The Intermediate Project Report of Product Recommendation Model

Xiaoting Zhou
Illinois Institute of Technology
Chicago, Illinois, USA
xzhou70@hawk.iit.edu
A20503290

Nevil Jack Denis
Illinois Institute of Technology
Chicago, Illinois, USA
psubramanian@hawk.iit.edu
A20474215

Hongbo Wang
Illinois Institute of Technology
Chicago, Illinois, USA
hwang218@hawk.iit.edu
A20524770

ABSTRACT

Product Recommendation Model (PRM) [1] is a process of identifying associations among entities and objects that frequently appear together, such as the collection of items in a shopper's Basket. When used appropriately, PRM can be an effective tool for Businesses/Companies in understanding consumer behavior better and influencing it. PRM is one of the key techniques used by large retailers that uncover associations between items by looking for combinations of items that occur together frequently in transactions [3]. In other words, it allows retailers to identify relationships between the items that people buy. For example, if customers are buying bread, how probably are they to also buy milk in the same transaction? This information may lead to increase sales by helping the business by doing product placement, shelf arrangements, upsell, cross-sell, and bundling opportunities. Association Rules are widely used to analyze retail basket or transaction data, is intended to identify strong rules discovered in transaction data using some measures of interestingness, based on the concept of strong rules [2]. In the process of online shopping, you have probably seen a section called "suggestions for you" or "customers who bought this item also bought" in which Product Recommendation Model plays an important role. The implementation of this analysis was aided by the initiation of electronic point of sales systems. Store owners used handwritten records and digital records of the customer transactions which were generated by point of sales system. This was effectively used to analyze ample amount of data to know about customer purchasing behavior and pattern. In this paper we are going to understand and help Instacart to make use of their customer transaction data and focus on descriptive analysis on the customer purchase patterns, items which are bought together and units that are highly purchased from the store to facilitate reordering and maintaining adequate product stock. Also, to identify the clusters and subgroups of customers possessing similar purchase behavior and to visualize the data to provide productive recommendations which focus on improving the revenue and customer experience through segmentation and prediction models. This paper will enable Instacart to enhance the user experience by suggesting the next likely product to purchase to the customer during the order process. Further, this paper will outline a marketing strategy for Instacart and similar retailers including sending personalized communications to customers reminding them to order again, by highlighting the predicted products in those communications.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

CCS CONCEPTS

• **Computing methodologies → Inductive logic learning**.

KEYWORDS

items, association, analyze, store, customer, transaction, data, predict

ACM Reference Format:
Xiaoting Zhou, Nevil Jack Denis, and Hongbo Wang. 2018. The Intermediate Project Report of Product Recommendation Model. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Online food orders are a novel method to refill on groceries and other necessities. Online grocery shopping is a hassle-free pastime that can be done at any time of day, including the wee hours of the morning. So what happens if you forget a few things as you're adding things to the cart or wish to receive better recommendations for your purchases? Will you place your order after a few hours of waiting? Users are given options based on their prior orders or user preferences to handle such circumstances.

With over 500 million products and 40000 locations, the grocery delivery and ordering app Instacart services customers in the United States and Canada. With the help of Instacart, you can receive product recommendations based on your prior purchases. We used transactional data from Instacart's historical customer orders to forecast which previously bought items will appear in a user's upcoming order. This information is provided as a Kaggle challenge and is open-sourced.

Understanding the transaction is a must to any form of business and its effect will lead to increase in sales. Especially in a retail store, it can be achieved by understanding the purchasing pattern of customer and related products which were sold together. This enables impulse buying from customers and also to understand their usual purchasing pattern and their effects towards the retail market.

Predicting which products will be in a user's future order is the goal. An anonymous sample of more than 3 million grocery orders from more than 200,000 Instacart users makes up the dataset. Between 4 and 100 orders from each user were made available by

Instacart, along with the order in which the items were added to the cart.

2 PROBLEM DESCRIBTION

According to a report by Marist Poll - Digital Economy Poll, approximately 76% of adults in the United States have shopped online, a trend that has been growing steadily over the years. Despite the benefits of e-commerce, however, statistics from Baymard Institute show that the average cart abandonment rate is a staggering 69.99%. One reason for this is that customers can become distracted by the sheer amount of options available to them, causing them to lose sight of their original intention. As a result, they may add unnecessary items to their cart, only to abandon them before finalizing their purchase and checking out. The product recommendation model is a method to solve the problem of user cart abandonment. The e-commerce giants have unique models to help them sell their products, like Amazon, eBay, and Walmart. However, these models cannot fit well in small e-commerce platforms. Large companies (Giants) utilize vast amounts of data to train their models, with an abundance of features. However, small e-commerce companies don't typically have access to data sets in the tens of billions necessary for model training. As a result, if they blindly apply a pre-existing model, over-fitting will show.

