ARISTOTLE UNIVERSITY OF THESSALONIKI
DEPARTMENT OF INFORMATICS

“Personalization of Supermarket Product Recommendations”

Panteli Iris 744
Pantelidou Kyriaki - Nektaria 765

Sequence of steps...

- Preprocessing
- Customer vector
- Product vector
- Similarity between customer and product vector
- Recommendations

Preprocessing

- Inputs:
  1. products-categorized.csv
     - → products
     - → classes
     - → subclasses
  2. groceries.csv
     - → customer_id
     - → basket

	product	class	subclass
0	abrasive cleaner	Paper-Cleaning-Home	Cleaners-Supplies
1	artif. sweetener	Condiments-Spice-Bake	Baking-Ingredients
2	baby cosmetics	Baby-Care	Baby-Bath-Skin-Care
3	baby food	Baby-Care	Formula-Baby-Food
4	bags	Paper-Cleaning-Home	Tableware-Trash-Bags
5	baking powder	Condiments-Spice-Bake	Baking-Ingredients
6	bathroom cleaner	Paper-Cleaning-Home	Cleaners-Supplies
7	beef	Meat-Seafood	Beef
8	berries	Fruits-Vegetables	Fresh-Fruits
9	bottled beer	Wine-Beer-Spirits	Beer-Ciders-Malts

Customer Vector (1)

- Creation of subclass vectors with countVectorizer.
- Aggregate the baskets of customers who had the same id.
- Assign its subclass vector to every product.
- Creation of basket vectors of the customers.
    - by aggregation of all subclass vectors of basket's products
- Normalization of that customer vectors.

Customer Vector (2)

$$\hat{\hat{c}}_{m,s} = \frac{\hat{c}_{m,s}}{\frac{1}{M} \cdot \sum_{m'=1}^{M} \hat{c}_{m',s}}$$

$$\hat{c}_{m,s} = \frac{c_{m,s}}{\sum_{s'=1}^{S} c_{m,s'}}$$

Product Vector (1)

- Use of Apriori algorithm for finding association rules between products
- Class/Subclass level
- $A \Longrightarrow B$: A is associated with B
- 9835 transactions (5-6 products on average in each)
- Parameters:
  - min support = $P(A \cap B) = 0.05$
  - min confidence = $P(B/A) = 0.3$
  - 44 rules for classes & 25 for subclasses

Product Vector (2)

$$P_s^{(n)} = \begin{cases} 1.0 & \text{if } s = \mathcal{S}(n) & \text{(within same subclass)} \\ 1.0 & \text{if } \mathcal{S}(n) \Longrightarrow s & \text{(within associated subclass)} \\ 0.5 & \text{if } \mathcal{C}(s) = \mathcal{C}(n) & \text{(subclass within same class)} \\ 0.25 & \text{if } \mathcal{C}(n) \Longrightarrow \mathcal{C}(s) & \text{(within subclass of associated class)} \\ 0 & \text{otherwise} & \end{cases}$$

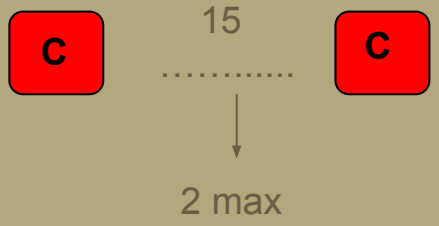
Similarity

The score $\sigma_{m,n}$ between customer m and product n is computed using a cosine coefficient between the corresponding vectors, $C^{(m)}$ and $P^{(n)}$:

$$\sigma_{m,n} = \rho_n \cdot \frac{C^{(m)} \cdot P^{(n)}}{\left\| C^{(m)} \right\| \cdot \left\| P^{(n)} \right\|} = 1 \cdot \frac{\sum_{i=1}^{k} \left(C_i^{(m)} \cdot P_i^{(n)} \right)}{\sqrt{\sum_{i=1}^{k} \left(C_i^{(m)} \right)^2} \cdot \sqrt{\sum_{i=1}^{k} \left(P_i^{(n)} \right)^2}}$$

Recommendations

- For every customer, we score all the products that are not contained in his basket.
- For every product assign its class/subclass.
- Group products by class/subclass.



- Aggregate all, sort by score and return top-10.

Thank you!