International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 11, Issue 05 (May 2015), PP.52-56

System For Product Recommendation In E-Commerce Applications

Usharani.S[1] and Anirban Basu[2]
Department Of CSE, East Point College of Engineering & Technology Bangalore, Karnataka, India

Abstract:- Recommendation technology, is an important method for information filtering in E-Commerce applications, and can effectively reduce information overload in Internet. It narrows down the choice of products from a large number of product offerings. With increase in the number of E-commerce users and products, the original recommendation algorithms and systems face many challenges namely in modeling user's interests more accurately, providing more diverse recommendation modes, and supporting large-scale expansion of data. To address these challenges, and meet the present demands of E-commerce applications, a personalized hybrid recommendation system, which can support massive data set, has been designed and implemented on Hadoop.

Keywords:- E-Commerce,Bigdata,Apache Hadoop,Personalized Hybrid Recommendation.

I. INTRODUCTION

With the increase in use of internet, E-commerce is gaining wide popularity. In recent years, several E-commerce websites have become very popular. Amazon, eBay, Netflix, etc. are examples. Learning about the interests of the consumers facilitates consumer shopping, and has become key issue in customer relationship management. There is a need to incorporate this in E-commerce applications. It is challenging to find out the product that an user really needs from a large number of product offerings. Although, E-commerce is being used widely, the concept of personalized recommendation is becoming crucial these days. There is a need to extract the characteristics of the products, and potential preferences, of consumers from his/her online browsing patterns and from purchase records, and to recommend appropriate products to the consumer which he/she is most likely to procure based on this.

In industry, personalized recommendation has become the core technology in E-commerce applications. Typical systems include: the book recommendation system of Amazon [2], the movie recommendation system of Netflix [3] and video recommendation system of YouTube [4] etc..

However, with the explosive growth of the number of E-commerce users and products, the amount of data of recommendation systems have undergone major changes. Users with diverse interests and more personalized demands, are using E-Commerce platforms and the amount of data to be processed is growing rapidly. The voluminous data to be processed is mostly in unstructured format which is not easy to analyse. All three characteristics namely: Volume, Velocity and Variety of Big Data are present in the data to be processesd. In such situations, earlier recommendation algorithms and systems face several **c**hallenges:

- Accurately modelling user's interests
- Providing more diverse recommendation modes
- Supporting large-scale expansion of data

To address these challenges, a method has been proposed in this paper. The existing methods do not support the analysis of Big Data. In the proposed method the performance of the system can be speeded up by using map reduce concept.

II. RELATED WORK

In recent years, research on recommendation technologies has attracted attention due to the "information overload". There are many companies who have designed their own recommendation system to support their Web applications, such as the Google news recommendation [6], FOFs system of Facebook [7] and the music recommendation of Yahoo! [8], etc. In these systems, generally the collaborative filtering (CF) is the commonly used core recommendation technology. CF is based on Analyzing historical data,

Research is being carried out to improve the different aspects of CF. For example, the papers [9] and [10] are focused on the sparsity issue of CF. In [9], Wang et al. proposed a unifying user-based and item-based approach by similarity fusion, and in [10] Sarwar et al. proposed a Latent Semantic Indexing (LSI) to reduce the dimension space and increase the data density, making the user similarity much more obviously. On the other hand, Mehta et al.[11] have discussed the attack resistance and trust issue of CF algorithms. Many other new CF recommenders such as Bayesian network-based [12] theoretic approach to collaborative filtering [13] and item-based [14] technologies and algorithms have been proposed to improve the accuracy and performance. But they did not consider the dynamic changes in consumer's interests.

Content based filtering is based on tha Profile attributes that is Similar Item Related to past. A key issue with content-based filtering is whether the system is able to learn user preferences from user's actions regarding one content source and use them across other content types.

When the system is limited to recommending content of the same type as the user is already using, the value from the recommendation system is significantly less than recommending other content from other services. For example, recommending news articles based on browsing of news is useful, but it's much more useful when music, videos, products, discussions etc. from different services can be recommended based on news browsing.

Again all the above mentioned methods and algorithms are centralized so that they cannot satisfy the scalable requirement of massive data processing in E-commerce applications.Because of this, a hybrid approach is used which is a combination of collaborative and content-based filtering.

To face today's new challenges as we identified earlier, the existing mechanisms also have some other limitations. These are : the current model of consumer interest cannot effectively reflect the change of consumers' interest for products, lack
of a hybrid framework for different demands and logics,and current distributed algorithms cannot support Big Data processing[5].

III. RECOMMENDATION SYSTEM

In this section, we describe a recommendation system. The architecture of the system is shown in Figure 1. In order to respond to user's product requirements within a short time, the system has been implemented in Hadoop.

The following steps are involved as shown in Figure 1:
Step1: The system summarizes the recommendation information gathered periodically based on the previous transactions of the user. This data is collected in an offline background mode.
Step2: The system uses a preprocessor to extract the useful information from this data and stores them in Hadoop Distributed File Sysem [HDFS].
Step3: The preprocessed data is clusterd by using Clustering algorithm and an User Preference Tree is constructed, which is used as the input to recommendation algorithm in the next step. Clustering is used for grouping similar types of users based on the similarity of their preferences. Users choosing same type of products are grouped together. This helps in building User Similarity Matrix.

