) = P(xW) = P(xYw)$$

Con esta afirmación que llevamos desde la demostración a podemos decir que el denominador en ambas expresiones serán iguales por lo que el análisis se limita al numerador

$$1 - P(w)$$
$$1 - P(Y)$$

$$sup(w) = P(w) \leq sup(Y) = P(Y)$$
$$1 - P(w) \geq 1 - sup(Y)$$

Como vimos en la demostración b el soporte de Y es mayor o igual al soporte de W, por definición del soporte, este puede ser representado como la probabilidad de dicho itemset, por lo que la probabilidad de Y también es mayor o igual a la de W, entonces si a estas probabilidades se les busca su complemento, es decir restarle a 1 la probabilidad, por lo tanto el complemento de P(W) es mayor que el complemento de P(Y), estos complementos son los que hacen pate de la conviction, demostrando la afirmación

$$conv(x \to w) \geq conv(x \to Y)$$

Dado que $Y \subseteq W$ se considera que el sup(Y) es un subset del sup(W) por lo tanto en este contexto la mejor regla es probable que sea en la que se asocia X con W, esto se puede decir por dos razones principales, la primera es que W es un itemset más largo comparado a Y, haciendo que sea probable que se tenga un soporte mayor, la confianza y el lift de la regla X --> W se espera que mínimo sean tan altos como los de X --> Y debido al soporte más grande de W