Interactive Network Graph using Association Rule Mining in E-commerce

David Lysko Jessica Chan Vargas Ji Ho Ahn Kunwoo Hong

Summary:

E-commerce industry has grown significantly over the recent years generating billions of transactions every single day. We created an interactive network graph generated by association rule mining which discovers products commonly purchased together. The interactivity truly empowers users to tailor the results based on their specific needs, and the friendly UI enables users to easily interpret the results.

Data Source:

UCI Machine Learning Repository - “Online Retail Dataset”.

Characteristics:

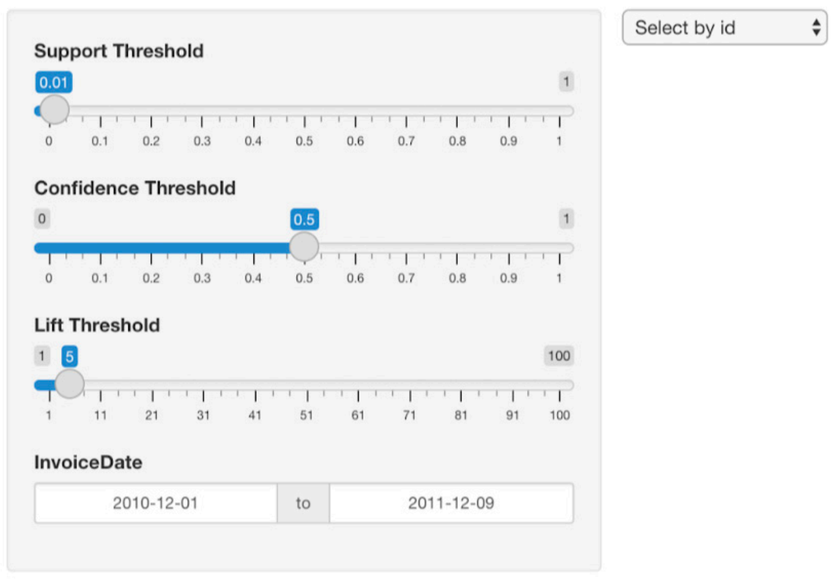
541,909 transactions which include: Invoice #, Stock_Code, Description, Quantity, Date, Price, Customer_ID, and Country (43.9MB)

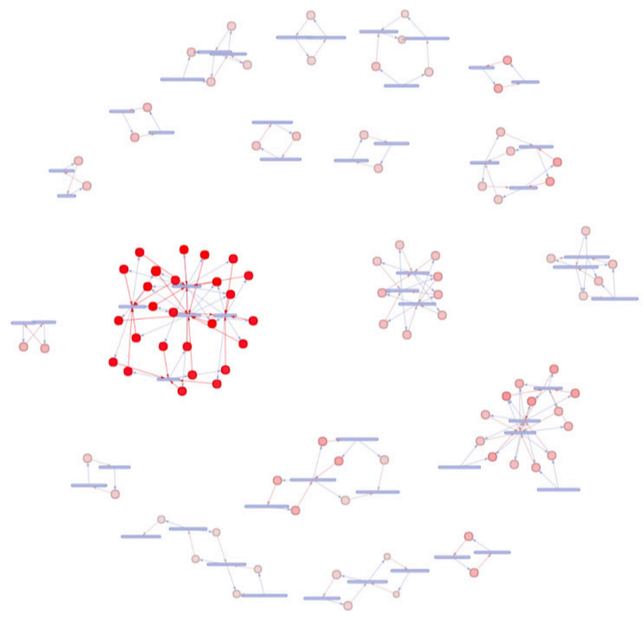
Who cares?

With increasing competition, it is essential for businesses to have a good understanding of their products and customers. Businesses can make strategic decisions based on the learnings from the association rules and offer discounts/ promotions to increase customer satisfaction.

Data Cleaning & Transformation Process:

1. Transform dates to a consistent format
2. Remove invalid product descriptions such as “Sale Error” or “AMAZON FEE”
3. Remove all columns except Invoice #, Description, Date, and Country
4. As a result, 519,419 transactional data remained (32.7MB)





Fine-tune your search to maximize your result:

Our application allows users to have full control over the association rules by enabling the ability to adjust thresholds for $Support_1$, $Confidence_2$, $Lift_3$, and Date.

1. Support: Proportion of all transactions containing items A & B
2. Confidence: Conditional probability that item A is purchased given that item B is purchased
3. Lift: Confidence / Support

Easily interpret your results:

Users are given a tabular view of the top 20 association rules which are shown below the network graph to easily understand which products are commonly purchased together along with the support, confidence, and lift calculations.

rules	support	confidence	lift
{HERB MARKER PARSLEY,HERB MARKER ROSEMARY} => {HERB MARKER THYME}	0.01	0.94	79.51
{HERB MARKER PARSLEY,HERB MARKER THYME} => {HERB MARKER ROSEMARY}	0.01	0.95	79.25
{HERB MARKER BASIL,HERB MARKER THYME} => {HERB MARKER ROSEMARY}	0.01	0.95	79.20
{HERB MARKER BASIL,HERB MARKER ROSEMARY} => {HERB MARKER THYME}	0.01	0.93	78.70
{HERB MARKER THYME} => {HERB MARKER ROSEMARY}	0.01	0.93	77.54

Experiments	Results
Add Country along with Support, Confidence, Lift, and Date filters	91% of the transactions in the dataset were from the UK, so Country was removed
Ensure that the filters are working correctly and association rules are properly displayed	When filtering for dates in December, our application displayed Christmas related items
Analyze the commonly purchased products	Top association rules contained similar products but in different variations such as green, pink, or rose pattern plates
Host the application through shinyapps.io allowing users to use the application online	Our application successfully deployed by limiting the dataset to 100,000 rows. The reason for the reduction is because the free version only allowed 1 GB of memory.

How do we compare to other methods?

Our application uses the apriori algorithm which is the most frequently used algorithm in association rule mining. What makes our application different from the traditional approach is that we focus on interactivity and ease of use where the traditional methods were static tables with limited rules.