

Optimizing Approach of Recommendation System using Web Usage Mining and Social Media for E-commerce

Anurag Singh
Research Scholar

Subhadra Shaw, PhD
Department of Computer Science & Application
AKS University, Satna, M.P

ABSTRACT

The recommendation system is very popular and plays an important role in the information system or web pages these days. The recommendation system can personalize its website with persons who suggest things to the user's needs. In the field of recommendation systems, the performance of all the recommendation algorithms is limited and each has its strengths and weaknesses, so much attention is paid to hybrid recommendation algorithms. In the proposed work, prediction using collaborative filtering, socio-demographic methodology, and sentiment analysis are integrated into a weighted system that is consistent with producing a single recommendation. This optimization approach improves the effectiveness of the recommendation process.

Keywords

Recommendation System, Collaborative Filtering, Basic Similarity Methods, Demographics, sentiment analysis.

1. INTRODUCTION

The success of e-business often depends on the ability to deliver personalized products to potential customers [1,2]. With the rapid development of the Internet and computer technology, a large number of online data sources offer many simple functions. Due to a large number of data sources, it is difficult for users to quickly find information that meets their needs and interests from these large data sets [4]. Finding the right item/information that will be useful for the user is a difficult and slow task, as data on the web has grown to a greater level over the past decade [1].

As recommendation has become a part of day-to-day life, we rely on external information before taking any decision about an artifact of interest. After getting user's preferences, if an accurate prediction algorithm is applied personalized recommendation can be done more correctly [2,3].

It's quite normal that e-commerce websites give their visitors a large number of very similar objects. Although users enjoy browsing multiple options, they sometimes find it difficult to select the right item in the plethora of presented choices. To help customers deal with such information, supplier systems are very important [5]. E-business sites allow us to automatically collect click-through data, which is a record of user page viewing history. This information is the value shown in predicting online shopping behavior [6]. A recommendation system is a tool, methodology, software, or system capable of making suggestions based on assessing the interests of customers. For example, websites such as Flipkart and Amazon use a referral system to refer products to buyers, while websites such as Wynk use a system to refer to music that is of interest to the listener [7].

1. Collaborative Filtering

Collaborative filtering (CF) was first proposed by Goldberg in

1992, [8] and its main idea is to recommend information that interests users by using the preferences of similar-minded groups. The collaborative filtering algorithm only considers the behavior of users to recommend items. They exploit the daily routine of other users' rating knowledge. Other users' behavior and preferences are used across the elements of what to recommend to new users [9]. Collaborative filtering is a technique where users can filter their favorite items based on similar customer responses. It works by looking for a larger group of people and finding smaller customers with similar interests to a specific customer. It looks at the items they like and combines them to create a list of ranked suggestions. CF is a relatively new algorithmic approach that has already entered many commercial applications and has become a key component of recommendation systems [10]. In general, the CF algorithm has two parts one is memory-based CF recommendation, and the second is model-based CF recommendation. [11].

1.1 Memory Based CF recommendation

Memory-based CF recommends the full use of historical data to find similar products. Memory-based CF recommendation can be categorized as item Based CF recommendation and User-Based CF recommendation. Item-based CF will find the equivalent of the target item based on the similarity between the recommended items. The user-oriented CF recommendation relies on active user neighborhood information to provide estimates and recommendations [10]. Neighborhood Selection is the key mechanism of a customer-centric CF policy, which selects a group of customers from the candidate environments as an active user.

1.1.1 Basic Similarity Methods

User-based CF is additionally called nearest-neighbor based Collaborative Filtering [13]. It first finds the target user's nearest-neighbors, then combines the preferences of neighbors to supply a prediction or top-N recommendation for the target users. Similarity computing which measures the similarity between two users is that the foremost vital area of user-based CF. Choosing an accurate similarity method can obviously improve the performance of user-based CF. The three basic similarity methods are as follows:

1.1.1.1 Cosine Similarity, In this case, two users are considered two dimensions in the object space. The similarities between them are measured by integrating the cosine of the angle between these two vectors. Formally, similarities between users i and j are given [13].

$$sim(i, j) = \frac{I \cdot J}{\|I\| \|J\|} = \frac{\sum_{c \in Item} RicRjc}{\sqrt{\sum_{c \in Item} Ric^2} \sqrt{\sum_{c \in Item} Rjc^2}}$$

where I, J represent the n -dimensional vectors that users i and j rated on the n items; $Item$ represents the whole items; denote the rating user i and j on item c .

