Smarter Product Recommendation System for Bigbasket

Final Paper
Nov 2, 2018
Ali Taheri
University of the Pacific

"For such a model there is no need to ask the question "Is the model true?". If "truth" is to be the "whole truth" the answer must be "No". The only question of interest is "Is the model illuminating and useful?" George Box

INTRUDUCTION

Bigbasket is the largest online food and grocery store in India. It was launched 17 years ago in December 2011. It is operating in 25 big cities of India for providing daily grocery requirements of families. Bigbasket is offering about 18,000 products and more than 1,000 brands to his customers all over the country. They have 35,000 online orders per day on average that means approximately 1 million order monthly that is huge in a country same to India. With the online grocery market expects to grow at 55% CAGR (Compound Annual Growth Rate) over the next four years, optimization of all required processes is of vital importance.

They are offering a broad range of products to their customers, so Indian people could find all of their regular daily requirements in the website or mobile application of Bigbasket and then customers would place an online order that contains all of their demands. One essential factor that is very important is the convenient ordering of customers. Ordering must be straightforward for users, and even they must enjoy it to increase the level of customer satisfaction.

Problem Definition

Based on current experiences, Bigbasket has some challenges that owners are trying to resolve them. One of them is multiple orders of an individual customer per day that would increase the cost of logistics for Bigbasket due to the high level of traffic in big cities of India. Sometimes clients forget to include some of their needs, so they have to start a new order from Bigbasket, and even they would order it from other online store or purchase it from the local store that would decrease total revenue of Bigbasket. Another reason for multiple orders is the unorganized list of products. About 30% of clients are ordering by their mobile phone via application of Bigbasket. Finding all required products on the small display of a smartphone could take about 30 minutes, so busy customers would leave some products for next orders.

These multiple daily orders of customers would increase costs of Bigbasket's supply chain, so obviously it will decrease total revenue of Bigbasket. This issue has main business implications that need to be paid attention by business owners. Firstly, each order has to be delivered separately, and multiple orders to the same address would cause redundancy in packaging, delivery time, delivery cost and so on. Additionally, it needs inefficient planning of headcount and required resources of Bigbasket that would make a considerable increase in the operational cost of Bigbasket. Secondly, the average size of the basket would decrease because of multiple orders due to forgetfulness and untidy shopping platform, so the revenue of Bigbasket will decline. Finally, this inconvenient ordering could lead to potential overstocking of perishable items that can go bad in the warehouse because some of them would not be purchased by potential customers. Additionally, prediction and preparing of inventory for future would be very hard due to the inconsistent ordering of clients.

Because of mentioned business implications, Bigbasket has asked a solution based on recommendation system on their online ordering systems to address following improvements:

- All online customers will forget less to order their required products and groceries, so the average size of basket would increase, and the total number of daily orders for each customer would be almost one.
- Customers might make more purchases based on recommendations, and even they would buy something that will require shortly (next days).
- Number of Daily orders or even weekly orders to the same address would decrease, so the cost of logistics decreases, and total revenue improves.
- By this convenient ordering of customers, the team of warehousing in Bigbasket could predict the required products for all distribution centers more efficiently for the improvement of total profit.

After implementing a smart recommendation system, the ultimate goal could be to offer a subscription-based model where customers spend less time for ordering. The main concept is that Indian people used to purchase their requirements daily. This service would optimize their delivery time and cost drastically.

DATA Gathering and Preparing

There are different approaches to the implementation of recommendation systems same to KNN (K Nearest Neighbors) and collaborative filtering. Collaborative filtering has been selected as a more efficient algorithm for the recommendation system of Bigbasket. The concept of collaborative filtering is straightforward: “If Alice loves A and B, and Bob loves A, B, and C, then Alice is more likely to love C.” It means that if people have common interest about something, they might have common interest about other things too. For example, if two clients of Bigbasket like dairy products and one of them likes potato chips, maybe other one likes potato chips too.

The primary requirement of every machine learning algorithm is data. Data for collaborative filtering could be brought as a large matrix where each row represents a customer and columns represents products. Data of matrix is a rating that each customer gave to products of Bigbasket that could be in the range of 0 to 5. The primary source of data is Bigbasket and data engineering team of Bigbasket could provide purchasing history of all users. Additionally, they could deliver a database about customer ratings for their purchased products. Another source could be search and click history of customers. This search history is a measure that would show the interest of user on some products. For example, if a customer makes frequent searches on an individual product, it shows his/her interest in the product that could increase customer rating for that product.

