Apriori

Saturday, December 14, 2024 10:49 AM

ARL (Association Rule Learning)

- What is it all about?
  - People who bought also bought…
  - Analyzing when things come in pairs, and looking for those rules

User ID	Movies liked
46578	Movie1, Movie2, Movie3, Movie4
98989	Movie1, Movie2
71527	Movie1, Movie2, Movie4
78981	Movie1, Movie2
89192	Movie2, Movie4
61557	Movie1, Movie3

Potential Rules:

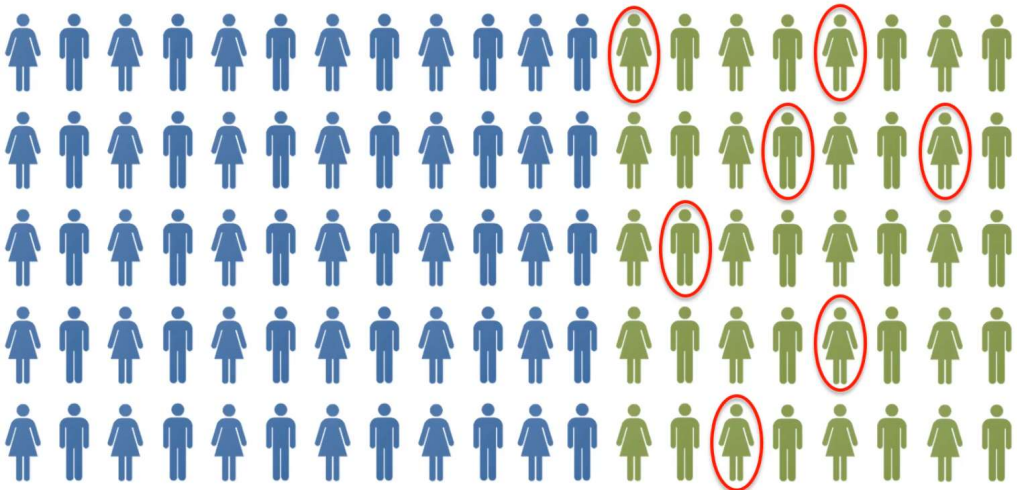
- Movie1 ⇨ Movie2
- Movie2 ⇨ Movie4
- Movie1 ⇨ Movie3

Some rules are going to be strong, and some are going to be weaker

Apriori - Support

- Movie Recommendation : $Support\ (M) = \frac{\#\ user\ watchlists\ containing\ M}{\#\ user\ watchlist}$
- Market Basket Optimization: $Support(I) = \frac{\#\ transaction\ containing\ I}{\#\ transactions}$
- Support is the percentage of people who did the first part
- Measure how frequently an item or combination of items appears in a data set

Apriori - Confidence

- Movie Recommendation: $Confidence\ (M1 \rightarrow M2) = \frac{\#\ user\ watchlists\ containing\ M1\ and\ M2}{\#\ user\ watchlists\ containing\ M1}$
- Market Basket Optimization: $Confidence(I1 \rightarrow I2) = \frac{\#\ transactions\ containing\ I1\ and\ I2}{\#\ transactions\ containing\ I1}$
- Testing a rule that people who have done one thing, have are also likely to do another thing
- Measures how often an association rule is true
- 
  - Given 40 people have seen movie A, and of the 40 people, 7 of them have seen Movie B, the confidence is 7 / 40 = 17.5%

Apriori - Lift

- Movie Recommendation: $lift\ (M1 \rightarrow M2) = \frac{confidence(M1 \rightarrow M2)}{support(M2)}$
- Market Basket Optimization: $lift\ (I1 \rightarrow I2) = \frac{confidence(I1 \rightarrow I2)}{support(I2)}$
- Confidence / Support
- Lift measures how much two items are to be associated with each other compared to if they were independent
- Higher lift is better

Apriori - Algorithm

- STEP 1: Set a minimum support and confidence
    - Don't waste your time building a model that don't really correlate with each other
- STEP 2: Take all the subsets in transactions having higher support than minimum support
- STEP 3: Take all the rules of these subsets having higher confidence than minimum confidence
- STEP 4: Sort the rules be decreasing left
    - The highest lift has the strongest association rule

Recommendation Systems like Amazon or Netflix is an example of Apriori, but much more complex. Apriori can be simple or very complex

Support - Frequency that items were associated with each other
Confidence - Out of all transactions that bought <left hand side>, the confidence level % also buys <right hand side>
Lift - The likelihood of <right hand side> being bought when <left hand side> is bought is lift times higher than if <right hand side> was bought alone

Eclat

Saturday, December 14, 2024 12:58 PM

Simplified version of Apriori

People who bought also bought…

In Eclat model we only have support

Elcat - Support

- Movie Recommendation: $Support\ (M) = \frac{\#\ user\ watchlists\ containing\ M}{\#\ user\ watchlists}$
- Market Basket Optimization: $Support\ (I) = \frac{\#\ transactions\ containing\ I}{\#\ transactions}$
- The frequency of items within the dataset

Elcat - Algorithm

- STEP 1: Set a minimum support
- STEP 2: Take all the subsets in the transactions having higher support than minimum support
- STEP 3: Sort these subsets by decreasing support