1.1.1.2 Adjusted Cosine Similarity, basic cosine ratio has one important drawback that the difference in the rating scale between different users is ignored. The adjusted cosine similarity match removes this setback by subtracting the corresponding user rating from each of the two pairs. In the past, the similarity between user i and j is given by [13].

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)(R_{jc} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{jc} - \bar{R}_j)^2}}$$

where I_{ij} represents the items that user i and j co-rated; \bar{R}_i denote the average rating of user i and \bar{R}_j .

1.1.1.3 Pearson's Collection Similarity, In this case, the similarity between users i and j is measured by computing the Pearson correlation. to form the correlation computation accurate we isolate the co-rated cases. The correlation similarity is given by [13].

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)(R_{jc} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{jc} - \bar{R}_j)^2}}$$

The demerit of CF recommendation algorithms is additionally very obvious, which causes that these researchers make efforts to enhance the algorithm or to mix with data mining techniques so as to unravel it's those problems. CF recommendation algorithm mainly has the subsequent issues.

Sparsity Problem: Due to the tremendous sparsity of the data, collaborative filter recommendation algorithms often lose information when forming the nearest neighbor sets of target users, thus reducing the recommended effectiveness.

Cold Start Problem: When a new item first appears, the CF recommendation algorithm is not able to predict the score because the user is not evaluating the new item. [14]

1.2 Model-Based CF recommendation

The model-based CF recommendation often takes advantage of knowledge mining, machine learning, and other techniques. Model-based CF methods include the neural network model, the Bayesian belief network model, and the clustering model [11].

2. SENTIMENT ANALYSIS

It is jointly known as opinion mining that's designed to support user's feelings with varied categories like positive negative neutral and mixed. It's accustomed to getting people's scans and angles towards one thing with services and products. Sentence analysis and response mining is a study that analyzes attitudes and emotions from written communication. The increasing importance of sentiment analysis in relation to social media development, such as reviews of small twitter blogs and discussions on social networks. Within the world businesses and organizations perpetually wish to seek out client or public opinions regarding their product and services [15]

The opinion mining method consists of three steps: Pre-processing, my association rule. Summarization. The transaction O is comprised of extracted features and opinion from each product review.

$O = (\text{customer product}[\text{feature1}, \text{opinion1}], [\text{feature2}, \text{opinion2}], \dots, [\text{feature}_n, \text{opinion}_n])$

product: name of product , feature: at product review measure product reviews ,opinion: thinking of customer about product

or feature , [feature, opinion]: feature-opinion combination [16].

3. DEMOGRAPHIC ANALYSIS

Demographics is that the study of a population supported factors like age, race, and sex. It's the gathering and study of information relating to the final characteristics of specific populations. It's oftentimes used as a business selling tool to see the most effective thanks to reaching customers and assess their behavior. Segmenting a population by using demographics permits firms to see the dimensions of a possible market. The utilization of demographics helps to see whether or not its products and services area units being targeted there to the company's most vital consumers [18]. In our proposed work demographic knowledge is collected through Facebook, twitter.

4. REVIEW OF LITERATURE

There are completely different approaches and techniques were developed by the researchers for effective product recommendations.

P. Lops and B. Roy [17] planned the action primarily based relative recommendation system. This recommendation system created for all registered or unregistered visitors of the web site. It created the use of lexical patterns to get recommendations and compared the results of the proposed system with the user's techniques and merchandise based technique that showed that the planned system reduced the restrictions of the normal recommendation system and provides smart quality accuracy.

