University of Virginia

Project proposal for University of Virginia
SYS 6014 Decision Analysis Spring 2020

Product-Recommender-System

Songminghuang Huang

sh4wd

April 12, 2020

1 Introduction

Online shopping is all over the internet. All our needs are just a click away. The biggest online shopping website is Amazon. Amazon is known not only for its variety of products but also for its strong recommendation system. For two decades, Amazon has been working to build a store with thousands of faces. Everyone who comes to Amazon sees it differently because the site is personalized to their personal interests. Just like when you walk into a store, the merchandise on the store shelves begins to re-arrange, placing you may need to be in front, you are unlikely to like to be in the back.Based on your current scenario and your past behavior, Amazon's recommendation system picks out a small number of items that may interest you from a library of hundreds of millions of products. The algorithm behind it is not magic, it just shares information that others have discovered with you. Everything is done automatically by algorithms. With the help of computers, people help each other implicitly and anonymously.

Now, I want to build an "Amazon-like" recommender system. In this project, the decision-maker is a shopping website who try to emulate Amazon's advertisement recommendation model. Nowadays, an increasing number of online companies are taking advantages of recommendation engines to attract user's attention and enrich shopping potential. Over the last 4-5 years, use cases of recommender system have been expanding rapidly across many fields, and this trend is expected to continue. Everyday people receive messages and emails about product or service they might be interested in, therefore they will feel easier to purchase the right product. Companies also make personalized recommendation to spread out good product and increase the profit.

The popularity and usefulness of recommendation motivate me to build a recommender system based on my knowledges

I select Amazon dataset because Amazon pays much attention collecting user data by asking them to rating their purchase and provide feedback to their experience. Therefore it's possible to gather plenty of data about each user as well as their purchase history. The more user data collected, the wider range of algorithm can be applied and the output would be more personalized.

2 The predictive tool

The ItemCollaborationFilter algorithm based on items (hereinafter referred to as ItemCF algorithm) pushes the recommendation system to an unprecedented scale of serving millions of users and processing millions of products. The success of this algorithm comes from the following aspects: simple and scalability, can often give surprising and useful recommendations, recommendations can be updated immediately based on new user information, interpretable. The tool can provide the recommended suggestion to the decision-maker, and the decision-maker will use the information generated by this tool to recommend the similar product to increase benefits and profits.

3 Model of the decision problem

3.1 The decision-maker's action sets $\mathbb{A}$

A complete itemization of all the options available to the decision-maker, from which the decision-maker selects a unique choice
In this project, from the dataset, the decision-maker is foucsing on the amazon review dataset considering the reviews and ratings given by the user to different products as well as his/her reviews about his/her experience with the product(s). To find the 2 most similar items, thus
$a^* \in \mathbb{A}$, the similar product can be recommended by the test data.

3.2 Sample space $\mathbb{X}$

The space from which observation data are drawn.
In the sample space $\mathbb{X}$, we can observe the below attributes: 'asin', 'date', 'categories', 'description', 'reviewe-text', 'reviewer', 'start-rating', 'price', 'developer'.

3.3 Data generating process

This might take the form of a statistical model describing the statistical properties of random variables $X_1, X_2 \ldots X_n \in \mathbb{X}$
Based on the below steps:

- Dataset processing
- KNN with LSH modeling

3.4 The parameter space Θ

The set of possible values for the unobserved parameters.
All possible values will depend on attribute, the only thing we believe is that it will make a small influence on our project.

3.5 The decision-maker's prior beliefs

These beliefs describe the information or beliefs the decision-maker has about the values of the unobserved parameters before collecting data. These beliefs can generally be represented as a probability distribution over the parameter space.
1. People's desire to buy is unlimited. You recommend product to them, they might buy. 2. The possibility of buying is 80

3.6 Payoffs

Payoffs will be a function of the selected action, and the realized value of an uncertain random variable. In this model, the select action based on the function: $Payoff = Accuracy_{recommendation} \times Profit_{recommendation} \times 0.8$

3.7 Utility function or Loss function

$$\begin{cases} 1, right \\ 0, false \end{cases}$$

3.8 The rule the decision-maker will use to choose a preferred action

Based on the accuracy of the model, the decision-maker will choose the prediction results.

4 The expected value of information

1) For shopping website, it can improve the return profit of advertisement

2) For buyers, it can let you buy more or better product you like

5 Results

6 Conclusions

7 References

- J. McAuley and J. Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. RecSys, 2013.
- http://jmcauley.ucsd.edu/data/amazon/
- https://www.kaggle.com/snap/amazon-fine-food-reviews