Beside search history, the duration that individual customer spent on each page for viewing specification and reviews of other products could be helpful for estimation of customer interest on those products. In addition to all of the above data sources, history of searching and purchases of customers outside of Bigbasket would be great for the algorithm of collaborative filtering. It means access to purchase history of users in other online stores or some search activity of customer in search engines same to Google or even in social networks same to Facebook and Instagram.

Data needs to be prepared for feeding into the implementation of the recommendation system that is based on collaborative filtering. We need to fill the matrix with ratings of each customer for all of their purchased products. The first group of purchased products by an individual user is products which customer gave a rating to them before. The second group is purchased products without any rating by individual customer. The challenging problem is assigning a reasonable rating to that product on behalf of the individual customer. The simplest solution is sending the survey to every user and ask them about the level of their satisfaction (rating from 0 to 5) about their purchased products. Responses of customers could be inserted in matrix directly, and it would help algorithm to recommend products better than before. After this process, undoubtedly there would be a lot of purchased products by a customer without any rating. For addressing this issue, the first solution is the usage of average ratings of other members of the household for that product. It is the case because grocery requirements of an individual household's members are almost the same. Secondly, product ratings could be estimated by volume and frequency in the purchase history of the mentioned product by the customer. As an example, the rating of a branded milk that client purchases every week could be 4 or 5, because the weekly purchase of that milk shows the high interest of customer about that milk. Besides all of these procedures, we could find almost identical products by metrics same to "Cosine" and "Pearson" and share their ratings from individual customers. By all mentioned methods, we would assign ratings to almost all purchased products by a customer. The next step is the prediction of new preferences of customers based on all of their purchasing pattern and ratings.

Analysis and Solution

As mentioned before, rows of the matrix are representing users and columns are standing for products of Bigbasket. Each element of the rating matrix shows the level of user satisfaction from the product on a scale of 0 to 5. It is evident that many matrix items are blank, and the foremost problem is the prediction of those blank spots by existing information in non-blank spots as you could see in following table.

	Organics Milk- Amul	Noodle - Maggi	Chocolate Bar- Snickers	Walnuts- Borges	Shampoo - J&J	Battery - Duracell	Bread - Fresho	Cauliflower - Fresho	Diaper - Pampers	Suda - Kinley
Arda	3		2					5		
Alan	5		5		2	4	1	5	3	2
Ali	5	4	4		4	3	2	3		1
Brendon	4				1	4	3	5		
Brandon		1	1		3	4	1		2	3
Joe		2	2	3	1	2			2	3
Mike	4	4	5	4	1	5	2		1	
Max	3	1				4	1	4	1	4
Denise	3	3				3	1		3	
Christy		1		1		1	1			
Austin	1	3			3	2	5	4	2	
David	2	4	2	1		3	4		3	4
Mike			5	4		3	3	1		2
Jeremy	3	3	2					1		

For applying collaborative filtering, customers and products would be modeled by some parameters. For this target, each of them could be represented by a scalar as bias and a vector as embedding. It means that each customer would be represented by customer bias and customer embedding and each product could be modeled by product bias and product embedding. Customer bias could show optimism and pessimism of him or her. It means that an optimistic customer usually gives better ratings compared to a pessimistic customer. Product bias means that some products could receive higher ratings despite their low quality. As an example, inferior products of famous brands could get unreasonably high rating due to their brands. Embedding vector is capturing different characteristics of the entity it is representing. In the case of Bigbasket, embedding of a customer could be a vector that shows the level of interest of user on the different type of products. For example, the first element of user embedding could show the interest of customer on dairy products, the second one on fruits/vegetables and so on. Embeddings of products would show properties of products in the same order of customer embeddings. In the above example, the first element of product embedding would show to what extent that product has dairy nature and the second one represents the level of its dependence on fruits and vegetables. By this modeling, we could calculate the rating of a product from each customer by the dot product of both embeddings plus both biases. In other words:

Predicted rating = Embedding vector dot product (user, product) + user bias + product bias

