ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 4, April 2018

Product Recommendation from Textual Reviews Using Hybrid Filtering Approach

[1] Sunil Jadhav, [2] Ritambhara Rajeshirke
[1] RDTC SCSCOE,Bhor, [2] RDTC SCSCOE,Bhor

Abstract- **The amount of information on the internet grows rapidly and people need some system to find and access appropriate information. Recommender Systems (RS) are currently used in both the research and in the commercial areas. RS are information filtering systems that deal with the problem of Information Overload. The existing recommendation is based on POI (Point of Interest), Geographical location, User Preference learning and algorithms like LDA, OGRPL (Online Graph Regularized User Preference learning) are used for information extraction and also use the Attribute Pruning (AP), Frank-Wolfe algorithm for improving the performance of the system with some limitation like high retraining cost, unable to capture change in preferences, work for specific value of k. The proposed system recommendation is base on sentiment prediction from textual review using the LDA for feature extraction and is originally based on Hybrid Filtering Approach. K numbers of products are recommended to the user. Hence, the proposed system improves the prediction accuracy of the recommendation system.**

Keywords— RS, Sentiment prediction, Hybrid Filtering Approach, Conflation algorithm.

I. INTRODUCTION

As the data on the internet grows, the number of choices is overwhelming, there is need to filter, priorities and efficiently deliver relevant information in order to solve the problem of information overload. The Recommender system solves this problem by searching through large volume of dynamically generated information to provide user with personalized content and services.

Traditional RS works in three phases as:

1) Information Collection phase
2) Learning Phase

3) Prediction/Recommendation phase

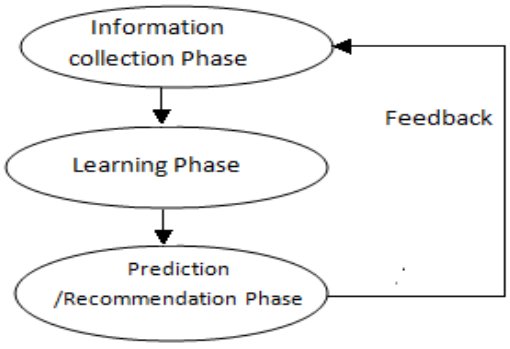

Fig1: Recommendation System

The information collection phase collects relevant information of user to generate a user profile or model for the prediction task including user's attribute, behaviors or content of the resources the user accesses. Information collection phase also includes user feedback and for e-commerce sites need collection of personal information associated with specific user. The learning phase applies learning algorithm and utilize information gather from the information collection phase. The Recommendation or Prediction phase predicts what kind of items of the user may prefer. The use of efficient recommendation technique is very important for system that will provide good and useful recommendation to its individual users.Recently, various approaches are developed which can utilize content base and collaborative filtering technique. Both techniques have strengths and challenges that can be resolved by using hybrid approach to improve the performance of RS.

II. REVIEW OF LITERATURE

A. Content-based Filtering:

The content base technique is a domain depend approach which focuses onanalyzing the attribute of the item for recommendation [9]. The content base considers two things First, user profile: feature extracted from the content of the item that evaluated in the past by user. Second, Item that is positively rated and mostly related to user profile. The CB is having ability to recommend new item even if there is no rating provider by user. The major disadvantage of CB is the need to have an in-depth knowledge and description of the features of the items in the profile i.e. it is dependent on items metadata.CBF uses different models to find similarity as: Vector Space model such as TF/IDF, Decision tree. LIBRA is a content based book RS.

B. Collaborative Filtering:

Collaborative filtering is the most commonly user approach in [1],[3],[5],[6],[7],[8] Collaborative filtering is a domain-independent prediction technique for content that cannot easily and sufficiently described by metadata.CF builds the database of preferences for item by user. Recommendation

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 4, April 2018

is made by calculating the neighbourhood of the user (group of similar user) i.e. it matches the user with relevant interest and preference by calculating similarity between their profile. The technique is divided into two categories: Memory base and Model base. Model base use the previous rating to learn a model in order to improve the performance of CF. Model base use the machine learning technique. Memory base can be further classified into user-base and item-base; in user base it form a group of user having same taste by comparing their rating on same item. Whereas in item base it computes prediction using similarity between items.