Singh Anurag, et. al [18]: Recommendation systems play an important role in marketing strategy because of increasing online sales. In this paper, different approaches like Collaborative Filtering, Demographic Approach, and Sentiment analysis are described and we focus on providing good quality product recommendations to all the users of an e-commerce site with help of different approaches. The recommendation system is a part of machine learning, which automatically learns from the experience rather than the predefined data. The proposed methodology of this paper is going to optimize the performance of the recommendation system using Collaborative Filtering, Demographic Approach, and Sentiment analysis. **Omar Zaarour, Mohamad Nagi [19],** proposed an improvement of the weblog mining procedure and to the prediction of the online navigational pattern. It proposed for session identification employing a refined time-out based heuristic. After detecting the navigational pattern by employing a specific density-based algorithm. Now finally, a replacement proposed method for efficient online prediction is additionally recommended for applicability. **Y. Koren, R. Bell [20],** Many efforts have been devoted to the CF problem, for details we refer the readers to the comprehensive survey of Koren et al. [17]. According to the working scheme, all these works can be categorized into offline models and online models. Roughly, the offline models can be classified into three types: The neighborhood models, the factor models, and the deep models. **Yadav Sambhav et. al. [21],** In this the paper proposes an approach to give weights to the Items in the user-item rating matrix to find the better neighborhood of the active user using the BAT algorithms. This method helped in providing personalized recommendations to all users as it generated a different set of weights for each user. **Joshi V Ameet [22],** author specified in a particular chapter of the e-book that recommendation systems represent a relatively new entry into the field of machine learning and it is a cutting edge of the area and is

under strong development. **Bo Song, et al.[23]**: the proposed algorithm of the user base, collaborative filtering recommendation is based on the Mahout framework, combined with the Naive Bayes classifier, to establish a user model and form a personalized recommendation specified in this paper. From the analysis of the three algorithms in this paper, it is concluded that the collaborative filtering recommendation algorithm interaction based on the user model is an algorithm of machine learning. Recommendation algorithms that work on the Hadoop platform not only solve practical problems but also include a practical and personalized recommendation for users. **Lee H. et al.[24]** In this paper, a real neural network specific to the analysis of time series data is used in the recommended model. The recommended LSTM(long-term short-term memory) recommendation model is evaluated by comparing the following: two models, an item-based filtering model, which is widely used as a rating system in the recommendation system, and the popular model, which is simple but applicable in the business environment. This paper extended the complimentary intervals as measures of different time intervals and examined the effectiveness of multiple intervals over time. Unlike previous studies of recommendation systems, which focus on the accuracy of recommendation from a single point of view, this paper compares the accuracy of the recommendation model with multiple times and shows that the proposed model has better performance even in a more time-consuming view. Multipoint system is a kind of decision support system that helps customers save their time and money and make decisions quickly. In this study, the use of Internet-based online sales data, but using the LSTM-based recommendation model on a variety of transaction data sets is a promising area for future research. **B. Chilehaoui, et.al[25]**, Collaborative Filtering, content-based filtering, and demographic filtering have been combined to solve the cold start problem.

5. PROPOSED WORK

This work is based on information theory. Our work brings a concept that helps analyze customer buying patterns and identify future customer needs. Today everyone is faced with the problem of massive selection bias called the most prevalent problem. Over the past decades, numerous studies have been conducted to develop methods for recommendation systems and to improve their accuracy. The goal is to optimize the performance of the recommendation system and overcome their problem of sparsity and cold –start and their scalability problems. Input to the system includes user, product, rating information, user demographic information, sentiment data, and visits counting on a new product. In the proposed work as shown in the following fig-3. Collaborative filtering is integrated with the demographic approach, sentiment analysis with a hybrid approach.

The proposed work can be divided into some parts.

Step-I: The unstructured data will be converted into the structured form after applying pre-processing and a user-item matrix will be generated.

