Data classification for improvement of product recommendation system using hybrid algorithm

[1]Saim Ahmed, [2]Mr. Anand Singh Rajavat,
[1]Student, [2]Assistent professor,
[1]CSE Department,
[1]SVVV, Indore, India

Abstract: Recommender systems have made significant utility in daily routing life. Online shopping and Social networking sites are playing crucial role in routine life. Over 3.5 billion people uses internet for various purpose. Online shopping retail sales are predicted to grow steadily in upcoming years. Product recommendation is one of the major requirements of e-commerce portals. This feature can help to increase shopping value with minimum shopping time. Logical recommendation not only helps to customer for purchasing but also increases total sales value. Generally, consumers need to search a lot to find a product of interest. Consequently, conventional recommender service systems often suffer from lack of scalability and efficiency problems when processing or analysis of this data on a large scale. To avoid these problems, a novel recommendations system using collaborative filtering algorithm and customer behavior is proposed to implement with Apache Hadoop server for BigData Processing. Proposed solution will attempt to recommend product based on similarity and popularity index of each product with respect to customer review. This project will attempt to develop own customer behavior analysis and classification algorithm to provide more perfect results. Amazon dataset will be used recommendation and evaluation purpose. Computation time for single and multimode cluster will be primary concern for performance measurement.

Index Terms **- Big Data, Amazon , Product Recommendation, Collaborative Filtering, Customized customer review analysis, Content Based Filtering.**

1. Introduction: On the Internet, where the number of choices is overwhelming, there is need to filter, prioritize and efficiently deliver relevant information in order to alleviate the problem of information overload, which has created a potential problem to many Internet users. Recommender systems solve this problem by searching through large volume of dynamically generated information to provide users with personalized content and services. This paper explores the different characteristics and potentials of different prediction techniques in recommendation systems in order to serve as a compass for research and practice in the field of recommendation systems. A recommender system or a recommendation system (sometimes replacing "system" with a synonym such as platform or engine) is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. Recommender systems have become increasingly popular in recent years, and are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. There are also recommender systems for experts, collaborators, jokes, restaurants, garments, financial services, life insurance, romantic partners and Twitter pages. Recommender systems typically produce a list of recommendations in one of two ways – through collaborative filtering or through content-based filtering (also known as the personality-based approach). Collaborative filtering approaches build a model from a user's past behavior (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users. This model is then used to predict items (or ratings for items) that the user may have an interest in. Content-based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties. These approaches are often combined.

1.1 Collaborative Filtering Algorithm: One approach to the design of recommender systems that has wide use is collaborative filtering. Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users. A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an "understanding" of the item itself. Many algorithms have been used in measuring user similarity or item similarity in recommender systems.

1.2 Content Based Algorithm: Another common approach when designing recommender systems is content-based filtering. Content-based filtering methods are based on a description of the item and a profile of the user's preferences. In a content-based recommender system, keywords are used to describe the items and a user profile is built to indicate the type of item this user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past (or is examining in the present). In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research.

2. Literature Survey:

A New Collaborative Filtering Recommendation Algorithm Based on Dimensionality Reduction and Clustering Techniques [1], in this paper, they proposed a new method for recommender systems that benefits from the potentialities provided by the k-means clustering algorithm and SVD technique. Firstly, the k-means clustering algorithm was adopted to cluster users in the same partition according to their preferences, and then the SVD was used in each cluster not only as a dimensionality reduction technique but also as a powerful mechanism, which could efficiently help in finding the most similar users. To evaluate the performance of the proposed method, they conducted experimentations on two real-world datasets for movies recommendation called Movie Lens 1M and Movie Lens 10M, which contain about 1 million and 10 million ratings made by anonymous users,

respectively. In addition, RMSE metric was adopted to evaluate the predictive accuracy of the proposed method in comparison with well-known k-nearest neighbor based recommendation and k-means-based recommendation methods.

Recommendations for All: Solving Thousands of Recommendation Problems Daily [2], in this paper, Sigmund, a system that allows us to solve instances of recommendation problems as a service. This system has been in production since early 2014 and serves tens of thousands of retailers daily by serving product recommendations. The system allows these retailers access to sophisticated product recommendations without having to design and deploy a recommender system. The objective of Sigmund is to provide the best possible recommendations at a low cost to thousands of retailers using a self-serve infrastructure. This raises several interesting challenges.

Riyaz P A, Surekha Mariam Varghese "A Scalable Product Recommendations using Collaborative Filtering in Hadoop for Bigdata" ICETEST- 2015, published by ScienceDirect-Elsevier-2015. Proposed solution developed a novel recommendations system using collaborative filtering algorithm. It is implemented in Apache Hadoop leveraging MapReduce paradigm for Bigdata. Proposed solution uses Data Extraction, Data Analysis, Algorithm Realization, Data Storage approach for recommendation purpose. This paper gives a scalable product recommendations collaborative filtering for Bigdata based on Amazon dataset.