3 OVERVIEW

Don't you hate it when you go shopping and forget to pick up something you meant to? To address the relationship between what items to purchase and, as a result, to increase and improve the company's sales and comprehend consumer behavior. In this project, we develop capabilities of reordering data for specific products and user preferences for products. Additionally, we construct future orders based on the users' previous order records. For Example, everyone enjoys an Apple, so the metrics for each product should reflect the quality of a product getting reordered on its own merits. The reorder metrics for Mr. A should account for this preference as well as the reordering measures that are unique to him because Mr. A like some unusual meal that is rarely enjoyed by anyone else. Other metrics based on orders might include ordering patterns, preferred times (day of the week/hour of the day), etc. For Product Recommendation Model we are using support, confidence, and lift to understand the association rules of the products and items.

4 DATA PREPARATION/PRE-PROCESSING

4.1 Data Preparation

Information about customer purchases and transaction details are delivered to us through six different datasets. Order and Product dataset form the base of the complete transactions and were merged to a single dataset through the common Product and Order ID variables accordingly. Later, aisles and departments datasets were merged with the order and product combined dataset through aisle ID and department ID to form a master dataset to commence the analysis. SAS Studio was used for preparing the data and other manipulation operations to proceed further for the analysis.

The datasets were provided by Instacart Technology Company and was taken from Kaggle to perform the analysis. The datasets provided by Instacart had complete information of over 3 million grocery orders from more than 200,000 Instacart users. Both product data and customer data from Instacart includes 50,000 unique products, week and the time of purchase, different product aisle and departments. Understanding the data, dairy products, fruits and vegetables were purchased the most across all the departments and people tend to purchase and reorders 60% of their previous orders mostly on Sunday and Monday.

4.2 Data Preparation

The data that we used in this project are collected from Kaggle, Instacart Market Basket Dataset is found on Simplified Instacart Market Basket Dataset | Kaggle. The data set is a relational group of files that tracks the orders that customers place over time. An anonymous sample of more than 3 million grocery orders from more than 200,000 Instacart users make up the data set. Each user receives between 4 and 100 of their order details, including the order of the things they purchased in each order, the day and week it was placed, and the amount of time between orders.

4.3 Data Preparation

Orders (3.4m rows, 206k users). Column definition:

- `order_id`: order identifier.
- `user_id`: customer identifier.
- `eval_set`: which evaluation set this order belongs in.
- `The eval set has three values`: train, test and prior.
- `order_number`: the order sequence number for this user (1 = first, n = nth).
- `order_hour_of_day`: the hour of the day the order was placed on.
- `days_since_prior`: days since the last order, capped at 30 (with NAs for order_number = 1).

Products (50k rows). Column definition:

- `product_id`: product identifier.
- `product_name`: name of the product.
- `aisle_id`: foreign key.
- `department_id`: foreign key.

Aisles (134 rows). Column definition:

- `aisle_id`: aisle identifier.
- `aisle`: the name of the aisle.

Departments (21 rows). Column definition:

- `department_id`: department identifier.
- `department`: the name of the department.

Prior (30m+ rows). Column definition:

- `order_id`: foreign key.
- `product_id`: foreign key.
- `add_to_cart_order`: order in which each product was added to cart.
- `reordered`: 1 if this product has been ordered by this user in the past, 0 otherwis.

5 WHAT WE HAVE DONE SO FAR

- Collecting and preparing the Data.
- Project Analysis.

- ML Coding Language: Python.
- Encapsulate Built-in Functions Selection.
- Linear/Classification Model Selection.
- PyTorch Tutorials: learn the basics and introduction to PyTorch - Tensors.

6 WHAT REMAINS TO BE DONE

- Data Preprocessing: Apr 3, 2023
- Model Training: Apr 6, 2023
- Model Evaluation: Apr 13, 2023
- Parameter Tuning: Apr 18, 2023
- Making Predictions: Apr 24, 2023
- Final Project Report Writing: Apr 28, 2023

REFERENCES

[1] Zhen-Song Chen, Lan-Lan Yang, Rosa M Rodríguez, Sheng-Hua Xiong, Kwai-Sang Chin, and Luis Martínez. 2021. Power-average-operator-based hybrid multiattribute online product recommendation model for consumer decision-making. *International Journal of Intelligent Systems* 36, 6 (2021), 2572–2617.

[2] Haein Kim, Geunho Yang, Hosang Jung, Sang Ho Lee, and Jae Joon Ahn. 2019. An intelligent product recommendation model to reflect the recent purchasing patterns of customers. *Mobile Networks and Applications* 24 (2019), 163–170.

[3] Rong-Ho Lin, Wei-Wei Chuang, Chun-Ling Chuang, and Wan-Sin Chang. 2021. Applied big data analysis to build customer product recommendation model. *Sustainability* 13, 9 (2021), 4985.

Received 18 January 2023; revised 27 March 2023; accepted 31 March 2023