Tutorial 11 Solution

1. A local retailer has a database that stores 10,000 transactions of last summer. After analyzing the data, a data science team has identified the following statistics:

   - {battery} appears in 6000 transactions
   - {sunscreen} appears in 5000 transactions
   - {sandals} appears in 4000 transactions
   - {bowls} appears in 2000 transactions
   - {battery, sunscreen} appears in 1500 transactions
   - {battery, sandals} appears in 1000 transactions
   - {battery, bowls} appears in 250 transactions
   - {battery, sunscreen, sandals} appears in 600 transactions

   (a) What are the support values of the preceding itemsets?

   Answer: The support values are as follows:
   $\text{Supp}\Big(\{\text{battery}\}\Big) = 0.6$
   $\text{Supp}\Big(\{\text{suncreen}\}\Big) = 0.5$
   $\text{Supp}\Big(\{\text{sandals}\}\Big) = 0.4$
   $\text{Supp}\Big(\{\text{bowls}\}\Big) = 0.2$
   $\text{Supp}\Big(\{\text{battery, sunscreen}\}\Big) = 0.15$
   $\text{Supp}\Big(\{\text{battery, sandals}\}\Big) = 0.1$
   $\text{Supp}\Big(\{\text{battery, bowls}\}\Big) = 0.025$
   $\text{Supp}\Big(\{\text{battery, sunscreen, sandals}\}\Big) = 0.06$

   (b) Assuming the minimum support is 0.05, which itemsets are considered frequent?

   Answer: All except {battery, bowls} are frequent.

   (c) What are the confidence values of {battery} $\rightarrow$ {sunscreen} and
   {battery, sunscreen} $\rightarrow$ {sandals}? Which of these two rules is more interesting, i.e. has higher values of confidence?

   Answer:

$$\text{Confidence}\ \Big(\{\text{battery}\} \rightarrow \{\text{sunscreen}\}\Big) = \frac{\text{Supp}\Big(\{\text{battery, sunscreen}\}\Big)}{\text{Supp}\Big(\{\text{battery}\}\Big)} = 0.15/0.6 = 0.25.$$

$$\text{Confidence}\ \Big(\{\text{battery, sunscreen}\} \rightarrow \{\text{sandals}\}\Big) = 0.06/0.15 = 0.4.$$

   Clearly, the rule {battery, sunscreen} $\rightarrow$ {sandals} has higher confidence.

Some extra:

$$\text{Confidence}\Big(\{\text{sunscreen}\} \to \{\text{battery}\}\Big) = 1500/5000 = 0.3.$$

$$\text{Confidence}\Big(\{\text{battery}\} \to \{\text{sandals}\}\Big) = 1000/6000 = 0.167.$$

$$\text{Confidence}\Big(\{\text{sandals}\} \to \{\text{battery}\}\Big) = 1000/4000 = 0.25.$$

$$\text{Confidence}\Big(\{\text{battery}\} \to \{\text{bowls}\}\Big) = 250/6000 = 0.042.$$

$$\text{Confidence}\Big(\{\text{bowls}\} \to \{\text{battery}\}\Big) = 250/2000 = 0.125.$$

2. Suppose for three products A, B and C, $\text{support}(\{A\}) = 0.6$, $\text{support}(\{B\}) = 0.6$, $\text{confidence}(\{B\} \to \{A\}) = 0.9$ and $\text{confidence}(\{C\} \to \{A, B\}) = 0.5$. Compute the following quantities.
   (a) $\text{Lift}(\{A\} \to \{B\})$
   (b) $\text{Leverage}(\{A\} \to \{B\})$
   (c) $\text{Confidence}(\{A\} \to \{B\})$
   (d) $\text{Lift}(\{A, B\} \to \{C\})$

Solution:

(a)

$$\text{Lift}(\{A\} \to \{B\}) = \frac{\text{Supp}\big(\{A, B\}\big)}{\text{Supp}(\{A\})\,\text{Supp}(\{B\})} = \frac{\text{Confidence}\Big(\{B\} \to \{A\}\Big) \times \text{Supp}(\{B\})}{\text{Supp}(\{A\})\,\text{Supp}(\{B\})} = \frac{0.9}{0.6} = 1.5$$

(b)

$$\begin{aligned}\text{Leverage}(\{A\} \to \{B\}) &= \text{Supp}(\{A \wedge B\}) - \text{Supp}(\{A\}) \times \text{Supp}(\{B\}) \\ &= \text{Confidence}(\{A\} \to \{B\}) \times \text{Supp}(\{B\}) - \text{Supp}(\{A\}) \times \text{Supp}(\{B\}) \\ &= (0.9)(0.6) - (0.6)(0.6) = 0.18\end{aligned}$$

(c)

$$\text{Confidence}(\{A\} \to \{B\}) = \frac{\text{Supp}(\{A \wedge B\})}{\text{Supp}(\{A\})} = \frac{0.54}{0.6} = 0.9$$

(d)

$$\begin{aligned}\text{Lift}(\{A, B\} \to \{C\}) &= \frac{\text{Supp}(\{A, B \wedge C\})}{\text{Supp}(\{A, B\}) \times \text{Supp}(\{C\})} \\ &= \frac{\text{Confidence}\Big(\{C\} \to \{A, B\}\Big) \times \text{Supp}(\{C\})}{\text{Supp}(\{A, B\}) \times \text{Supp}(\{C\})} \\ &= \frac{0.5}{0.54} = 0.93\end{aligned}$$