3. Problem Statement: The study of existing solutions observes that most of the product recommendations implement ranking and popularity index value estimated by consumer view and ranking algorithms. Analysis of product category and classification of shopping is still a big challenge. Subsequently, e-portals provide web personalization based on product sale not the product nature. Although, limited work has been done in this field and a wide range of improvement is expected. The most important challenge into existing field is small scale data analysis. All the traditional recommendation work has been evaluated and created for small data size. Huge data size not only change the data nature but also increase lots of hurdle during mining work. Another ways, integration of customer behavior analysis and product nature is also absent into existing solutions.
The complete problem statement concludes into following points:
1. Enhancement in product recommendation algorithm for large volume dataset.
2. Integration of customer behavior with Product nature
3. Item cold start problem

A solution based on similarity, popularity and customer nature index for product recommendation is big challenge to overcome all the above

4. Proposed system: The complete proposed solution is divided into three modules which are listed below:
Module 1: Parsing Technique
Module 2: Filtering Technique
Module 3: Sentiment Analysis Technique
Step 1: Raw Dataset has been considered with variable memory size from electronics and video
Game field.
Step 2: Next, complete dataset has been moved from normal location to HDFS.
Step 3: Data Collection and Preparation module load desire dataset from HDFS to MySQL for cleaning and processing purpose. Afterwards, complete solution has been parsing through three different modules to generate results that are more accurate.
Module 1# Parsing Module
Step 4: Data Cleaning and Input Preparation method has been developed to remove unwanted anomaly and prepare input for mapper in form of Key and Value.
Module 2# Item Recommendation Module
Step 5: Item recommendation algorithm has been developed. Here, Collaborative Filtering algorithm has been used to estimate the similarity between user characteristics and product type. Threshold value has been maintained to retrieve upper level recommendation. Here we use Apache Mahout Collaborative Filtering algorithm which use Pearson correlation coefficient is used to find similarity of the user for item recommendation to the user.

A popular similarity measure in user - based CF: Pearson correlation
a,b : users
ra,p: rating of user a for item p
sim(a,b): set of items, rated both by a and b

$$sim(a,b) = \frac{\sum_{p \in P}(r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P}(r_{a,p} - \bar{r}_a)^2}\sqrt{\sum_{p \in P}(r_{b,p} - \bar{r}_b)^2}}$$

Module 3# Sentiment Technique
Step 6: At last sentiment analysis through Stanford API has been used to observe sentiments of user review. Here we find positive and negative item on the basis of user review. Then we recommend item to the user which has only positive review. By applying this accuracy of CF based recommendation is increase as we recommend only those products which are likely similar to the user and has only positive response from other user.
Module 4# Content based algorithm

Step 7: After sentiment analysis the result is take as input for content based algorithm. In Content based algorithm phase we find the similarity between items on the basis of metadata of the item.
Step 8: Hybrid Recommendation has been proposed by collaborative results of sentiment technique and Item recommendation module.
Step 9: Final Recommendation based on both algorithm is generated as Outcome.
Step 10: The complete solution is implemented using Hadoop 2.7.1 server and Java technology.
Step 11: A Swing based user interface has been developed to seek user input.

5. Conclusion: The complete work expects an integrated solution based on product characteristics and customer behavior for product recommendation. Few points to describe the expectations are illustrated below;

- ❖ To develop customer classification algorithm for customer behavior analysis.
- ❖ To calculate product similarity and popularity index to estimate sales trend and popular product list.
- ❖ Product recommendation based on product and customer nature.

REFRENCES:

[1] Hafed Zarzour, Ziad Al-Sharif, Mahmoud Al-Ayyoub, Yaser Jararweh, "A New Collaborative Filtering Recommendation Algorithm Based on Dimensionality Reduction and Clustering Techniques" IEEE 2018 9th International Conference on Information and Communication Systems (ICICS).

[2] Bhargav Kanagal, Sandeep Tata., "Recommendations for All: Solving Thousands of Recommendation Problems Daily", IEEE 2018 34th International Conference on Data Engineering.

[3] Riyaz P A, Surekha Mariam Varghese "A Scalable Product Recommendations using Collaborative Filtering in Hadoop for Bigdata" ICETEST- 2015, published by ScienceDirect-Elsevier-2015.

[4] Anindita A Khade "Performing Customer Behavior Analysis using Big Data Analytics" 7th International Conference on Communication, Computing and Virtualization, published by ScienceDirect-elsevier- 2016.

[5] Z. Huang, D. Zeng, and H. Chen, "A comparative study of recommendation algorithms for e-commerce applications," IEEE Intelligent Systems, 2006.

[6] Jyoti Gupta, Jayant Gadge, Performance Analysis of Recommendation System Based On Collaborative Filtering and Demographics, IEEE International Conference on Communication, Information & Computing Technology (ICCICT), 978-1-4799-5521-3, 2015 DOI: 10.1109/ICCICT.2015.7045675.

[7] Yanhong Guoetal., "An improved collaborative filtering algorithm based on trust in e-commerce recommendation system", IEEE 2010.

[8] B.M. Sarwar et al., "Item-item Collaborative Filtering Recommendation Algorithms," 10th Int'l World Wide Web Conference, ACM Press, 2001, pp. 285-295.

[9] Sanjeev Dhawan, Kulvinder Singh, Naveen Kumar "Collaborative Filtering Based Product Recommendation System for Online Social Networks", International Journal of Scientific Research in Computer Science, Engineering and Information Technology.

[10] Iman Barjasteh, Rana Forsati, Farzan Masrour, Abdol-Hossein Esfahanian, Hayder Radha "Cold-Start Item and User Recommendation with Decoupled Completion and Transduction".

[11] Samundeeswary K, Vallidevi Krishnamurthy "Comparative Study of Recommender Systems Built Using Various Methods of Collaborative Filtering Algorithm" International Conference on Computational Intelligence in Data Science (ICCIDS).

[12] Le Hoang Son "Dealing with the new user cold-start problem in recommender systems: A comparative review".

[13] Mohammad-Hossein, Nadimi-Shahraki, and Mozhde Bahadorpour "Cold-start Problem in Collaborative Recommender Systems: Efficient Methods Based on Ask-to-rate Technique "Journal of Computing and Information Technology 2.