Homework 7: Walmart Coupon Optimization

Siqin Yang

Tackle the Problem

Walmart would like to offer coupons for cash register receipts so as to increase sales. Therefore, I think coupons should focus on items that customers are highly likely to buy, which means I should find out relationships among products and then make use of these relationships.

Model Design

- ✧ Model: An association rule model.
  - Association rule learning is used for discovering interesting relations between variables, which is implemented by exploiting measures of interestingness.
  - One classical method is Apriori algorithm, which could be summarized as below:

$$
\begin{aligned}
&\text{Apriori}(T, \epsilon) \\
&\quad L_1 \leftarrow \{\text{large } 1-\text{itemsets}\} \\
&\quad k \leftarrow 2 \\
&\quad \textbf{while } L_{k-1} \neq \emptyset \\
&\qquad C_k \leftarrow \{c = a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a, \{s \subseteq c \mid |s| = k-1\} \subseteq L_{k-1}\} \\
&\qquad \textbf{for transactions } t \in T \\
&\qquad\quad D_t \leftarrow \{c \in C_k \mid c \subseteq t\} \\
&\qquad\quad \textbf{for candidates } c \in D_t \\
&\qquad\qquad count[c] \leftarrow count[c] + 1 \\
&\qquad L_k \leftarrow \{c \in C_k \mid count[c] \geq \epsilon\} \\
&\qquad k \leftarrow k + 1 \\
&\quad \textbf{return } \bigcup_k L_k
\end{aligned}
$$

- ✧ Input: Cash register receipts data.
  - For each receipt, I would mainly make use of product names, which would be transformed to category names later.
  - I would use one-hot encoding to deal with categories.
- ✧ Output: Rules among certain combinations of products.
- ✧ Observations: Records about receipts. Each receipt is one record.

Below is the framework of data.

Receipt No.	Category 1	Category 2	…	Category n
1	1/0	1/0	…	1/0
2	1/0	1/0	…	1/0
3	1/0	1/0	…	1/0
……		……		

Proposed Framework

- First I would assign each product a category.
  - Finding relationships among products would be too broad and too costly. By assigning products categories, I could more efficiently find relationships.
  - Besides, coupons on specific categories would be more attracting than

coupons on specific products, as customers would have more freedom to choose which specific product to buy.
    - I would talk to the management to create the list of categories.
- Then I would use one-hot encoding to transform products to categories.
- With above preprocessing, I could build the association rule model and find out interesting relationships among categories.
    - I would only use data of recent 2 years, which should be enough to build the model, and much representative.
    - I would talk to the management on deciding parameters like minimum confidence and minimum support.
- Then I could make use of these relationships.
    - For each new receipt, by examining all products (and transforming them to predefined categories) I could find out if they fit any relationships.
        - Relationships that are fully satisfied are not considered any more.
            - For example, if the model tells that there exists a relationship among {Category A, Category B, Category C}, and the receipt has already all Category A, B and C in it, then this relationship is no longer useful for this receipt.
        - If there exists some unsatisfied relationships in the receipt, then I would make a list of all the "unsatisfied part".
            - "Unsatisfied item" is defined as the category that is part of the relationship, yet not appeared in the receipt.
                - For example, if the model tells that these exists a relationship among {Category A, Category B, Category C}, and the receipt has already Category A and B in it, then Category C is the "unsatisfied item" of this relationship for this receipt.
            - Since Walmart is only going to offer coupons on one product, I would only focus on relationships with only one "unsatisfied item".
            - I would rank all the "unsatisfied item" according to their unit profit (averaged among all products of the category the product belongs to), and offer the coupon on the most profitable category.
        - If there exists no relationship in the receipt, then I would just offer the coupon on the mostly-bought category.
            - The mostly-bought category is the one that has the highest frequency of appearing in receipts. In calculating frequency, I would take the amount of products bought into account.
- I would incorporate newly occurred receipts and update my model on a regular basis, in order to track updates.