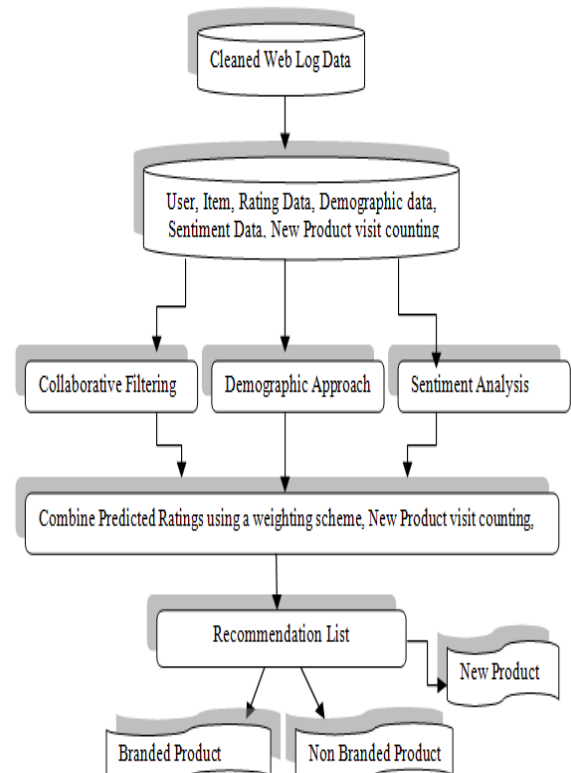


Fig.-1 Optimizing approach of Recommendation system

Step-II: In this step we generate user-product rating Matrix, consider a set of n users $U=\{u_1,u_2,u_3,\dots,u_n\}$ and a set of m items $I=\{p_1,p_2,p_3,\dots,p_m\}$. Let R_{ij} , $1 \leq i \leq n$, $1 \leq j \leq m$, denotes the rating value of user “I” on product j. If $R_{ij} \neq 0$ indicates that the user “I”

have not rated the product j [26]. The rating matrix R is given in the following (1).

$$R = \begin{pmatrix} & I1 & I2 & \dots & Im \\ U1 & R11 & R12 & \dots & R1m \\ U2 & R21 & R22 & \dots & R2m \\ \dots & \dots & \dots & \dots & \dots \\ Un & Rn1 & Rn2 & \dots & Rnm \end{pmatrix} \quad (1)$$

Step-III: The main concept of collaborative filtering is to calculate similarity products or users. At present, scholars have proposed many similarity measures, cosine similarity, Adjusted Cosine Similarity, Pearson coefficient, and other methods. Among them, the Adjusted Cosine Similarity is the best among many calculation methods. In this approach, we analyze the Adjusted Cosine Similarity method.

Step-IV: In this approach, the demographic profile of customers is considered in computing item ratings. This consideration is based on the fact that demographics contribute to differences in people’s tastes or preferences. A clustering algorithm such as K-means partitions users based on user demographic Data using Euclidean distance as the distance measure. The prediction for an item using demographics based user clusters will be taken as the average

rating of the user cluster for that item.

Step-V: Sentiment classification techniques can be divided into machine learning, dictionary-based approach, and hybrid approach [27]. The Machine Learning Approach applies popular machine learning algorithms and uses language features. The Lexicon-Based Approach relies on a sentiment dictionary, a collection of known and pre-edited sentiment words. It is a dictionary-based approach and a corpus-based approach that uses statistical or semantic methods to find sentimental polarity. The hybrid Approach combines both approaches and is very common with sentiment lexicons playing a key role in the majority of methods. The reviews have been categorized based on the polarity into positive and negative classes by carrying out a sentiment analysis on the preprocessed dataset. To achieve this a pertained Baise Bayes classifier is run on the product reviews dataset by using a dictionary-based approach. The model is trained on features like several positive and negative words which help in determining the polarity of an input review/ sentence is shown in the following.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Here, A is the class(positive/negative) that the review B will be classified into on the basis of the features that B contains and have been learned by the approach[28].

Step-VI: This proposed work will also work for a fresh new product which is produced by small scale enterprises. There is not much information about that new product and company. How to the recommendation of that new product. First of all new product is upload with features and company name on a particular e-commerce website. Whenever registered or unregistered users visit on the website, then count the click on that product of every user. On the basis of click and comparison, the recommendation system shows that product with click count to every visited user. The recommendation system recommends a particular new product on the basis of the maximum visit.

