

Prediction of E-Commerce Product Ratings Based on Similar Users

Pola Shashank¹, M Venkatesh², K Ravi Chandra Reddy³, Dr.P.Indira Priyadarsini (Assoc Professor)⁴

^{1,2,3,4}Department of Information Technology, Vardhaman College of engineering, Hyderabad

Abstract

Together with the fast advancement of continuous expansion and the Internet of E-commerce scope, product quantity, as well as assortment, boost fast. Merchants offer many goods via going shopping customers and websites generally consider a huge amount of moment to discover the products of theirs. Within e-commerce sites, the item rating is among the primary key ingredients of an excellent pc user expertise. Many methods are working with whose users to consider the goods they wish. A comparable item suggestion is among the favorite modes working with whose customers look for items in line with the item scores. In general, the suggestions aren't personalized to a particular pc user. Exploring a great deal of solutions tends to make customers runoff as a result of the info clog but not offering proper reviews for solutions. Traditional algorithms has data sparsity and cold start issues. To overcome these problems we use cosine similarity method to identify the similarity between those vectors. The nearest similar vector ratings will be used during the estimation of the unknown ratings. The proposed methodology records ratings of each product from users and those are represented by a vector, and the cosine similarity is used a measure to identify the similarity between those vectors. The nearest similar vector ratings will be used during the estimation of the unknown ratings. Hence, By using the above approach it can overcome the above problems and also it can achieve high efficiency and accuracy in a simple manner

Keyword: Cosine similarity data sparsity cold start

Introduction

The E commerce business is replacing the going shopping encounters of individuals. They receive a great deal of options to choose products coming from numerous sites. However the primary headache sits with these. There are various items and users to address. A lot of e commerce businesses have stopped due to poor engagement of clients. With all the latest enhancements as well as inexpensive option of computational strength, a great deal of information is produced[1]. Appropriate utilization of this particular information might eventually result in the profits of an E commerce enterprise.

The information produced as a result of an E commerce site is equally unstructured and structured. The organized data consists of consumer grow older, place of the buyer, group of products purchased, rating info of people on things as well as Time and Date of purchase and browsing[2]. The patterns produced against organized structure isn't adequate to fully understand someone's you're shopping requirements not to mention there's a requirement to mine helpful patterns coming from unstructured platforms. They comprise critical reviews of people on products, person simply click patterns, tweets, social networking commentary as well as products seen by clients[3].

You'll find numerous kinds of E commerce industries. Amazon, Myntra, Paytm mall, Flipkart, that offer products straight and therefore are referred to as immediate E commerce sites[4]. These things are available as well as delivered towards the consumers as soon when the purchase is positioned. On the flip side, Music streaming sites as Amazon music, Video and Spotify streaming sites as Netflix, Amazon key video clip are additionally a kind of E commerce sites that interact the clients with great tips as well as cost

clients over a prepaid schedule. No matter what could be the type of E commerce business, the supreme objective is engaging clients as well as get great earnings.

Literature Survey

Among the powerful customization solutions operating the adaptive net is collaborative filtering. Collaborative filtering (CF) is the procedure of filtering or perhaps analyzing Items thru the views of other individuals. CF engineering creates in concert the views of big interconnected towns along with the net, supporting the filtering of sizable Quantities of information[6].

Computer systems and also the net enables us to progress beyond basic word-of-mouth. Rather than restricting ourselves to hundreds or tens of people, the web permits us to think about the views of countless numbers. The pace of computer systems permits us to think to do the opinions for real-time and find out not just what a significantly bigger society believes of a gadget, but additionally, produce a really personalized perspective of that particular product while using views best for a certain group or user of consumers.

For being much more professional, a score covers the connection of two points - item and user frequently making use of several worths. One method to imagine ratings as being a matrix. Without any loss of generality, a scoring matrix comprises a dining room table in which every row belongs to a user, each and every column belongs to a certain video, so the amount in the intersection of a column and a row belongs to the user's score printer. The lack of a score rating at this intersection suggests that the person hasn't but ranked the product. The word User describes any person that offers scores to a product[5]. Usually, we drive the phrase to relate to individuals by using a phone system to get info (e.g., recommendations) even though it also describes people who offered the information (ratings) used within producing the info.

Collaborative filtering methods create predictions or maybe suggestions for a certain pc user as well as 1 or maybe lot more things. can include anything at all for that a man can actually a piece score, like an art form, journal articles, CDs, books, or even getaway destinations. Scores within a collaborative filtering structure are able to undertake an assortment of types.

