Recommendation System for Ecommerce using Big Data

Prof. Alka Leekha

alka.leekha@gmail.com

Shreyas Upadhye, Pratik Ahire, Pranav Pawar

Bharati Vidyapeeth College of Engineering

Sector-7, C.B.D. Belpada, Navi Mumbai-400614, India.

shreyas2june@gmail.com, ahirepratik4.94@gmail.com, pranavpawar7@gmail.com

Abstract

E-commerce and web portals are facing fierce completion as they strive to win customer and enterprise business. E-commerce is growing overall yet the competitive pressure is significant since it is a scalable business and the winner takes it all. Customers are not loyal and mostly arrive from search engines and online advertisements. They wait for few second on portal looking for intended product and less pricing and many of them are motivated by friends circle on social media. With advent of web 2.0 most of the people start using internet. Many people share their feeling, needs and emotions on social media. To keep customer visiting e-commerce portal and attract customer to portal must understand customer needs, show them appropriate personalize recommendation content in each content module on portal. While browsing on site customer provides us many clues by means of behavior on site, various actions of site, feedback and customer care data. So we can mine this data to understand interest of customer. Traditionally many content provider use content base filtering which take single use historical data for recommendation and collaborative filtering which consider the preferences from other users with similar tastes. With these traditional approaches we can get only limited knowledge of customer needs hence we have proposed to use user's historical data and social media data such as status, personal info, and interpersonal influence to get better insight knowledge of customer need. All the methods have their own strengths and weakness. So we adopt a hybrid recommender system by combining two or more recommendation techniques to get better content optimization. The most popular hybrid method combines Content based predictor and Collaborative filtering recommender systems. By combining content and collaboration an elegant and effective framework can be developed. Hence the name Content Based Collaborative Filter (CBCF). While considering the large customer space and variety and velocity of data traditional data processing technologies cannot cope up with the problem. Hence we have propose hadoop distributed processing to rescue from big data problem.

I.   Introduction

In order to improve customer experience on E-commerce and portal sites, measure and model customer experience is essential. Recent years have witnessed rapid growth of the Internet, which has become an important medium to deliver digital content to Web users instantaneously. E-commerce and portal site now have vast amount of data about their visitors profile, visitor behavior and action on web portal, customer care, feedbacks and search queries. Social media site such as Twitter, Facebook etc. provide

great venues for customer to share joy and struggle, vent emotion and stress and seek social support. On social media site people share their everyday encounter in an informal and casual manner. People digital footprint provides vast amount of implicit knowledge and whole new prospective for E-commerce business. Customer shares their opinion about products and their experience of e-commerce portal. So e-commerce business can understand data about customer from various touch point to provide personalize experience at portal by analyzing customer need, can use that data to analyze customer service experience and to track dissatisfied customer, for predicting sale, for dynamic pricing offers. But considering sheer data volumes, diversity of internet slang and unpredictability of location and timing of student posting on internet and complexity of customer experience, the pure manual analysis of customer data cannot give us deep insight knowledge of customer.

One of the promising solutions to rescue from high velocity, high variety and volume customer data is Big data analytics. Measuring customer experience holistically is fundamentally big data problem. Gartner define, Big data as having tree attributes: high volume, high velocity and high variety. High volume means growing quantity of data, high velocity means acceleration in speed of data and high variety indicate increase in data of data, which for the e-commerce provider new challenge for managing data related to search queries, web logs, user action and behavior, purchase history and social media data.

In this project, we focus attention on to improve customer experience on e-commerce portal by providing personalize online content optimization and tracking dissatisfied customer by analyzing profile information, user behavior and action on ecommerce portal, feedback and social media data.

Traditional content recommendation divides into two measure approaches: content base filtering and collaboratively filtering. In content base filtering customer profile is generated for user base on product description of product items previously rated by user, the main drawback of this approach is limited capability of recommending product contents. Collaborative filtering is one of the most successful and widely uses technique, analyzes user rating to recognize the commonality and recommend the items by leveraging the preferences from other user. These techniques can be view as first generation of traditional recommendation technique to predict user interest. All the methods we described have their own strengths and weakness. So we adopt a hybrid recommender system by combining two or more recommendation techniques to get better content optimization. The most popular hybrid method combines Content based predictor and Collaborative filtering recommender systems. By combining content

