UNIVERSITY of WASHINGTON

A Priori Rule Mining (Basket Analysis)

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Association Rules

Given subgroups of items or experiments, we want to find the most likely group given an initial item(s).

- If a customer buys car insurance for a minivan, (s)he is likely to buy car insurance for a second car.
- If a patient has condition x and y, they are likely to have condition w and z.
- If a customer buys bread and milk, they are very likely to buy eggs.

The last example is where the term “Basket Analysis” originates from.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Data Example

We can put customer transactions into a matrix:

Transaction	Purchases
1	Bread, milk, eggs, beer
2	Beer, ping pong balls, cups
3	Eggs, cups, bread
4	Beer, ping pong balls, wine

Transaction	Bread	Milk	Eggs	Beer	PingPong Balls	Cups	Wine
1	1	1	1	1	0	0	0
2	0	0	0	1	1	1	0
3	1	0	1	0	0	1	0
4	0	0	0	1	1	0	1

Association Rules

Transaction	Bread	Milk	Eggs	Beer	PingPongBalls	Cups	Wine
1	1	1	1	1	0	0	0
2	0	0	0	1	1	1	0
3	1	0	1	0	0	1	0
4	0	0	0	1	1	0	1

Let S be the set of all possible purchases, and n be the number of transactions.

Each rule can be written: $(x_1, x_2, \ldots, x_j) \rightarrow (y_1, y_2, \ldots, y_k)$

Where x and y are elements of S.

Given a specific rule, we can write the 'Support' of the rule:

$$Supp((x_1, x_2, \ldots, x_j) \rightarrow (y_1, y_2, \ldots, y_k)) = \frac{\#trans(x_1, x_2, \ldots, x_j, y_1, y_2, \ldots, y_k)}{n}$$

$$Supp(bread \rightarrow milk) = \frac{1}{4}$$

Interpret as 'The proportion of transactions that contain all the items

Association Rules

Transaction	Bread	Milk	Eggs	Beer	PingPongBalls	Cups	Wine
1	1	1	1	1	0	0	0
2	0	0	0	1	1	1	0
3	1	0	1	0	0	1	0
4	0	0	0	1	1	0	1

Given a specific rule, we can write the 'Confidence' of the rule:

$$Conf((x_1, x_2, \ldots, x_j) \rightarrow (y_1, y_2, \ldots, y_k)) = \frac{Supp(x_1, x_2, \ldots, x_j, y_1, y_2, \ldots, y_k)}{Supp(x_1, x_2, \ldots, x_j)}$$

$$Supp(bread \rightarrow milk) = \frac{1}{4}$$

$$Conf(bread \rightarrow milk) = \frac{0.25}{Supp(bread)}$$

$$Conf(bread \rightarrow milk) = \frac{0.25}{0.5} = 0.5$$

This is interpreted as how good of a predictor the rule is.

Association Rules

To even start considering a rule, we impose that it must have a minimum support.

- I.e., the items must appear together a minimum # of times.

We also want strong rules, so we specify a minimum confidence as well.

Association Rules

Support and confidence does not mean that it will have a big impact. To look at impactful rules, we consider the 'lift':

$$\text{Lift}((x_1, x_2,\ldots, x_j) \rightarrow (y_1, y_2, \ldots, y_k)) = \frac{Supp(x_1,x_2,\ldots, x_j,y_1,y_2,\ldots,y_k)}{Supp(x_1,x_2,\ldots, x_j)\times Supp(y_1,y_2,\ldots, y_k)}$$

If the association of x and y happen by chance, we would expect this lift term to be around or less than 1.

If lift > 1, then there is a positive correlation between the two groups.

Association Rules

Sometimes association rules are not helpful.

- Customers who buy car warranties also buy cars.

Searching all combinations of rules is computationally intensive, so we use an algorithm called "A Priori".

- We restrict our search to item sets that have a minimum support.
- Also, we know: $Supp(x_1, x_2,\ldots, x_j, y_1) \leq Supp(x_1, x_2,\ldots, x_j)$

Summary

Support
- The proportion of transactions that contain all the items

Confidence
- how good of a predictor the rule is

Lift
- Measures the impact of support & confidence

Correlation
- If lift > 1, then there is a positive correlation between the two groups

A priori
- Restrict the search to have a minimum support

A Priori Rule Mining

Concept Review

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON