Presentation on Machine Learning

Advanced Programming Language

Under the guidance of Prof. Chang

Presented by Anuradha - 阿洛拉
(M07158023)

Dated : 27th may, 2019.

Topic :

- **Association Rule Learning (ARL)**

Content of presentation

- Introduction
  - Data Mining
  - Associated Rule Learning (ARL)
  - Apriori model
- Explanation with Example
- Explanation of Program (How it works)
- Results & Interpretation
- Conclusion
- Reference

Data Mining

- The discovery of new information in terms of patterns or rules from vast amounts of data.
- The process of finding interesting structure in data.
- The process of employing one or more computer learning techniques to automatically analyze and extract knowledge from data.
- Data mining is actually one step of a larger process known as knowledge discovery in databases (KDD).
- The KDD process model comprises six phases
  - Data selection
  - Data cleansing
  - Enrichment
  - Data transformation or encoding
  - Data mining
  - Reporting and displaying discovered knowledge

- **Prediction**:
  - Determine how certain attributes will behave in the future.
- **Identification**:
  - Identify the existence of an item, event, or activity.
- **Classification**:
  - Partition data into classes or categories.
- **Optimization**:
  - Optimize the use of limited resources.

Aim of Data mining

Association Rule learning (ARL)

- Association rules are if-then statements that help to show the probability of relationships between data items within large data sets in various types of databases. Association rule mining has a number of applications and is widely used to help discover sales correlations in transactional data or in medical data sets.

- **What Association Rule Mining Aims to Achieve?**

  Association Rule Mining is one of the ways to find patterns in data.

  It finds:

- features (dimensions) which occur together
- features (dimensions) which are “correlated”

Association Rules

- Association rules are frequently used to generate rules from market-basket data.
  - A market basket corresponds to the sets of items a consumer purchases during one visit to a supermarket.
- The set of items purchased by customers is known as an item set.
- An association rule is of the form X=>Y, where X ={x_1, x_2, …., x_n }, and Y = {y_1,y_2, …., y_n} are sets of items, with x_i and y_i being distinct items for all i and all j.

For an association rule to be of interest, it must satisfy a minimum support and confidence.

Support:

The minimum percentage of instances in the database that contain all items listed in a given association rule.

Support is the percentage of transactions that contain all of the items in the item set .

Confidence:

Given a rule of the form A=>B, rule confidence is the conditional probability that B is true when A is known to be true.

Lift :

Lift controls for the *support* (frequency) of consequent while calculating the conditional probability of occurrence of {Y} given {X}.

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Total\ number\ of\ transactions}$$

$$Confidence(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Transactions\ containing\ X}$$

$$Lift(\{X\} \rightarrow \{Y\}) = \frac{(Transactions\ containing\ both\ X\ and\ Y)/(Transactions\ containing\ X)}{Fraction\ of\ transactions\ containing\ Y}$$

https://towardsdatascience.com/association-rules-2-aa9a77241654

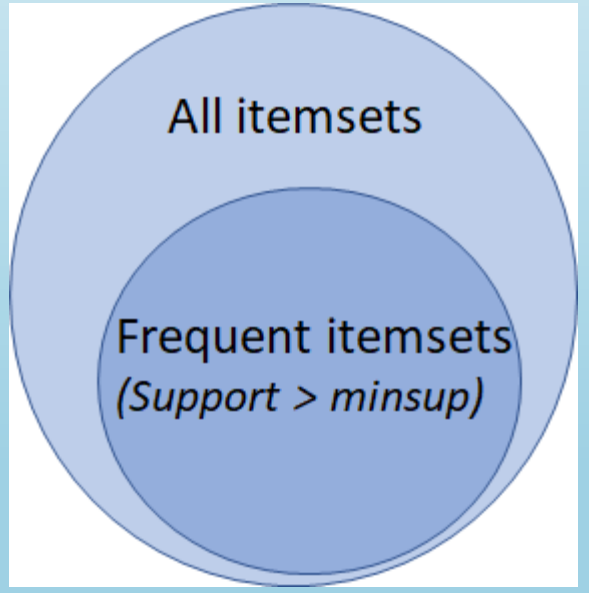
1. Association Rule: Ex. {X → Y} is a representation of finding Y on the basket which has X on it.

2. Item set: Ex. {X,Y} is a representation of the list of all items which form the association rule.

3. Support: Fraction of transactions containing the item set.

4. Confidence: Probability of occurrence of {Y} given {X} is present.

5. Lift: Ratio of *confidence* to baseline probability of occurrence of {Y}.



Apriori Intuition

“People who bought something also bought something else.”

Example :


Pampers
swaddlers
#1
N
20


XXXX
GOLD
FURPHY
REFRESHING ALE
Asahi
SOUKAI
爽快
Heineken
MID STRENGTH LAGER
PREMIUM QUALITY
PURE BLONDE

Apriori Algorithm works :-

Step 1 : Support

Movie Recommendation :

Support(M) = # user watch lists containing M / # user watch lists

Market Basket Optimisation :

Support(I) = # transactions containing I / # transactions

Step 2 : confidence

Movie Recommendation :

Confidence(M_1-M_2) = # user watch lists containing M_1 & M_2 / # user watch lists containing M_1

Market Basket Optimisation :

Confidence(I_1-I_2) = # transactions containing I_1 & I_2 / # transactions containing I_1

Step 3 : Lift

Movie Recommendation

$$\text{Lift}(M_1\text{-}M_2) = \frac{\text{Confidence}(M_1\text{-}M_2)}{\text{Support}\ (M_2)}$$

Market Basket Optimisation :

$$\text{Lift}(I_1\text{-}I_2) = \frac{\text{Confidence}(I_1\text{-}I_2)}{\text{Support}\ (I_2)}$$

Example

Five transactions from a supermarket

TID	List of Items
1	Beer,Diaper,Baby Powder,Bread,Umbrella
2	Diaper,Baby Powder
3	Beer,Diaper,Milk
4	Diaper,Beer,Detergent
5	Beer,Milk,Coca-Cola

Step 1

- Min_sup 40% (2/5)

C_1 ➔ L_1

Item	Support
Beer	"4/5"
Diaper	"4/5"
Baby Powder	"2/5"
Bread	"1/5"
Umbrella	"1/5"
Milk	"2/5"
Detergent	"1/5"
Coca-Cola	"1/5"

Item	Support
Beer	"4/5"
Diaper	"4/5"
Baby Powder	"2/5"
Milk	"2/5"

Step 2 and Step 3

C_2 ➔ L_2

Item	Support
Beer, Diaper	"3/5"
Beer, Baby Powder	"1/5"
Beer, Milk	"2/5"
Diaper,Baby Powder	"2/5"
Diaper,Milk	"1/5"
Baby Powder,Milk	"0"

Item	Support
Beer, Diaper	"3/5"
Beer, Milk	"2/5"
Diaper,Baby Powder	"2/5"

Step 4

C_3 ➔ empty

Item	Support
Beer, Diaper,Baby Powder	"1/5"
Beer, Diaper,Milk	"1/5"
Beer, Milk,Baby Powder	"0"
Diaper,Baby Powder,Milk	"0"

- Min_sup 40% (2/5)

Step 5

- *min_sup=40% min_conf=70%*

Item	Support(A,B)	Suport A	Confidence
Beer, Diaper	60%	80%	75%
Beer, Milk	40%	80%	50%
Diaper,Baby Powder	40%	80%	50%
Diaper,Beer	60%	80%	75%
Milk,Beer	40%	40%	100%
Baby Powder, Diaper	40%	40%	100%

Results

$$Beer \Rightarrow Diaper$$

- support 60%, confidence 70%

$$Diaper \Rightarrow Beer$$

- support 60%, confidence 70%

$$Milk \Rightarrow Beer$$

- support 40%, confidence 100%

$$BabyPowder \Rightarrow Diaper$$

- support 40%, confidence 70%

Interpretation

- Some results are believable, like baby powder and diaper bought together.
- Some results need more analysis, like milk and beer bought together.
- Some results are unbelievable, like diaper and beer.
- It contains unreal results because of the small data, so data interpretation plays a huge role in this.

Conclusion

From this concept of associated data learning (ARL), we learn many things : -

- How to short out the data from a number of data's.
- There are lot of information hidden in the data, if we can extract them well it will be very beneficial.
- The information can be extracted from the data on the daily basis without any interruption.
- We can make assumption and interpret the result by seeing the clear analysis of data.
- Data plays a huge role in marketing business, so if handled well, can make an effortless effect in the surrounding environment.

Reference

1. https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/
2. https://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining
3. https://github.com/topics/apriori-algorithm?l=python
4. https://en.wikipedia.org/wiki/Association_rule_learning
5. https://towardsdatascience.com/association-rules-2-aa9a77241654
6. https://www.analyticsindiamag.com/understanding-association-rule-learning-its-role-in-data-mining/
7. https://www.r-bloggers.com/association-rule-learning-and-the-apriori-algorithm/
8. https://github.com/topics/association-rule-learning
9. https://towardsdatascience.com/complete-guide-to-association-rules-2-2-c92072b56c84

Thank you