For each user, UPT is constructed based product similarity based on their product preferances And create product similarity matrix.

UPT is described in details in section 3.1.
Step 4: The recommendation algorithm is accelerated using MapReduce technique on two matrices: one based on similarity of users and another based on similarity of products.

Step5: Finally recommendation is provided to the active consumers immediately in online mode.

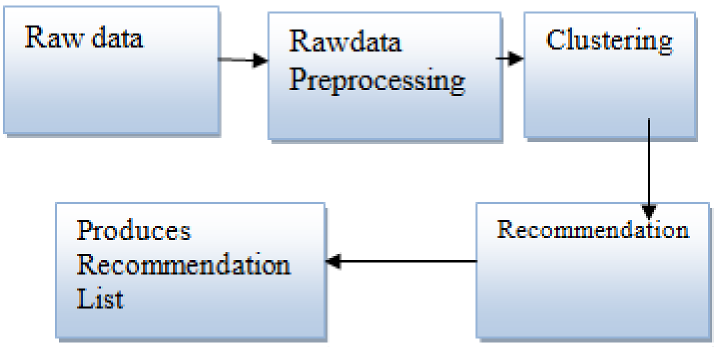

Figure 1. The architecture of Recommendation System.

3.1 UPT

In recommendation system, the field Classification Vector CV and the Interest Energy(IE) are defined and a tree structure is proposed for the modeling of user's interest, called User Preference Tree (UPT)[9].

Step 1 : The field Classification Vector CV of a product is defined as the summation of product name and product weight.

CV=<(CVK1,CW1),(CVK2,CW2),......, (CVKm,CWm)>,

where CVKx denotes the x-th dimensional attribute's name of a product and CWx denotes its relative weight i.e., classification weight in the range of 0 to 1.

The attributes of a Product Pj can be defined as CV(Pj) = <CVK1,CVK2,…,CVKM>,

where CVKX denotes the x-th dimensional attribute's value of Pj.

For example in Figure 2:

CV = <(First Category), (Second Category), (Brand), (Style)>,

CV(Pj)= <Clothes, Jacket(Man), Adidas, Black >.

Step 2 : Interest Energy(IE):

Interest Energy is defined as the degree of interest of an User Ui to Product Pj. This is determined by the frequency of visit of user Ui for a Product Pj.

Step 3 : User Preference Tree (UPT):

UPTfor an user is defined as a tree of depth |CV|+1. Where CV is Classification Vector.The leaf node which represents a product by user Ui is defined by five-tuple in the leaf node by {PID, IE, IW, CR, level}
where ,

- PID denotes Product ID.
- IE denotes Interest Energy.
- IW denotes the interest weight of certain product.
- CR denotes the rating of User Interface Ui to certain product.
- Level denotes the final choice of the Product .

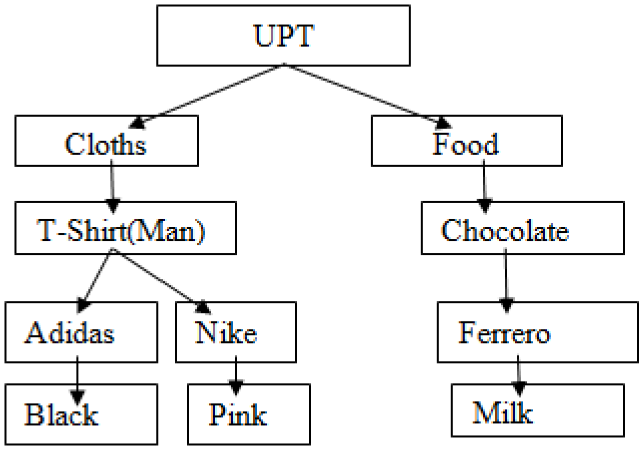

Figure 2. User preference tree based on a four level Classification.

UPT is shown in Figure 2. UPT Tree is defined based on the product selected and here two types of products are choosen i.e.,Cloths and Food. In Cloths, Men T-Shirts selected and in that Adidas (black color) & Nike (pink color), are chosen. In Food,Chocolate is chosen and in that Ferrero is selected. We define products based on selection of user, where a visited product Pj uniquely corresponds to a path from root to corresponding leaf node,where each keyword corresponds to the relevant attribute of product Pj.

User Similarity is defined as the cluster of users interested in similar recently products.

Product similarity is defined as the cluster of similar products based on their features.

These two matrices help in recommending products to the consumers..

IV. PERFORMNCE ON MAP-REDUCE FRAMEWORK.

With the great explosion of the number of products and the increase in the popularity of online purchases, the efficiency and scalability of recommendations are proving to be important. If we still use the traditional centralized processing methods, the consumer's requirement can not be satisfied for datasets with size in Tera Bytes, as the response time of recommendation may be up to several hours. Therefore, in order to greatly reduce the recommendation response time, we adopt MapReduce to reduce the time for clustering and for recommendation mechanism. Parallel processing methods of user similarity and products similarity calculation are proposed.

