Refined concept note

Songminghuang Huang

February 2020

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.

1 The decision problem

In this project, the decision-maker is a shopping website who try to emulate Amazon's advertisement recommendation model. This project tries to solve problems confronting present recommended systems such as how to improve the degree of automation, how to make algorithms run faster and more robust with parallel computation.
Based on the Amazon reviews Dataset Amazon Review Dataset, I am 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).

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 Formalism

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 for Clothes, shoes and jewelleries and Beauty products and 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 most 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: u'asin', u'overall', u'reviewText', u'reviewTime', u'reviewerID', u'reviewerName', u'summary', u'unixReviewTime', u'HelpfulnessNumerator', U'HelpfulnessDenominator'.

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:

- Content-based Recommender System
- Collaborative Filtering
- Weight Learning
- Latent Factor Model
- Bias Extension
- Ensemble Model

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 big 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.
The data span a period of 18 years, including 35 million reviews up to March 2013. Reviews include product and user information, ratings, and a plaintext review.
Number of reviews : 34,686,770/Number of users : 6,643,669 (56,772 with more than 50 reviews)/Number of products : 2,441,053
Based on these input factors, sentiment analysis is performed on predicting the helpfulness of the reviews.

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:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2(y_i - \overline{y})^2}}$$

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

- Content-based Recommender System
- Collaborative Filtering
- Weight Learning
- Latent Factor Model
- Bias Extension
- Ensemble Model