Tutorial 3A: Support, Confidence & Lift in Association Rule Mining

In the following table (table 1), there are nine baskets containing varying combinations of milk, cheese, apples, and bananas.

Basket	Product 1	Product 2	Product 3
1	Milk	Cheese	
2	Milk	Apples	Cheese
3	Apples	Banana	
4	Milk	Cheese	
5	Apples	Banana	
6	Milk	Cheese	Banana
7	Milk	Cheese	
8	Cheese	Banana	
9	Cheese	Milk	

The next step is to determine the relationships and the rules. For explanation purposes, the following table shows some of the relationships. In total there are 22 rules for the nine baskets.

Basket		How many Baskets Containing The product	Total # Baskets	Support	Confidence	Lift
CALCULATIONS →		(A&B)	Total	(A&B)/Total	(A&B)/(A)	[(A&B)/(A)]/[B/Total]
	Milk	6	9	0.666666667		
	Cheese	7	9	0.777777778		
1	Milk >>Cheese	6	9	0.666666667	1	1.285714286
	Apple, Milk	1	9	0.111111111		
2	(Apples, Milk) >> Cheese	1	9	0.111111111	1	1.285714286
3	(Apples, Cheese) >>Milk	1	9	0.111111111	1	1.5
	Apple, Cheese	1	9	0.111111111		

The first measure called the support is the number of transactions that include items in the {A} and {B} parts of the rule as a percentage of the total number of transactions. It is a measure of how frequently the collection of items occur together as a percentage of all transactions. The support formula written out would look something like:

$$\textbf{Support} = \frac{(\textbf{A} + \textbf{B})}{\textbf{Total}}$$

$$\textbf{Support for Basket 1} = \frac{(\textbf{Milk} + \textbf{Cheese})}{\textbf{Total}} = \frac{6}{9} = .6666667$$

Interpreted as:

Fraction of transactions that contain both A and B.

The second measure called the confidence of the rule is the ratio of the number of transactions that include all items in {B} as well as the number of transactions that include all items in {A} to the number of transactions that include all items in {A}.

The confidence formula written out would like something like:

$$\text{Confidence} = \frac{(A + B)}{A}$$

$$\text{Confidence for Basket 1} = \frac{(\text{Milk} + \text{Cheese})}{\text{Milk}} = \frac{6}{6} = 1.000$$

Interpreted as:

How often items in B appear in transactions that contain A only.

The third measure called the lift or lift ratio is the ratio of confidence to expected confidence. Expected confidence is the confidence divided by the frequency of B. The Lift tells us how much better a rule is at predicting the result than just assuming the result in the first place. Greater lift values indicate stronger associations.

The lift formula written out would look something like:

$$\text{Lift} = \left(\frac{\left(\frac{(A + B)}{A} \right)}{\left(\frac{B}{\text{Total}} \right)} \right)$$

$$\text{Lift for Basket 1} = \left(\frac{\left(\frac{(\text{Milk} + \text{Cheese})}{\text{Milk}} \right)}{\left(\frac{\text{Cheese}}{\text{Total}} \right)} \right) = \left(\frac{\left(\frac{6}{6} \right)}{\left(\frac{7}{9} \right)} \right) = \left(\frac{1}{.7777778} \right) = 1.2857$$

Interpreted as:

How much our confidence has increased that B will be purchased *given* that A was purchased.