- Scalar reviews are able to include possibly numerical scores, including the 1 5 stars offered around MovieLens, or maybe ordinal reviews like firmly decide, disagree, neutral, agree, really don't agree.
- Binary score version options involving good/bad or agree/disagree.
- Unary reviews are able to signify that a person has noticed or even bought something, or even usually rated the product favorably.
- The lack of a score suggests we've absolutely no info relating to the end-user on the product (perhaps they bought the product someplace else).

Ratings might be gathered via explicit ways, implicit ways, or perhaps each. Explicit scores are all those in which a person is required to present a viewpoint on a gadget. Implicit ratings are all those inferred out of a user's measures. For instance, an end-user that visits a program page maybe has a bit of fascination with that particular item while a person that consequently buys the item might have a significantly more powerful fascination with that service

Proposed System

Fig.1 describes the proposed ensemble modeling architecture of our system. First step is identify null values then estimating those null values using cosine similarity method,

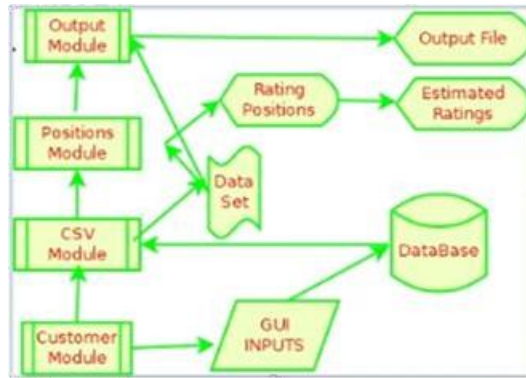


Figure 1. Proposed model of our system

Cosine similarity

Cosine similarity is a metric used to determine exactly how comparable the vectors are regardless of the size of theirs. Mathematically, it measures the cosine on the perspective somewhere between 2 vectors projected inside a multi dimensional room. The more compact the perspective in between the vectors, greater the cosine similarity. Using cosine similarity, we are that 2 vectors attain optimum similarity whenever the angle in between them is 0° (they are oriented inside an equivalent direction), they've zero similarity when the angle between them is 90° (they are orthogonal to just one another), they've -1 similarity when the angle between them is 180° (they are oriented around diametrically opposing directions). In this particular task, merchandise scores of every person is represented by a vector, so the cosine similarity is utilized a degree to determine the similarity involving all those vectors. The closest related vector scores will likely be utilized throughout the evaluation of unfamiliar reviews. General formula of cosine similarity is $\text{Cos}(x, y) = \frac{x \cdot y}{\|x\| * \|y\|}$

Where,

$x \cdot y$ is dot product of vectors

$\|x\|$ and $\|y\|$ are the length of two vectors

$\|x\| * \|y\|$ is the cross product of two vectors

The detailed algorithm of cosine similarity is shown in the below figure

```

Algorithm 5 Cosine Similarity

1. procedure SIMILARITY(query_vector, doc_vectors)
2.   i <- 0, similarity <- []
3.   while i ≠ total_docs do
         similarity[i] =  $\frac{\text{doc\_vectors}[i] \cdot \text{query\_vector}}{\|\text{doc\_vectors}[i]\| \|\text{query\_vector}\|}$  (6)
4.   i <- i + 1
5.   top_n <- (sort)similarity(: -n)
6.   return dictionary, all_words

```

Figure 2. Algorithm of cosine similarity

Implementation

Below Fig.3. represents the data flow diagram of this project. Initially, the customer details are stored and an unique id is assigned to that customer then he will enter the ratings of different products and then those details are stored in ".csv" file then we will estimate the null values in the csv file and those values are filled using cosine similarity algorithm.

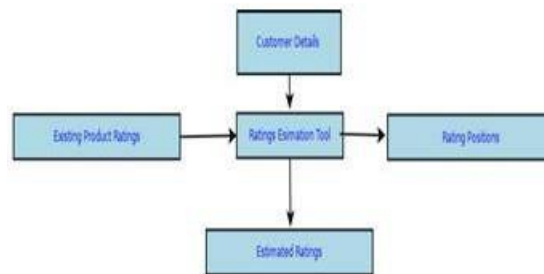


Figure 3. Dataflow Diagram

Here the GUI for this project is created using Qt designer. Qt designer is a simple GUI interface can be built easily. The GUI screen is visible like shown in the fig below



Figure 4. GUI Screen

But Here instead using GUI inputs a dataset has been taken into consideration and applied algorithm on it. This Experiment is made on product rating dataset in which it contains 3176 rows and the columns are customer id and ratings of 10 different products.