Step-VII: Weight computation is one in which the score of a recommended product is computed from the results of all the available recommendation techniques present in these approaches. This integrates the scores from each technique using a linear formula. Therefore, the various techniques must be able to produce their recommendation score, which can be linearly combinable. It is very useful that all of the system's capabilities are brought to bear on the recommendation process. The benefit of a weighted hybrid is that all the recommender system's strengths are utilized during the recommendation process in a straightforward way. Hence the ratings of several recommendation techniques are combined together to produce a single recommendation.

6. OUTCOME

The success of a recommendation system is depending on user satisfaction. Service recommendations according to the user needs will define the success rate of this system. This proposed system definitely provides an accurate response to the user and it will be the best outcome of this system.

7. CONCLUSION

This paper focuses on providing good quality product recommendations to all the users of e-commerce. Recommendation systems play an important role in marketing strategy because of increasing online sales. At present,

customers' expectations are high as cost and quality are concerned and at the same time manufacturers may compromise on profits as to competitors, due to the dynamic business scenario. In this paper, an optimizing approach recommender system is proposed to give recommendations to both the registered and unregistered users. There are different approaches like collaborative filtering, demographic approach, and sentiment analysis are described. The prediction accuracy of every single recommendation algorithm is insufficient and each has its own merit and demerit, so we pay more attention to hybrid recommendation algorithms. In the weighted strategy, we need to decide how to the optimization of the weight of each recommendation algorithm. Hence the ratings of several recommendation techniques are combined together to produce a single recommendation.

8. FUTURE WORK

We can work on models to improve their accuracy and involves in discovering an advanced combination of data mining techniques that provide good results than existing combinations. In the future, the same approach can be used for prediction and recommendation, with a large database. The proposed system can be developed in an experimental environment in the future.

9. REFERENCES

- [1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, pp. 734-749, 2005.
- [2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, 1994, pp. 175-186.
- [3] McNee, S., Riedl, J., Konstan, J.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: 24th International Conference Human Factors in Computing Systems, Montréal, Canada, pp. 1097-1101 (2006)
- [4] Huang X, Pan S, "Study on Grid Scheduling of Super-Peer Model Based on QoS", *Fourth International Symposium on Information Science & Engineering*, IEEE Computer Society, 2012.
- [5] Jiro Iwanagaa, Naoki Nishimurab, Noriyosh Sukegawac, Yuichi Takano, "Improving collaborative filtering recommendations by estimating user preferences from clickstream data", 1567-4223/ © 2019 Elsevier B.V. All rights reserved, doi.org/10.1016/j.jelerap. 2019 100877.
- [6] Zuoquan Lin, Hanxuan Chen, "A Probabilistic Model for Collaborative Filtering", *WIMS2019*, June 26- 28, 2019, Seoul, Republic of Korea.
- [7] P. Lopes and B. Roy, "Dynamic Recommendation system using web usage mining for e-commerce users" *Procedia Computer Science* 45, pp 60-69, 2015.
- [8] Singh Anurag, Wao A Akhilesh, "Optimizing Recommendation system using web usage mining and social media for e-commerce", ©2020, IJSREM, www.ijsrem.com.
- [9] Gupta Shefali and menu dave, "An overview of recommendation system method and technique",