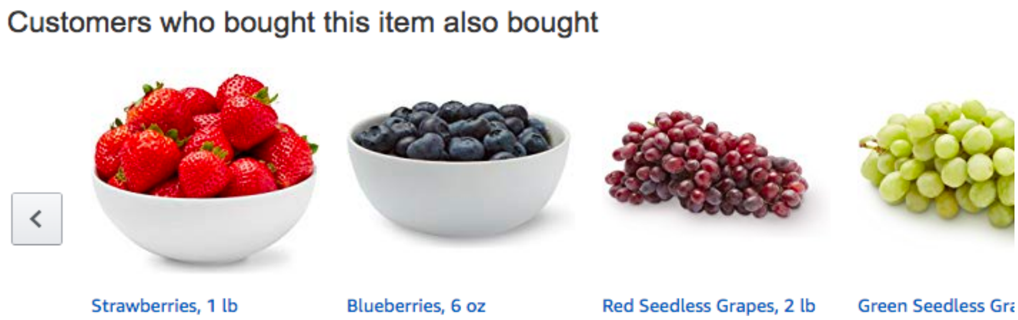
At first step, a dimension for embedding vectors would be considered as a hyperparameter. Then, random numbers would be assigned to all biases and embeddings for starting a collaborative filtering algorithm. At this phase, all existing ratings would be calculated by above formula, and total Root Mean Square Error (RSME) will be computed. Now it is the time to change existing values of biases and embeddings for decreasing and finally minimizing total RSME. Gradient descent could be an excellent choice for this optimization. All optimized values of biases and embeddings would appear after multiple steps that will lead to a local minimum for total RSME as following figure.

	Organics Milk- Amul	Noodle - Maggi	Chocolate Bar- Snickers	Walnuts- Borges	Shampoo - J&J	Battery - Duracell	Bread - Fresho	Cauliflower - Fresho	Diaper - Pampers	Suda - Kinley
Product Bias	1.09	1.24	1.38	1.28	0.85	0.44	-0.24	0.91	-0.13	-0.36
Product Embeddings	1.31	0.38	-1.36	-2.55	-0.92	2.75	-2.92	2.26	0.93	0.29
	2.05	-0.37	-1.09	0.57	1.28	2.13	2.91	0.21	-2.31	1.98
	2.30	-0.48	-0.63	-0.25	-1.85	-2.95	-1.09	1.71	-1.59	2.70
	0.29	1.09	1.66	2.34	-1.00	2.49	0.42	-1.30	1.10	1.69
	1.91	-2.99	1.57	-1.77	1.85	-0.97	-2.47	-1.49	-1.10	1.95

User Bias	User Embeddings						Organics Milk- Amul	Noodle - Maggi	Chocolate Bar- Snickers	Walnuts- Borges	Shampoo - J&J	Battery - Duracell	Bread - Fresho	Cauliflower - Fresho	Diaper - Pampers	Suda - Kinley
-0.18	0.81	1.49	-0.97	-0.48	-1.42	Arda	3.05	-1.76	1.86	-0.81	-1.39	-0.86	-0.93	4.64	-0.60	-0.33
1.86	-1.69	-1.88	0.78	0.81	0.59	Alan	4.54	-0.98	4.89	-1.81	1.85	3.65	1.06	4.92	2.66	2.26
-0.26	1.08	1.59	1.51	0.78	-0.33	Ali	4.22	2.67	4.43	-1.77	4.50	2.99	2.30	2.64	-0.29	1.08
0.27	1.23	0.93	1.57	-0.37	-0.02	Brendon	3.71	0.38	-0.33	-1.88	1.21	4.31	2.62	4.65	0.42	-1.59
0.17	1.83	0.11	-0.63	0.30	-1.38	Brandon	-0.49	-1.73	0.90	0.34	2.84	3.58	0.98	0.07	2.07	3.01
0.66	-1.67	1.97	-1.38	-1.34	0.82	Joe	1.01	1.62	1.78	2.98	0.97	1.78	0.22	-1.44	2.28	2.51
0.12	1.67	1.19	-1.62	-0.74	-1.18	Mike	4.27	3.58	4.83	3.51	1.23	4.05	1.58	-0.79	0.93	0.33
-0.84	-0.39	1.02	-0.25	1.07	1.75	Max	1.67	1.00	-1.63	-1.16	-0.06	3.74	1.07	3.63	1.03	3.93
-0.27	-0.30	0.93	-0.60	-1.20	0.70	Denise	2.50	-1.94	-1.67	-1.54	-1.92	2.60	0.80	-1.85	2.78	-0.92
-0.69	-0.09	-0.97	-0.75	0.91	-1.10	Christy	0.20	1.12	-0.23	0.65	-0.40	1.20	1.35	-0.13	-0.38	-0.44
0.85	-0.52	1.47	-0.43	-1.96	-0.76	Austin	1.18	2.50	-1.82	-0.13	2.81	1.67	4.90	3.97	1.62	-0.30
1.81	0.24	-1.06	-0.91	-0.06	-1.56	David	1.43	4.04	2.15	0.80	-0.18	2.51	3.62	-1.07	3.02	4.27
0.24	1.95	-0.60	-1.74	-0.41	-1.96	Mike	-1.98	-0.85	4.65	4.21	0.25	2.97	2.64	1.16	-0.85	2.28
1.50	0.17	-0.09	0.67	0.56	0.82	Jeremy	2.84	2.53	1.96	-0.18	-0.65	-0.42	-0.15	1.37	-0.80	-0.03