Firstly the dataset consists of ratings in format of string eg: excellent, good etc, Then data manipulation is done on that dataset to map the string values to integers like Excellent => 5 likewise.

After Mapping the highest rating will be the digit 5 and the null values are represented as digit 0. The below line represents how the mapping function is done.

```
df["RatingProd1"].map('NoRating':0,'Poor':1,'Average':2, 'Good':3,'veryGood':4,'Excellent':5)
```

Then the null values in the dataset are identified and then we will note the customer id of that null value row then will identify the similar user ratings for that product upto 6 iterations then we will replace that non-rating values with that similar user ratings. then we will generate the final csv file in which all the null values are replaced with similar product ratings.

The below figure represents about the position of null values and value to be replaced for that null value by identifying the similar user.

```
rating is estimated at Position: 82 5 1.0
rating is estimated at Position: 84 3 3.0
rating is estimated at Position: 85 5 1.0
rating is estimated at Position: 88 7 2.0
rating is estimated at Position: 90 4 2.0
rating is estimated at Position: 93 7 1.0
rating is estimated at Position: 98 5 2.0
```

Figure 5. Null Values Position in Dataset

Result Analysis

Analysis is the procedure of taking into consideration a little something intelligently or even making use of statistical techniques to be able to fully understand it or perhaps describe it. An evaluation is a description or maybe explanation which results as a result of taking into consideration a little something very carefully. It's a procedure for collecting as well as examining information, within such a fashion which the ensuing

info may be utilized to find out if the software program or maybe system is efficiently implementing the planned pursuits and it is effectively obtaining its reported goals as well as anticipated outcomes.

Here After Performing the operations in our experiment,the final dataset is created without null values presented in it.The below fig. shows the final output i.e, dataset of this project.

A	B	C	D	E	F	G	H	I	J
4	5	4	3	4	2	2	3	4	3
2	1	5	3	2	2	4	3	5	4
2	5	5	4	4	4	2	3	4	1
5	4	4	3	1	5	4	1	5	4
4	5	5	3	5	5	4	2	1	3
2	3	3	3	3	4	3	5	4	4
3	1	2	2	4	5	4	4	5	4

Figure 6. Output

Conclusion And Future Scope

With this analysis, we are going to be in a position to foresee the unfamiliar score of products. The rating evaluation is applied using the ensemble method known as cosine similarity algorithm, and they examine the item ratings of the clients as well as discovers the unfamiliar ratings as well as estimates the unfamiliar ratings of things to appropriate worth. The rating on the item is believed about the scale of five. The undertaking is possible on the on the internet looking corporate stores and websites to calculate the scores of the products of theirs,The undertaking helps you to estimate the general score on the item plus it gets better the product sales on the merchandise. This particular undertaking eventually gets better the need on the item and also quality of all of the service and also in the enhancement of quality of the expertise of companies.

As of today, the method is effectively analyzed with the information group of an us multinational list corporation which works a chain of hypermarkets, lower price stores, as well food markets. The applicability of the task, to calculate the scores on the products and services of various other shops, have to become investigated

References

1. J. Liu, L. Sun, R. Higgs, Y. Zhang and Y. Huang, "The Electronic Commerce in the Era of Internet of Things and Big Data," 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou, 2017, pp. 360-363, doi: 10.1109/CSE-EUC.2017.252.
2. L. Li, "E-Commerce Data Analysis Based on Big Data and Artificial Intelligence," 2019 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2019, pp. 133- 138, doi: 10.1109/ICCNEA.2019.00034.
3. S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2018, pp. 1275-1279, doi: 10.1109/ICIRCA.2018.8597286.
4. M. Tabibian and H. Amindavar, "E-commerce: technical and market approach," EUROCON'2001. International Conference on Trends in Communications. Technical Program, Proceedings (Cat. No.01EX439), Bratislava, Slovakia, 2001, pp. 149-152 vol.1, doi:10.1109/EURCON.2001.937785.
5. Gediminas Adomavicius and Alexander Tuzhilin. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions". In: IEEE transactions on knowledge and data engineering 17.6 (2005), pp. 734–749.
6. Marko Balabanovi´c and Yoav Shoham. "Fab: content-based, collaborative recommendation". In: Communications of the ACM 40.3 (1997), pp. 66–72.
7. Minqing Hu and Bing Liu. "Mining and summarizing customer reviews". In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004, pp. 168–177.