

Association Learning

Association Analysis

- Simplest form of analysis
- Given a set of transactions, find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction
- Web usage mining, intrusion detection, continuous production, and bioinformatics.

Market Basket Analysis



Image Source: <https://www.analyticsvidhya.com/blog/2014/08/effective-cross-selling-market-basket-analysis/>

Concept

- Available Information

- List of purchases made by customer over the period of time (*No personal information of customer*)

- Problem Statement

1. Can we find purchase patterns?
2. What items tends to be purchased together?
3. Seasonal pattern of item purchase?



Benefits and Applications

1. Product suggestion to customers
2. Selecting items for promotion
3. Find promotional opportunity

Other Application Areas

1. Banking Services
2. Insurance Companies
3. Medical – Symptom Analysis

Association Analysis

- **Association Rule**

- An implication expression of the form $X \rightarrow Y$, where X and Y are itemset
- Example:
 $\{\text{Milk, Diaper}\} \rightarrow \{\text{Beer}\}$

Record	Items Bought
Record 1	Bread,Milk
Record 2	Bread,Diapers,Beer,Eggs
Record 3	Milk,Diapers, Beer,Cola
Record 4	Bread,Milk,Diapers,Beer
Record 5	Bread,Milk,Diapers,Cola

Purchase Profiles

Beauty
Conscious

Health Conscious

New Family

Personal Care

Sentimental

Smoker

Pet Lover

Casual Reader

Sports
Conscious

Tech
Enthusiast

Automotive

Photographer

Limitation

- Takes over 18 months to implement
- Only hypotheses, needs explicit testing
- Measurement of impact needed
- Difficult to identify product groupings
- Complexity grows exponentially

Association Analysis

Record	Items Bought
Record 1	Bread, Milk, Beer
Record 2	Bread, Diapers, Eggs
Record 3	Milk, Diapers, Beer, Cola
Record 4	Bread, Milk, Diapers, Beer
Record 5	Bread, Milk, Cola

Association Rule

$$\{X\} \Rightarrow \{Y\}$$

$$\{\text{Bread, Eggs}\} \Rightarrow \{\text{Milk}\}$$

Rule gives relation between Bread and Eggs versus Milk

Support - Rule Evaluation Metrics -

- Support (s) range[0,1]

Fraction of transactions that contain both X and Y

$$Supp(X) = \frac{\text{Occurance of } X}{\text{Total records in Dataset}}$$

$$\{\text{Milk}\} = 4/5 = 0.8$$

$$\{\text{Beer}\} = 3/5 = 0.6$$

$$\{\text{Beer, Milk}\} = 3/5 = 0.6$$

Confidence - Rule Evaluation Metrics

- Confidence (c) range[0,1]

Measures how often items in Y appear in transactions that contain X

$$conf(X \rightarrow Y) = \frac{Supp(X \cup Y)}{Supp(X)}$$

$$\{\text{Milk}\} \Rightarrow \{\text{Beer}\} = 3/4 = 0.75$$

$$\{\text{Milk}\} \Rightarrow \{\text{Bread}\} = 3/4 = 0.75$$

$$\{\text{Beer}\} \Rightarrow \{\text{Milk}\} = 3/3 = 1$$

Lift - Association Rule Metrics

- Lift (lift) range $[0, \infty]$

How much more often the antecedent and consequent of a rule

$$\text{lift}(X \rightarrow Y) = \frac{\text{Supp}(X \cup Y)}{\text{Supp}(X) * \text{Supp}(Y)}$$

$$\{\text{Milk}\} \Rightarrow \{\text{Beer}\} = 0.6 / (0.8 * 0.6) = 1.25$$

$$\{\text{Milk}\} \Rightarrow \{\text{Bread}\} = 0.6 / (0.8 * 0.8) = 0.93$$

$$\{\text{Beer}\} \Rightarrow \{\text{Milk}\} = 0.6 / (0.6 * 0.8) = 1.25$$

1 Implies both are Independent

Conviction - Association Rule Metrics

- Conviction (conv) range[0, ∞]

A high conviction value means that the consequent is highly depending on the antecedent.

$$\text{conv}(X \rightarrow Y) = \frac{1 - \text{supp}(Y)}{1 - \text{conf}(X \rightarrow Y)}$$

{Milk} => {Beer} = 1.60

{Milk} =. {Bread} = inf

{Beer} => {Milk} = 0.8

Available Algorithms

1. Apriori Algorithm
2. Eclat Algorithm
3. FP-growth Algorithm

Apriori algorithm

- Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases.
- It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database.

Transaction Table (Transaction Encoder-Apriori)

	Beer	Bread	Cola	Diapers	Eggs	Milk
Record 1	T	T	F	F	F	T
Record 2	F	T	F	T	T	F
Record 3	T	F	T	T	F	T
Record 4	T	T	F	T	F	T
Record 5	F	T	T	F	F	T

AssociationLearning MarketBasketAnalysis.ipynb

References

1. https://en.wikipedia.org/wiki/Association_rule_Learning
2. <http://cbafiles.unl.edu/public/cbainternal/facStaffUploads/KM3MarketBasket.ppt>
3. https://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/