The existing system consider different approaches for recommendation such as Point of Interest (POI)[3] recommendation mainly focuses on geographical and social effect.POI is used for graphical location in[3]. In Matrix factorization base approach it will form a matrix base on number of users and number of items can be predicted according to the Eq.(1)$\hat{R}_{u,i}$ denotes the predicted objective. $\bar{R}$ denotes the average value of all ratings.

$$\hat{R}_{u,i} = \bar{R} + U_u P_i^T \quad (1)$$

Sentiment based approach: Sentiment analysis consist of three things: Sentence level, Review level and phrase level. Sentence level classifies the review in different polarity as positive, negative and neutral. Review base approach: In this approach the user numeric ratings and product review are considered from the available number of opinion. In this paper the Hybrid Filtering approach is used for recommendation which is a combination of collaborative and content base filtering approach[2],[11],[12]. In collaborative filtering it will consider and process the data collated in the form of textual review and content base is for considering the rating given by the user to the product.

III. SYSTEM OVERVIEW

The proposed system uses the Hybrid Filtering approach for improving the performance of the Recommender System. The existing approaches consider the limited understanding of the user and item where as the propose approach build recommendation with good understanding of the user interest.

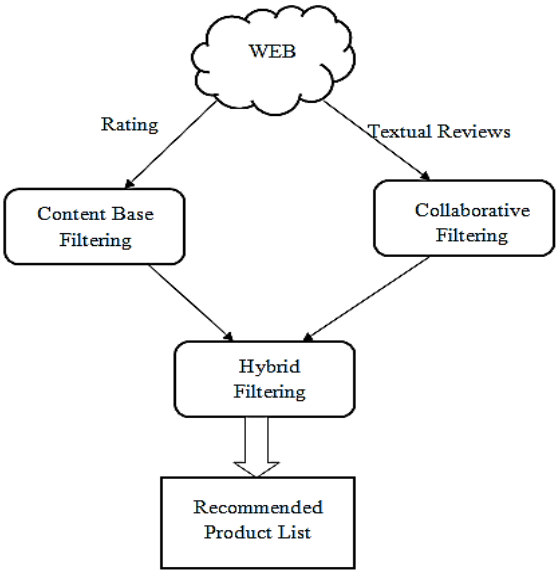

Fig2: Basic system overview

The Content base approach considers details i.e. metadata about the product, where as collaborative considers the textual reviews expressed by user. The result of both techniques is combined called hybrid approach and the product with high relevance is recommended.

IV.PROPOSED SYSTEM

The Recommendation System considers two parameters First, Ratings information and Textual reviews for producing the most relevant prediction according to user query.

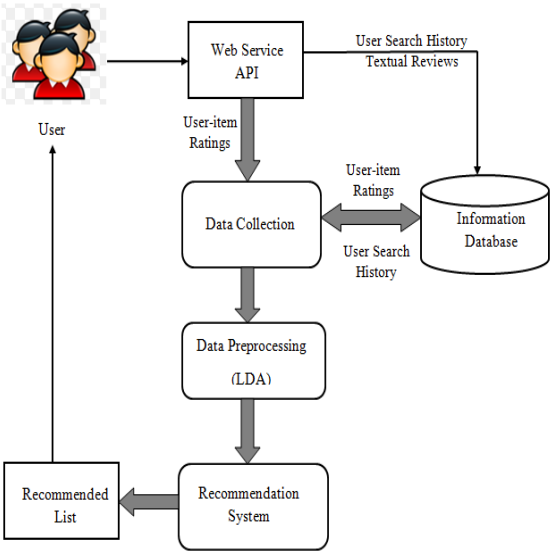

Fig 3: System Architecture

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 4, April 2018

The proposed system consists of three Steps:
1. Data Collection
2. Data Pre-processing
3. Recommendation System (RS)

User interacts with the system through user interface. User interaction details are recorded in the form of web log and stored in the information database.web logs are maintain in the form of plan text files so it can be used for pre-processing.

