Business Data Analytics

Lecture 13

The slides are available under creative common license. The original owner of these slides is the University of Tartu
Are you being tempted?
Lecture 13: Cross-sell/Up-sell recommendations: Market Basket Analysis
What Walmart did later?

Sales on Friday evenings
But what exactly Walmart did?

Data -> Information -> Recommendation

How to find association among the entities from large datasets?
Recommendation Examples

• Product recommendation – like Amazon’s “customers who bought that, also bought this”.
• Music recommendations – like Last FM's artist recommendations.
• Medical diagnosis – like cool stuff.
• Content optimization – like in magazine websites or blogs.
How we can do recommendations?

Market

Basket

Analysis

MBA
Retailers can use the insights gained from MBA in a number of ways, including:

- Grouping products that co-occur in the design of a store’s layout to increase the chance of cross-selling;
- Driving online recommendation engines (“customers who purchased this product also viewed this product”); and
- Targeting marketing campaigns by sending out promotional coupons to customers for products related to items they recently purchased.
Market Basket Analysis (MBA)

• It is a technique or algorithm to identify the association rules from your data
• It is association rules with “lot of business” outcomes

• Input
  • List of purchases by customers over different visits

• Output
  • What items purchased together?
  • What items purchased sequentially?
  • What items purchased in seasons?

Association rules -> Generates rules
Example: (X -> Y)

Market Basket -> Assigns business outcome to those rules
Example: X,Y could be sold together
MBA: Terminologies

• **Items**: Objects that we are identifying associations between.

• **Examples**:
  - For an online retailer, each item is a product in the shop.
  - For a publisher, each item might be an article, a blog post, a video etc.

• A group of items is an item set.
  - \( I = \{i_1, i_2, i_3 \ldots, i_k\} \)
MBA: Terminologies Cont..

• **Items**: Objects that we are identifying associations between.

• **Transactions**: Transactions are instances of groups of items co-occurring together.

• **Examples**:
  • For an online retailer, a transaction is, generally, a group of items bought together.
  • For a publisher, a transaction might be the group of articles read in a single visit to the website.
  • **NOTE**: It is up to the analyst to define over what period to measure a transaction.

• For each transaction, we have an item set.
  • $t_n = \{i_1, i_2, i_3, ..., i_k\}$
MBA: Terminologies Cont..

• **Items**: Objects that we are identifying associations between.

• **Transactions**: Transactions are instances of groups of items co-occurring together.

• **Rules**: are statements of the form
  
  • \{i_1, i_2, i_3, \ldots\} \Rightarrow \{i_k\}
  
  • if you have the items in item set (on the left hand side (LHS) of the rule i.e. \{i_1, i_2, \ldots\}) , then it is likely that a visitor will be interested in the item on the right hand side (RHS i.e. \{i_k\}).

  • In the example above, rule would be:
    
    • \{flour, sugar\} \Rightarrow \{eggs\}
MBA: Terminologies Cont..

• **Items**: Objects that we are identifying associations between.

• **Transactions**: Transactions are instances of groups of items co-occurring together.

• **Rules**: are statements of the form

• **Output of MBA?**

Market basket analysis is generally a set of rules, that we can then exploit to make business decisions (related to marketing or product placement, for example).

Association rules -> Generates rules
Example: (X -> Y)

Market Basket -> Assigns business outcome to those rules
Example: X,Y could be sold together
MBA: Terminologies Cont..

- It works basically on following concepts
  - **Support**: \( P(A \cup B)/T \)
    - What is co-occurrence of two items name A and B
    - \( (\text{Co-occurrence of A and B})/\text{Total Tran.} \)
  - **Confidence**: \( P(A \cup B)/P(A) \): 
    - The proportion of transactions which contain A and also contain B.
    - How confident we are that B is present in presence of A.
    - Ratio of Support of (A and B), and Support of A.
  - **Expected Confidence**: 
    - How confident we are that B is present in absence of A.
    - \# transactions where B is present /Total transactions
MBA: Terminologies Cont..

- **Lift:**
  - Ratio of Confidence and Expected Confidence
  - Ratio of (B in presence of A) and (B in absence of A)
  - Explains the change in probability of B over “presence of A” and “absence of A”
  - Lift $\leq 1$
    - A has no impact on B
  - Lift $> 1$
    - Relationship between A and B is significant
    - Larger the lift ratio, the more significant the association.

Source: https://www.youtube.com/watch?v=WxDV9WEYqPw
MBA: Example

• A customer who buys a milk, also buys a bread. Let’s say
• Support – 5% (5% transactions have shown that milk and bread is bought together)
• Confidence – 70% (70% of the customers who bought milk, also bought bread)

• Lift

• Example
  • $P(A) = 0.4$, $P(B) = 0.7$, $P(\text{A} \cup \text{B}) = 0.3$
  • Lift: $0.3/(0.4 \times 0.7) = 1.071$
Retail Case Study

- Possible shopping baskets \((T)\)
  - Transaction 1: Beer, Diaper, Pretzels, Chips, Aspirin
  - Transaction 2: Diaper, Beer, Chips, Lotion, Juice, BabyFood, Milk
  - Transaction 3: Soda, Chips, Milk
  - Transaction 4: Soup, Beer, Diaper, Milk, IceCream
  - Transaction 5: Soda, Coffee, Milk, Bread
  - Transaction 6: Beer, Chips

- Frequent Items (based on \(M_s = 30\%\))
  - \((\text{Beer, Diaper})\) : with support 50\%
  - \((\text{Beer, Chips})\) : with support 50\%

- Rule 1: Beer \(\rightarrow\) Diaper
  - Confidence = \(3/4\) and Expected Confidence = \(3/6\)
  - Lift = \((3/4)/(3/6) = 2*(3/4) = 1.5\)

- Rule 2: Beer \(\rightarrow\) Chips
  - Confidence = \(3/4\) and Expected Confidence = \(4/6\)
  - Lift = \((3/4)/(4/6) = (3/2)*(3/4) = 1.1\)
Let us summarize about the problem*

- Generate set of rules that link two or more products together.
- Each of these rules should have a lift greater than one.
- Also, we are interested in the support and confidence of those rules:
  - Higher confidence rules are ones where there is a higher probability of items on the RHS being part of the transaction given the presence of items on the LHS.
- Recommendations based on these rules to drive a higher response rate.
  - We’re also better off *actioning* rules with higher support first, as these will be applicable to a wider range of instances.

*Problem: How to find rules which can help us in finding patterns?
MBA is through association rules

• We have to generate rules
• 3 Types
  • Actionable Rules: On which you can take action.
  • Trivial Rules: Interesting and need to do more research on.
  • Inexplicable Rules: Complex or incomprehensible (does not make sense)

Association rules -> Generates rules
Example: (X -> Y)

Market Basket -> Assigns business outcome to those rules
Example: X,Y could be sold together
How can we do this in R?

- R has an excellent suite of algorithms for market basket analysis in the **arules package** by Michael Hahsler and colleagues.
- Includes support for both the 1) **Apriori algorithm** and the 2) **ECLAT**
- **Apriori**
  - A level wise, breadth-first algorithm which counts transactions to find frequent itemsets and then derive association rules from them.
  - `Apriori()` in package arules
- **ECLAT** (equivalence class transformation algorithm).
  - Finds frequent itemsets with equivalence classes, depth-first search and set intersection instead of counting
  - `Eclat` in the same package
Let us Practice in R!
Demo time!

https://courses.cs.ut.ee/2018/bda/spring/Main/Practice