System Architecture Diagram.
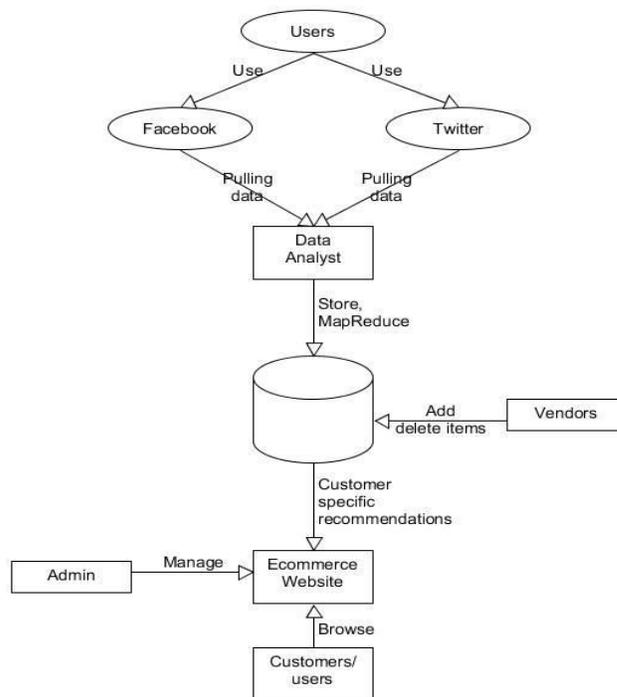
and collaboration an elegant and effective framework can be developed. Hence the name Content Based Collaborative Filter (CBCF).

II. Proposed System

We propose to use user profile, users' historic actions and behavior, feedback, customer care data, users' social media data to serve each customer with personalize and track users dissatisfaction. We use internal data about user to segment the whole traffic into different customer segment. Then we will refine user interest segments base on personal social media data and interpersonal influence on social media. And we use user sentiments on social media to predict user interest and combine it with customer care data to track dissatisfied customer. So we adopt a hybrid recommender system by combining two or more recommendation techniques to get better content optimization. The most popular hybrid method combines Content based predictor and Collaborative filtering recommender systems. By combining content and collaboration an elegant and effective framework can be developed.
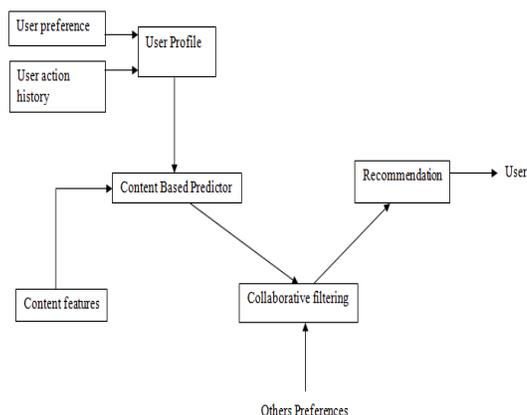
Content based collaborative filtering.

Steps:

1) Data Generation:

Data is generated whenever user visit ecommerce portal in form of log of pages visited, links click on particular page, time spent on particular page, product view, product rating, Feedback for particular project, customer queries at customer care department.

Customers often share their personal info, product experience, need and interest on Facebook and twitter.

2) Data Collection:

As data generated is reside at various places like customer care department, action and behavior log at web server and data at social network. We collect data using big data technology with HDFS.

To get social media data we will use facebook.api and twitter.api

3) Data Cleaning

After collection of data we will reduce some noisy data and store it in appropriate format for further processing like removing

4) Data mining

We will use text clustering e.g. Multi label Naïve Bayesian algorithm and regression method for predicting the user interest.

5) Knowledge discovery and presentation

Then we will use customer segment and user interest predicted in previous step to formulated appropriate recommendation for user at present time.

III. Conclusion

We propose to use user profile, users' historic actions and behavior, feedback, customer care data, users' social media data to serve each customer with personalize and

track users' satisfaction. The proposed system acts as a platform for many commercial organizations to understand their customer better. It helps the Ecommerce companies to capture the responses of their products/services from social networking sites and measure the number of responses in favor or against them. We use internal data about user to segment the whole traffic into different customer segment. Then we will refine user interest segments base on personal social media data and interpersonal influence on social media. And we use user sentiments on social media to predict user interest and combine it with customer care data to track dissatisfied customer.

References

[1] Jiang Bian, Anlei Dong, Xiaofeng He, Srihari Reddy and Yi Chang, "User action   interpretation for online content optimization," IEEE Transaction, Knowledge and Data Engineering, Vol.25, No.9, 2013.

[2] Wei Tan, M. Brain Black, Iman Saleh and SchahramDustdar," Social Network Source Big Data Analytics," publish by IEEE Computer Society, 2013.

[3] Jeffrey Spiess, Yves T'Joens, RalucaDragnea, Peter Spencer and Laurent Philippart, "Using Big Data to improve customer experience and business performance," Alcatel-Lucent, Bell Labs Technical journal 18(4), 3-17, 2014.

[4] Carlos E. Otero and AdrinPeter,"Research direction for Engineering Big data Analytics Software," publish by IEEE computer Society, Jan 2015.

[5] Rao, Khan, Begum and Divekar, "Design of E-commerce personalizes service basedon web mining classification technique" IEEE, Intelligence and Computing Research, 2013, pp.1-4.

Websites

1) www.wikipedia.com
2) http://www.smartdatacollective.com/seanmallonbizdaquk/410001/5-benefits-big-data
3) https://www.qubole.com/resources/solution/ecommerce-big-data/