Association Analysis in Detail

After this video you will be able to..

- Define the terms ‘support’ and ‘confidence’
- Describe the steps in association analysis
- Explain how association rules are formed from item sets

Association Analysis Steps

1. **Create item sets**

   {bread} {butter} {bread, milk} {bread, beer}

2. **Identify frequent item sets**

   {bread} {bread, beer}

3. **Generate rules**

   {bread, milk} => {diapers}

Analysis Association Dataset

ID	Items
1	diapers, bread, milk
2	bread, diapers, beer, eggs
3	milk, diapers, beer, butter
4	bread, milk, diapers, beer
5	bread, milk, diapers, butter

Item Sets

{bread, milk} => {diapers}
{milk) => {bread}

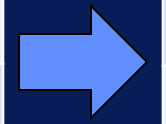
Rules

If bread and milk are bought, then diapers are also bought

Adapted from http://www-users.cs.umn.edu/~kumar/dmbook/index.php

Create Item Sets

ID	Items
1	diaper, bread, milk
2	bread, diaper, beer, eggs
3	milk, diaper, beer, butter
4	bread, milk, diaper, beer
5	bread, milk, diaper, butter



1-Item Sets

Item	Support
bread	4/5
butter	2/5
milk	4/5
beer	3/5
diaper	5/5
eggs	1/5

Support = frequency of item set

‘diaper’ occurs in all transactions

‘eggs’ occurs only once, in transaction 2

Create Item Sets

ID	Items
1	diaper, bread, milk
2	bread, diaper, beer, eggs
3	milk, diaper, beer, butter
4	bread, milk, diaper, beer
5	bread, milk, diaper, butter

minimum support = 3/5

1-Item Sets

Item	Support
{bread}	4/5
{butter}	2/5
{milk}	4/5
{beer}	3/5
{diaper}	5/5
{eggs}	1/5

Remove these item sets since they have low support.

Create Item Sets

ID	Items
1	diaper, bread, milk
2	bread, diaper, beer, eggs
3	milk, diaper, beer, butter
4	bread, milk, diaper, beer
5	bread, milk, diaper, butter

minimum support = 3/5

2-Item Sets

Item	Support
{bread,milk}	3/5
{bread,beer}	2/5
{bread,diaper}	4/5
{milk,beer}	2/5
{milk,diaper}	4/5
{beer,diaper}	3/5

1-item sets:
{bread}, {milk}, {diaper}

‘beer’ and ‘diaper’ occur together 3 times, in transactions 2, 3, & 4

Create Item Sets

minimum support = 3/5

ID	Items
1	diaper, bread, milk
2	bread, diaper, beer, eggs
3	milk, diaper, beer, butter
4	bread, milk, diaper, beer
5	bread, milk, diaper, butter

2-Item Sets

Item	Support
{bread,milk}	3/5
{bread,beer}	2/5
{bread,diaper}	4/5
{milk,beer}	2/5
{milk,diaper}	4/5
{beer,diaper}	3/5

1-item sets:
{bread}, {milk}, {diaper}

bread milk beer= 1/5
milk, diaper beer =2/5

Remove these item sets since they have low support.

Create Item Sets

ID	Items
1	diaper, bread, milk
2	bread, diaper, beer, eggs
3	milk, diaper, beer, butter
4	bread, milk, diaper, beer
5	bread, milk, diaper, butter

minimum support = 3/5

3-Item Sets

Item	Support
{bread,milk, diaper}	3/5

Only 3-item set with support > minimum support

1-item sets:
{bread},
{milk},
{diaper}

2-item sets:
{bread,milk},
{bread,diaper},
{milk,diaper},
{beer,diaper}

Frequent Item Sets

ID	Items
1	diaper, bread, milk
2	bread, diaper, beer, eggs
3	milk, diaper, beer, butter
4	bread, milk, diaper, beer
5	bread, milk, diaper, butter

1-Item Sets

Item	Support
{bread}	4/5
{milk}	4/5
{diaper}	5/5

minimum support = 3/5

2-Item Sets

Item	Support
{bread,milk}	3/5
{bread,diaper}	4/5
{milk,diaper}	4/5
{beer,diaper}	3/5

3-Item Sets

Item	Support
{bread,milk, diaper}	3/5

Rule Terms

Rule

Antecedent **X → Y** Consequent

If X, then Y

Rule Confidence

support for X & Y together

support for X

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

Itemset Support

$$\text{supp}(X) = \frac{\#\text{ transactions with X}}{\text{total \# transactions}}$$

Rule Generation & Pruning

frequent item sets → association rules

each k-item set → 2^k-2 rules!

frequent item sets → significant rules

Use rule confidence to
constrain rule generation

Keep rule if confidence > minimum confidence

Rule Example

ID	Items
1	diaper, bread, milk
2	bread, diaper, beer, eggs
3	milk, diaper, beer, butter
4	bread, milk, diaper, beer
5	bread, milk, diaper, butter

3-Item Sets

Item	Support
{bread,milk, diaper}	3/5

min confidence = 0.95

conf (X → Y) = supp (X U Y) / supp (X)

Candidate rule: {bread,milk} → {diaper}

$$\text{conf} = \frac{\text{supp (bread,milk,diaper)}}{\text{supp (bread,milk)}} = \frac{3/5}{3/5} = \frac{3}{3} = 1.0$$

Candidate rule: {bread,diaper} → {milk}

$$\text{conf} = \frac{\text{supp (bread,diaper,milk)}}{\text{supp (bread,diaper)}} = \frac{3/5}{4/5} = \frac{3}{4} = 0.75$$

Association Analysis Algorithms

- Use different methods to make efficient:
  - item set creation
  - rule generation efficient
- Algorithms:

  Apriori FP Growth Eclat

Association Analysis Steps

- Item sets created from data
- Frequent item sets identified using support
- Rules generated from frequent item sets and pruned using confidence