1. Data collection: consist of data from information database, Item-rating give by the user and the list of reviews expressed by the user in the form of textual review.

2. Data pre-processing: Pre-processing is the technique important for discovering the knowledge from the data collected in the form of text. The task of data pre-processing includes the task of extracting the product feature from the textual review using LDA(Latent Dirichlet Allocation)[1],[3] and calculating the sentiment polarity for the Textual review expressed by user.

A. LDA is used for modelling the relationship between the Textual review, words and topics.Table1 shows the terminology used in LDA.

Tab1: Terminology used in LDA

Term	Description
W	Wordbook, consist of N unique words and word is equated using label{1,2,…,N}
wi∈ {1,2,…,N}	Word, by using character matching each word plot to W having size N
Rm	Review of user or the document Document D = {R1,R2,….Rm}
Tp	Number of Topics
MD	Is the multinomial distribution of the topics specific to m
Ck	Component for each k topic
Am,n	Topic association, topic associated with n-th token in document m
x,y	Dirichlet priors.

LDA data preprocessing starts with the dictionary construction, in which it collect the words by considering the each users review without paying attention on the order of the words. The procedure starts with removing the noise word, stopwords. After done with the word filtering on data the text is clean and ready for generating topics. The dictionary of unique word is constructed in which each word have label wi ∈ {1, 2,…,N}.

Process of LDA:
Input : Users document set D
Output: Topic distribution and topic list per user.

1. Choose dimensional dirichlet random variable per document/review Rm. MD ~ Dirichlet (x).
2. For each topic Tp the conclusion base upon the observation

$$\rho(M, \varphi | D^{train}, x, y) = \sum_A \rho(M, \varphi | A, D^{train}, x, y) P(A, | D^{train}, x, y) \quad (2)$$

3. Repeat step 1 and to for getting output of LDA.

From the above steps we get the user topic distribution with topic list.

B. Next Step is to find the sentiment polarity using the HowNet sentiment dictionary. Sentiment Dictionary (SD) is used for classifying the word from review in positive and negative word. The SD consists of total 8938 words. The Sentiment Degree Vocabulary (SDV) consists of five levels as per the degree of the sentiment word. The dictionary used the 4379 of positive words (PW) and 4650 of negative words (NW).

$$s = \begin{cases} 1 & +ve\ word \\ -1 & -ve\ word \end{cases} \quad (3)$$

The Score for positive word is (1) while the score for negative word is (-1). Equation (4) is used for calculating the score of the Textual Review TRexpressed by User u for the item i.

$$s(T_R) = \frac{1}{N_c} \sum_{w \in c} Q \cdot V_w \cdot R_w \quad (4)$$

$s(T_R)$: Sentiment score of reviewT_R.
c is used for section. Review is split into sections.
Ncis for number of sections.
V_wAssign the value according to the five levels of the SDV.
V_w ={5,4,2,0.5,0.25} for level 1,2,3,4,5 respectively.
R_w : is nothing but the initial score of the word as assign by equation (3) i.e. either 1 or -1.
To increase the accuracy of the score calculation two linguistic rules are added: "and" rule & "but" rule.
"and" implies the similar polarity in two words while "but" implies the opposite polarity in two words.

3. Recommendation system:

After the data preprocessing the input for the recommendation system is the Ratings given by the user. Sentiment scores for the review, User and item details and Recommendation is made base on Hybrid Filtering Approach. Content base filtering is applied to process the Item details by finding the similar item and solves the problem of new item recommendation.
Collaborative filtering is used for making more relevance prediction by analyzing the customer reviews. In proposed

system uses nearest neighbor algorithm [4],[9],[12] for recommending product.

