Basic Terminology & Brute Force Method for mining association rules

Analytics Vidhya

Market Basket 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

Transaction 1	🍎 🍺 🍚 🍗
Transaction 2	🍎 🍺 🍚
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	🍼 🍺 🍚 🍗
Transaction 6	🍼 🍺 🍚
Transaction 7	🍼 🍺
Transaction 8	🍼 🍐

Example of Association Rules

{Baby Food} □ {Diapers}

{Milk, Rice} □ {Beer}

Analytics Vidhya

Support

Transaction 1	🍎 🍺 🍚 🍗
Transaction 2	🍎 🍺 🍚
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	🍼 🍺 🍚 🍗
Transaction 6	🍼 🍺 🍚
Transaction 7	🍼 🍺
Transaction 8	🍼 🍐

$$\text{Support}\{🍎\} = \frac{4}{8}$$

$$\text{Support}\{🍎, 🍺\} = \frac{3}{8}$$

Analytics Vidhya

Confidence

Transaction 1	🍎 🍺 🍚 🍗
Transaction 2	🍎 🍺 🍚
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	🍼 🍺 🍚 🍗
Transaction 6	🍼 🍺 🍚
Transaction 7	🍼 🍺
Transaction 8	🍼 🍐

$$\text{Confidence}\{🍎 \rightarrow 🍺\} = \frac{\text{Support}\{🍎, 🍺\}}{\text{Support}\{🍎\}} = \frac{3}{4}$$

$$\text{Confidence}\{🍺 \rightarrow 🍎\} = \frac{\text{Support}\{🍎, 🍺\}}{\text{Support}\{🍺\}} = \frac{3}{6}$$

Analytics Vidhya

Lift

Transaction 1	🍎 🍺 🍚 🍗
Transaction 2	🍎 🍺 🍚
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	🍼 🍺 🍚 🍗
Transaction 6	🍼 🍺 🍚
Transaction 7	🍼 🍺
Transaction 8	🍼 🍐

$$\text{Lift}\{🍎 \rightarrow 🍺\} = \frac{\text{Support}\{🍎, 🍺\}}{\text{Support}\{🍎\} \times \text{Support}\{🍺\}} = \frac{3}{\frac{1}{2}}$$

$$= 1$$

Analytics Vidhya

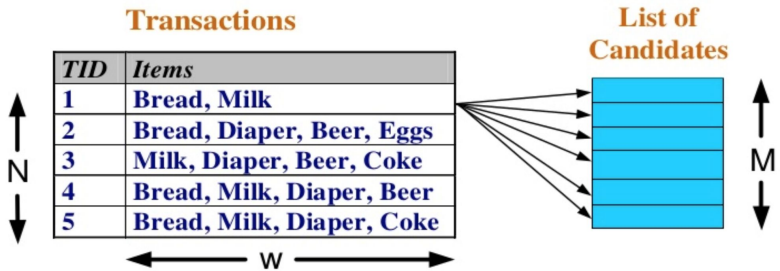
Terminologies (1)

- **Itemset**
  - A collection of one or more items (Example: {Milk, Bread, Diaper})
- **k-itemset**
  - a set of *k* items.
    - E.g. {beer, cheese, eggs} is a 3-itemset
    - {cheese} is a 1-itemset
    - {honey, ice-cream} is a 2-itemset
- **Frequent/Large Itemset (L_k)**
  - An itemset whose support is greater than or equal to a minsup threshold
- **Candidate Itemsets**
  - a set of *candidate* large *k*-itemsets.

Analytics Vidhya

Brute Force Method

- List all possible association rules
- Compute the support and confidence for each rule
- Prune rules that fail minimum support & minimum confidence thresholds
- Computationally expensive



Basket Data

Analytics Vidhya