Data Mining Algorithms 2020/2021

Next Basket Recommendation

Yana Zabrodskaya

First attempt

The very first thing I thought about was to calculate a purchase frequency (in days) for each customer and each item and to memorize the last purchase date of this item.

What I wanted my program to do:
to estimate if it's time to buy this specific item based on the prediction date, the last purchase date, and the frequency of purchasing, i.e. customer buys milk every Friday. Prediction date is Friday. Should they buy milk? Probably.

Why it didn't work:
people are erratic, they don't buy milk every Friday. It also was impossible to measure the frequency if a customer bought an item only once, i.e. customer likes ice cream but buys a different one every time. What prediction did I want to get? An ice cream (doesn't matter which one for now). What prediction did I get? None.

Why spinach

Next time I tried to measure every item's popularity among all customers. As a result, I got the likelihood of being purchased for every item for each month.

Score for a *j^th^* item with a prediction date *d* calculates this way: $w_j = \sum \frac{1}{|d-d_i|+1}$, where d_i is the date for an i^{th} purchase.

What I wanted my program to do:
to recommend the most popular items among the ones the customer bought in the past.

Why it didn't work: spinach.
Spinach is the most purchased item. So, imagine there's a strange customer that buys the only one (and the same) item every time and it's never a spinach. But ONCE they made a mistake and bought a spinach. What prediction did I want? That the only and the same item. What prediction did I get? Spinach.

Almost perfect

Then I normalized the popularity with purchase frequency. *#id*$_j$ is an amount of j^{th} item among customer's purchases and *#total* is a total amount of purchased items:

$$w_j = w_j \cdot \frac{\#id_j}{\#total}$$

What I wanted my program to do:

to predict next purchase based on popularity AND frequency.

Why it didn't work:

imagine, a customer has been a poor student for 5 years and has been eating only ready-meal noodles for 7 kč a pack. However, they have a job for some time now! With a salary! Finally, they can afford noodles for 10 kč. What will they buy more likely? More expensive food. What does my program think? It's 5 years against a few months.

Best solution

I remembered about the last purchase date I had in the first part. So, I combined everything together.

$$w_j = \frac{w_j}{|last\_purchase_j - d|}$$

What I wanted my program to do:

to work correctly with a score of 0.197.

What my program does:

ignores spinach and adapts to changing preferences.