Business Actions

Now, all blank spots could be calculated by values of biases and embeddings so that Bigbasket will have an estimation about the interest of his customers about different products. Bigbasket could use assigned ratings and estimated ratings for offering products to his customers after their login in time of ordering by them. Bigbasket could use different strategies for recommending products to their customers. One of them is considering last purchases of customer and offering missed products in the new order which have the highest rating among all products of the old basket. Another solution could be attention to some recipes that consists at least one or more products of new basket and offering some products of those recipes with the highest estimated ratings. By these processing's, the proposed model for recommendations will appear to the customer as a list of products which other customers have purchased before beside products of the current basket of customer or in combination with the item being looked at them. Bigbasket could use some expressions same to "Customers who bought this item also bought" or "Frequently bought together" over or under figures of recommended products same to following figure.



By implementation of this recommendation model, Bigbasket would see improvements in two important areas. Firstly, its total revenue will go up by a significant increase in average basket size of clients and a considerable decrease in mean time for all orders. Secondly, its total cost will go down by a moderate decrease in their headcount and a substantial improvement of their logistic cost. Both improvements prove that carrying out of recommendation system is a worthwhile project for an online retailer same to Bigbasket.

Future of Bigbasket

Based on the news, Bigbasket has recently taken an important step towards improving the shopping experience of customers through the acquisition of "RainCan" which is a subscription-based e-grocery startup. At RainCan, the customer could select all of his or her regular grocery requirements same to milk, bread, fruits & vegetables, and coconut water. Then, the customer will configure the start date when RainCan will start delivery. At the next step, the customer decides which time-slot is convenient for him to receive products and he could choose intervals of every day, alternate day or every third day as the frequency of delivery. This subscription-based delivery could improve micro-delivery services for essential household items.

Also, they acquired a stake in a business called “Kwik24” which sets up smart vending machines for selling fresh products. Additionally, it seems owners of Bigbasket are talking to merge with their main competitors in order to better compete with Amazon and other popular online retailers. All of these changes and prospects of Bigbasket show that Bigbasket needs a tremendous big-data analytics platform for being a leader in this competitive market.

Resources

- https://economictimes.indiatimes.com/opinion/interviews/predictability-a-challenge-in-subscription-model-hari-menon-ceo-bigbasket/articleshow/60214720.cms
- https://www.exchange4media.com/marketing-news/the-biggest-challenge-is-to-get-the-supply-chain-rightmarketing-headbigbasket-89507.html
- https://economictimes.indiatimes.com/small-biz/startups/amazon-in-talks-to-buy-indian-online-grocer-bigbasket/articleshow/59130524.cms
- https://economictimes.indiatimes.com/small-biz/startups/grocery-startups-subscription-based-services-may-prove-cost-efficient/articleshow/60214386.cms
- https://www.youtube.com/watch?v=RYTbIrjrJyM&t=0m30s
- http://www.kwik24.com/
- https://www.thenewsminute.com/article/3-acquisitions-bigbasket-now-offer-micro-delivery-vending-kiosks-and-more-90286
- https://towardsdatascience.com/collaborative-filtering-and-embeddings-part-1-63b00b9739ce
- https://www.vccircle.com/bigbasket-buys-raincan-to-boost-micro-delivery-business