Apriori based Association Rule Mining

Objective

Use the given data set to generate association rules using Apriori algorithm.

Association Rule Mining using Apriori –

The Apriori node discovers association rules in the data. Association rules are statements of the form

if antecedent(s) then consequent(s)

Support refers to the percentage of records in the training data for which the antecedents are true. **Confidence** is based on the records for which the rule's antecedents are true and is the percentage of those records for which the consequent(s) are also true.

Building Apriori model

1. Read in the data. Identify the fields that represent a product (we are mining association rules in between products), set their role as "Both" in Types tab. For remaining fields mark role as "None".
2. Connect it with a *"Apriori"* node from *Modeling* tab of nodes palette.
3. On *Fields* tab, select all fields in *Consequents* as well as *Antecedents*. (Goal is to extract every possible rule).
4. On *Model* tab uncheck *Use partitioned data* and supply *Min Support%* and *Min Rule Confidence%* as 10% both.
5. Select *Simple* mode on *Expert* tab. *Run* the stream and see the model nugget. Use filtering options provided there to select rules having support and confidence % more than 50%.
6. Connect source node to a "Web Node" from Graph tab of nodes palette. Select all product fields, check "Show True Flags Only" and select Line Values are "Overall Percentage". Specify appropriate values for "Weak links below" and "Strong links above" parameters and Run the node.
7. Observe the web carefully. Strong links will result in bi-directional rules coming out of Apriori node.

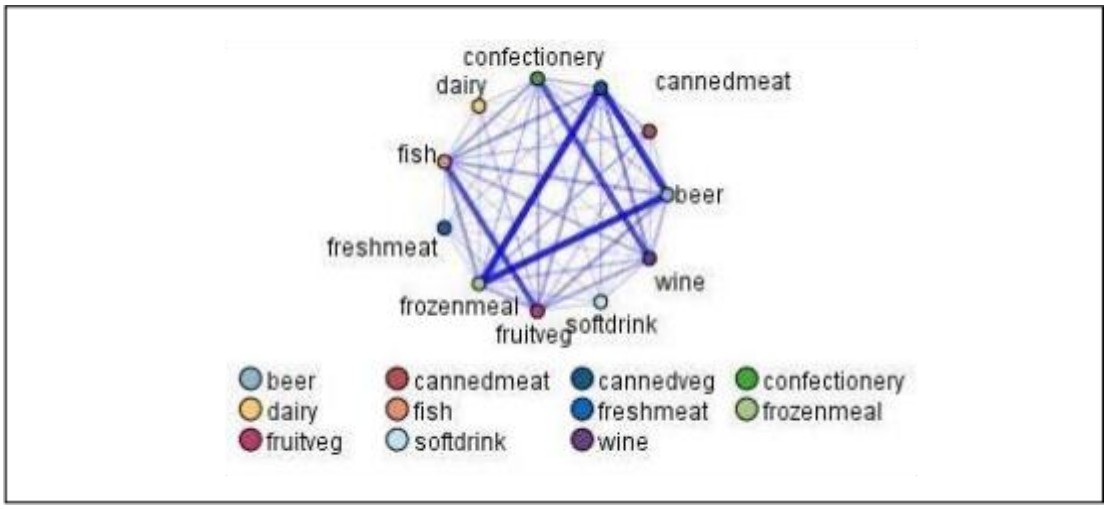

Fig 1- Output of a web node

Assignment

1. Go to *Expert* tab of model and select mode as *Expert*. Generate association rules through different *Evaluation measures* -

Evaluation Measure	**Description**
Rule confidence	The default method uses the confidence (or accuracy) of the rule to evaluate rules
Confidence difference	It is the absolute difference between the rule's confidence and its prior Confidence. Helps in discarding obvious rules
Confidence ratio	It is the ratio of rule confidence to prior confidence. It find the rules that predict the rare events.
Information difference	It is based on the concept of Information gain. It takes support into consideration such that the rules which cover more number of records are preferred for a particular level of confidence.
Normalized Chi-square	It is a statistical index of association between antecedents and Consequents. Depends on support.

2. Association rule mining for multidimensional data for categorical attributes.

Steps:

1. Use the zoo data in the source node. Select the correct data type of the fields. Select the Role of field 1 and field 18 as "None" and the remaining as "Both".

2. Convert relational table into presence-absence table. Use the "Set to Flag" Node from "Field Ops". Select "field 14" to be converted as flag field. Specify True Value as "1" and False Value as "0".

3. Apply Apriori and analyze the rules.