A. General outline for nearest neighbor algorithm:

Input: Number of Items to be recommended N
Number of neighbors used for ranking k
User to recommend items to u
List of all items Items
Rating Matrix R
Output: N item to be recommended.
foreach item ∈ items do
if item € u.rated_items then
item.rank ← rank according to
nearest neighbor(k,u,item);
descending_rank_sort(item);
return top(N,items);
The distance measure formula for K-nn algorithm as stated below:

$$d(x,y) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2} \quad (5)$$

Where
n: number of attribute.
x_k& y_k : kthattribute of x and y data object.

The main advantage of K-nn is the retraining cost is less as compare with the rule base algorithm like Apriori rule base algorithm [5],[10].

B. Evaluation Metrics for the RS are Root Mean Square (RMS) and Mean Absolute Error (MAE).

$$RMSE = \sqrt{\sum_{i \in R_{test}} (\hat{R}_{u,i} - R_{u,i})^2 / |R_{test}|} \quad (6)$$

$$MAE = \sum_{i \in R_{test}} |\hat{R}_{u,i} - R_{u,i}| / |R_{test}| \quad (7)$$

Where,
$R_{u,i}$: is the rating value given by the user u to the item i.
$\hat{R}_{u,i}$: is the recommended rating value.
R_{test}: Represents the test data set.

V.CONCLUSION

The Proposed system is hybrid recommendation system combines the collaborative and content base approaches of filtering and use the K-nn algorithm for recommending the product, which decrease the cost of data retraining and try to solve the problem of new item recommendation by incorporating the content base approach. The propose approach increase the performance by analyzing the user Textual review and providing the most relevant k number of Recommendation to the user of the system.

REFERENCES

[1] Xiaojiang Lei, Xueming Qian, Member, and Guoshuai Zhao,"Rating Prediction based on Social Sentiment from Textual Reviews",IEEE TRANS,2016

[2] Zhou Zhao, Deng Cai, Xiaofei He and Yueting Zhuang,"User Preference Learning for Online Social Recommendation",DOI 10.1109/ TKDE. 2016.2569096, IEEE Trans., 2016.

[3] Hongzhi Yin, Xiaofang Zhou, Bin Cui, Hao Wang, Kai Zheng and Quoc Viet Hung Nguyen," Adapting to User Interest Drift for POIRecommendation",IEEE Trans,2016.

[4] Xavier Amatriain, Alejandro Jaimes, Nuria Oliver, and Josep M. Pujol," Data Mining Methods for Recommender Systems", ICREA,Chapter 2,page no.41-49.

[5] Yining Teng, Lanshan Zhang, Ye Tian, Xiang Li," A Novel FAHP Based Book Recommendation Method by Fusing Apriori Rule Mining", International Conference, 2015

[6] Kishor Barman and Onkar Dabeer," Analysis of a Collaborative Filter Based on Popularity Amongst Neighbors," IEEE Transactions on information theory, vol. 58, no. 12, december 2012.

[7] Jianxun Liu, Mingdong Tang, Member,Zibin Zheng, Member, Xiaoqing (Frank) Liu, Member, Saixia Lyu,"Location-Aware and Personalized Collaborative Filtering for Web Service Recommendation", IEEE Transactions ,2015.

[8] G. Linden, B. Smith, and J. York," Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Comput., vol. 7, no. 1, pp. 76–80, 2003.

[9] Alastair A. Abbott and Ian Watson," Ontology-Aided Product Classification: A Nearest Neighbour Approach", Department of Computer Science, University of Auckland,Auckland, New Zealand.

[10] Kang Liu, Liheng Xu, and Jun Zhao," Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model", IEEE TRANS, VOL. 27, NO. 3, MARCH 2015.

[11] Kaustubh Kulkarni, Keshav Wagh, Swapnil Badgujar, Jijnasa Patil,"A Study Of Recommender Systems With Hybrid Collaborative Filtering", (IRJET) Volume: 03 Issue: 04 | Apr-2016

[12] Mohammad Aamir,Mamta Bhusry," Recommendation System: State of the Art Approach",International Journal of Computer Applications (0975 – 8887) Volume 120 – No.12, June 2015