Project Proposal

July 9, 2019

1 DATA 612 Project Proposal: Instacart Recommendations

1.0.1 Introduction

Instacart in one of the most popular grocery ordering and delivery apps. After browsing and selecting products through the app, personal shoppers review orders and do in-store shopping and deliveries for customers. Due to the vast majority of available retailers, inventory size, and diverse customer base, it is very important for Instacart to assist its customers in navigating the wealth of available options by providing useful personalized recommendations. It is in this light that I hereby propose to build a recommender system using the Instacart Online Grocery Shopping Dataset 2017 that they have made publicly available in an effort to improve their own system.

1.0.2 The Data

The dataset is anonymized and contains a sample of over **3 million** grocery orders from more than 200,000 Instacart users. The data is a relational database separated into 6 tables describing customers' orders over time. For each user, between 4 and 100 of their orders are provided, with the sequence of products purchased in each order.

The dataset is provided as-is for non-commercial use, and may be accessed from https://www.instacart.com/datasets/grocery-shopping-2017, or from the Kaggle competition "Instacart Market Basket Analysis" at https://www.kaggle.com/c/instacart-market-basket-analysis/overview. The data will be cited as specified.

1.0.3 The Proposed System

The proposed system will be based on a hybrid approach that leverages collaborative filtering and dimensionality reduction techniques such as SVD due to the large size of the dataset. The hybrid approach will be based on minimizing the RMSE each methodology and then to minimize the overall RMSE of combined approaches by exploring different weights (for each approach). This system will be built on Spark using Python. Efforts will be made to build a web application to demonstrate the system in an online environment.

1.0.4 Sources

https://tech.instacart.com/3-million-instacart-orders-open-sourced-d40d29ead6f2
https://www.instacart.com/datasets/grocery-shopping-2017
https://www.kaggle.com/c/instacart-market-basket-analysis/data

Submitted by Albert Gilharry