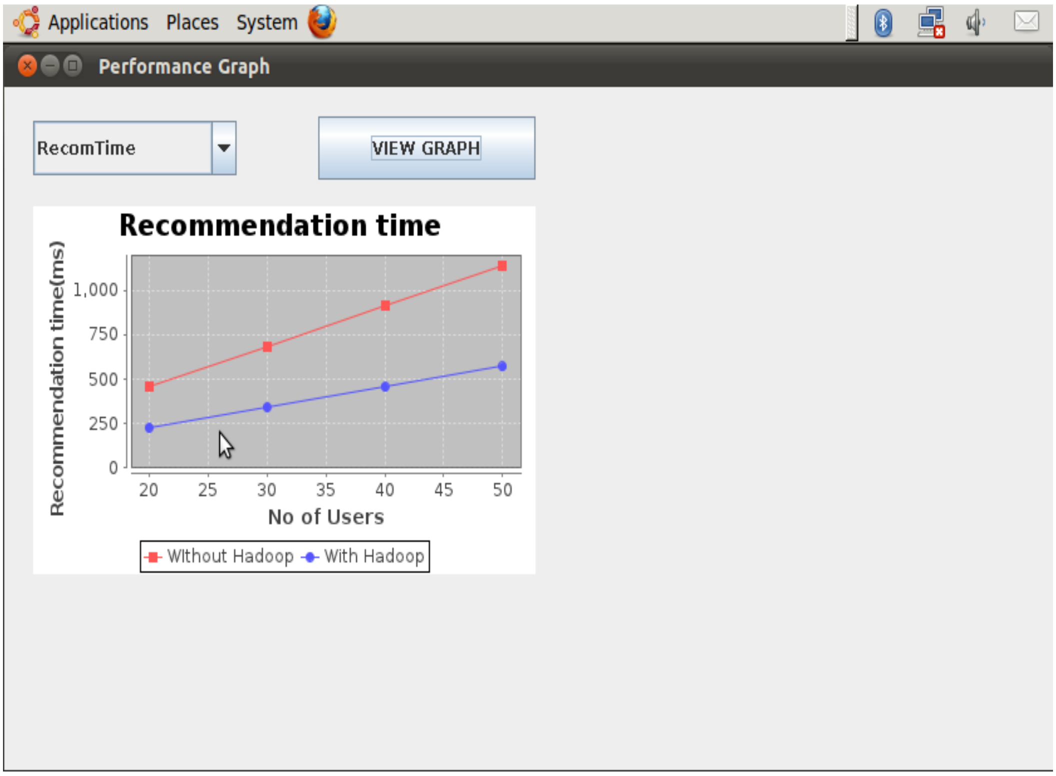

Fig 3:Showing the performance graph.

V. CONSLUSION

In this paper, a personalized recommendation system has been discussed which can support massive data set. Here recommendation algorithm is designed to satisfy user's diverse demands and supports the Big Data Set. The execution process of recommendation algorithms can be speeded up by using MapReduce. The system has been implemented on Hadoop. Performance has been analysed and results show the advantages.

REFERENCES

[1]. Z. Huang, D. Zeng and H. Chen. A Comparison of Collaborative- Filtering Recommendation Algorithms for E-commerce.
[2]. URL:www.amazon.com.
[3]. URL:www.netflix.com.
[4]. URL:www.youtube.com.
[5]. M. Armbrust, A. Fox, R. Griffith, et al. Above the Clouds: A Berkeley View of Cloud Computing[J]. Commun. ACM. 2010, 53: 50-58.

[6]. J. Liu, P. Dolan, E. Pedersen. Personalized news recommendation based on click behavior. Proceedings of the 15th international conference on intelligent user interfaces. ACM, 2010: 31-40.
[7]. L. Backstrom. Dealing with structured and unstructured Data at Facebook. The 8th Extended Semantic Web Conference ESCW 2011.
[8]. Aizenberg N, Koren Y, Somekh O. Build your own music recommender by modeling internet radio streams.
[9]. J. Wang, AP. Vries, MJT. Reinders. Unifying User-based and Itembased Collaborative Filtering Approaches by Similarity Fusion.
[10]. Sarwar, B.M., Karypis, G., Konstan, J.A., et al. Application of dimensionality reduction in recommender system
[11]. B. Mehta, W. Nejdl. Attack resistant collaborative filtering.
[12]. Breese, J.S., Heckerman, D., Kadie, C. Empirical analysis of predictive algorithms for collaborative filtering.
[13]. Aggarwal, C.C., Wolf, J.L., Wu, K., et al. Horting hatches an egg: a new raph-theoretic approach to collaborative filtering. [14] Sarwar, B., Karypis, G., Konstan, J., et al. Item-Based collaborative filtering recomm