

Assignment 02

Part_5: Conceptual Questions

Full Names: Achraf Cheniti, Vicky Leia, Farwa Sajjad

How does data sparsity affect performance?

Data sparsity is a situation with many missing or empty values in the data, which can occur when the dataset is not dense. In our case, this is common for our recommendation system because of the possibility of a small portion of possible interactions, such as user ratings, compared to the total possible combinations.

Since our recall measure evaluates how many relevant items a system can correctly identify. For our recommendation systems, it is about how many items that a user would like are actually recommended. data sparsity leads to lower recall because the system has less data to work with, which makes it harder to accurately predict which items the user would be interested in. For instance, if the user has only rated a few items (sparse data), the system might find it difficult to find and recommend other items that would be relevant to the user.

While recall gets affected the most in sparse data situations, other metrics like precision (how many of the recommended items are relevant) and F1-score (which combines precision and recall) can also be impacted. For example, Sparse data can lead to both low recall and potentially low precision, as the system might recommend items that the user isn't interested in due to a lack of interaction history.

What kinds of product bundles were discovered?

Methodology:

Using market basket analysis, we analyzed transaction data to discover frequent itemsets, or product combinations that frequently appear in customer purchases. The support metric, which indicates the frequency of a particular bundle relative to the total number of transactions, was used to determine the most common bundles.

Findings:

Through the analysis, several significant product bundles were discovered across various categories. The discovered bundles represent combinations of products that are frequently purchased together. Below are the top discovered product bundles:

1. Books:
 - a. Bundle 1: (P0050, P0006, P0052, P0011) – 4 products (Support: 0.06)
 - b. Bundle 2: (P0044, P0013, P0088, P0064) – 4 products (Support: 0.06)
 - c. Bundle 3: (P0044, P0088, P0079, P0064) – 4 products (Support: 0.06)

- d. Bundle 4: (P0099, P0098, P0088, P0083) - 4 products (Support: 0.06)
- 2. Electronics:
 - a. Bundle 1: (P0044, P0013, P0079, P0064) – 4 products (Support: 0.06)
 - b. Bundle 2: (P0079, P0013, P0088, P0064) – 4 products (Support: 0.06)
 - c. Bundle 3: (P0064, P0088, P0013, P0044, P0079) – 5 products (Support: 0.06)
- 3. Home:
 - a. Bundle 1: (P0070, P0039, P0015, P0022) - 4 products (Support: 0.06)
- 4. Clothing:
 - a. Bundle 1: (P0070, P0039, P0051, P0015) – 4 products (Support: 0.06)

Conclusion:

The analysis revealed several interesting product bundles across different categories. Notably:

- **Books** and **Electronics** show strong combinations of products that are commonly purchased together.
- **Home** and **Clothing** categories also revealed relevant bundles, although fewer were identified compared to Books and Electronics.

What improvements would you suggest for real-world deployment?

For real-world deployment of product bundle discovery, there are several areas where improvements can be made to enhance the accuracy, efficiency, and scalability of the solution.

1. Use More Advanced Algorithms

While we're already using Association Rule Learning such as Apriori Algorithm to identify frequent itemsets and product bundles, there are other ways like using Deep Learning by implementing a Neural Collaborative filtering Model to uncover non-linear patterns and relationships between products

2. Incorporate Customer Segmentation

Apply Customer Behavior Analysis that is based on tailoring bundles based on customer behavior and demographics. For instance, certain bundles may appeal more to frequent shoppers, while others may target first-time buyers

3. Data Enrichment

Incorporate External Data by integrating external data sources such as product reviews, social media trends, or inventory levels to refine recommendations. For example, trending products in a certain region or season might suggest new bundles. Furthermore, we can

also rely on User feedback to improve recommendations, which could be in the form of ratings, reviews or clickstream data that shows how customers interact with bundles.

4. Integration with E-commerce Platform

Ensuring that the product bundle recommendations are seamlessly integrated into the user interface of the e-commerce platform. For example, showing related product bundles or product pages at checkout, or in marketing emails. Moreover, integrating dynamic pricing for bundles based on factors like demand, seasonality, and inventory which can increase the attractiveness of the bundles.