- ©Springer Nature Singapore pvt.Ltd.,2020.H.Sharma et al. (eds.), “Advances in computing and Intelligent System”, doi.org/10.1007/978-981-15-0222-4
- [10] Goldberg D, Nichols D, Oki BM et al , “Using collaborative filtering to weave an information tapestry. *Commun ACM* ,1992,35(12):61–70,
- [11] Tareq U Shihab,et al., “ Framework of dynamic recommendation system for e-shopping”,Received:20February2018/ Accepted: 7 November 2019, © Bharati Vidyapeeth’s Institute of Computer Applications and Management 2019, charm: Springer , <https://doi.org/10.1007/s41870-019-00388-6> .
- [12] Yehuda Koren, “Tutorial on Recent Progress in Collaborative Filtering”, *RecSys’08*, October 23–25,2008,Lausanne, Switzerland, ACM 978-1-60558-093-7/08/10.
- [13] Bo Hong, Mengchen Yu , A collaborative filtering algorithm based on correlation coefficient, Received: 6 August 2018 / Accepted: 26 October 2018 , Springer-Verlag London Ltd., part of Springer Nature 2018
- [14] Xiao-Yu Huang a , Bing Liang b , Wubin Li ,” Online collaborative filtering with local and global consistency“, doi.org/10.1016/j.ins. 2019. 08.009 0020-0255/© 2019 Elsevier.
- [15] Lin Guo, Qinke Peng ,” A Combinative Similarity Computing Measure for Collaborative Filtering”, *Applied Mechanics and Materials Vols 347-350 (2013)* pp 2919-2925 Online: 2013-08-08, © (2013)Trans Tech Publications, Switzerland.
- [16] Mai Jianying, Fan Yongjian shen, ” A neural networks based clustering collaborative filtering algorithm in e-commerce recommendations system”, *ICWISM*, © 2009 IEEE.
- [17] Bing Liu , “ Sentiment Analysis and Opinion Mining”, Morgan & Claypool Publishers, May 2012.
- [18] Won Young Kim, Joon Suk Ryu, Kyu II Kim, Ung Mo Kim “A Method for Opinion Mining of Product Reviews using Association Rules”, *ICIS 2009*, November 24-26, 2009 Seoul, Korea Copyright © 2009 ACM 978-1-60558-710-3/ 09/11.
- [19] Omar Zaarour, Mohamad Nagi, “Effective Web Log Mining and Online Navigational Pattern Prediction”, ELSEVIER, 2013.
- [20] Y. Koren , R. Bell , *Advances in collaborative filtering*, in: *Recommender Systems Handbook*, Springer,2015, pp. 77–118 .
- [21] Yadav Sambhav, Shreyam Vikesh, Nagpal Sushama, “ An Improved Collaborative Filtering Based Recommender System using Bat Algorithm”, 1877-0509 © 2018 The Authors. Published by Elsevier Ltd.
- [22] Joshi V Ameet, “Machine Learning and Artificial Intelligence”, ISBN 978-3-030-26622-6 (eBook) <https://doi.org/10.1007/978-3-030-26622-6>, ©Springer Nature Switzerland AG 2020
- [23] Bo Song, Yue, et al, “Research of Collaborative Filtering Recommendation Algorithm based on Mahoutand user Model”, *Conference Series 1437 (2020) 012095*, IOP Publishing, 2020,doi:10.1088/1742-6596/1437/1/012095.
- [24] Lee In Hea , Choi Young, et al.,” a multi-period product recommendation system in online food market based on recurrent neural networks”, ©2020 by the authors, Licensee MDPI, Basel, Switzerland,Sustainability 2020, 12, 969; doi:10.3390/su12030969.
- [25] B. chilehaoui, M. chiazzaro and . Wang, ” An improved Hybrid Recommender System by ombing predictions” , *IEEE workshops of International conference on advanced information networking and application* PP. 644-649,2011.
- [26] Panda ku. Sanjaya, Bhoi ku Sourav, Singh Munesh, “A collaborative filtering recommendation algorithm based on normalization approach”, © Springer-Verlag GmbH Germany, part of Springer Nature 2020.
- [27] Walaa Medhat, Ahmed Hassan, Hoda Korashy ,” Sentiment analysis algorithms and applications: A survey”, 2090-4479 _ 2014 Production and hosting by Elsevier B.V. on behalf of Ain Shams University <http://dx.doi.org/10.1016/j.asej.2014.04.011>
- [28] Roul Ku Rajendra, Sahoo Keshari Jajati , “ Sentiment Analysis and Extractive Summarization Based Recommendation System”, © Springer Nature Singapore Pte Ltd. 2020,H.S. Behera et al. (eds.), *Computational Intelligencein Data Mining, Advances in Intelligent Systems and Computing* 990, https://doi.org/10.1007/978-981-13-8676-3_41.