ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

Recommendation of Product using Hybrid Filtering Approach from Textual Reviews

Deepali B. Jagtap[1], Prof.T.P.Vaidya [2]

P.G. Student, Department of Computer Engineering, Sinhgad College of Engineering, Vadgaon(BK), Pune, India[1]

Associate Professor, Department of Computer Engineering, Sinhgad College of Engineering, Vadgaon(BK), Pune, India[2]

ABSTRACT: Recommender Systems (RS) are currently useful in both the research and in the commercial areas. Traditionally there are three approaches for recommendation Content-base, Collaborative and Hybrid. RS are information filtering systems that deal with the problem of Information Overload. The existing recommendation systems are 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 propose system recommendation is based on Hybrid Filtering Approach. K numbers of products are recommended to the user. Hence, proposed system improves the prediction accuracy of recommendation system.

KEYWORDS: Recommendation System, Content-base, Collaborative, Hybrid Filtering Approach, Conflation algorithm.

I. INTRODUCTION

The explosive growth of available information brings the “information overload” problem. Recommender Systems (RS) are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to user's preferences, interest, or observed behavior about item. Recommender systems help users with item selection and purchasing decisions based on user’s tastes and preferences using a variety of information gathering techniques. Accurate recommendations enable users to quickly locate desirable items without being overwhelmed by irrelevant information. Generally, there are three variants of recommendation approaches: Content based and Collaborative Filtering (CF) based and Hybrid Filtering based approaches. The basic idea of the content based approach is to use properties of an item to predict a user's interests towards it. The key idea of collaborative filtering is to use the feedback from each individual user. CF approaches can be further grouped into model-based CF and neighborhood-based CF. Neighborhood based CF approaches use user-item ratings stored in the system to directly predict ratings for new items. In contrast, model-based CF approaches use user-item ratings to learn a predictive model. The hybrid filtering is nothing but the combination of any two approaches.

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: Information Collection phase, Learning Phase and the Prediction /Recomm - endation phase.

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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

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.

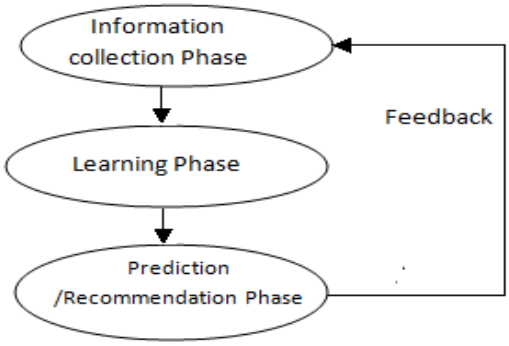


Fig1.1: Traditional Recommendation System

II. RELATED WORK

Generally, there are three variants of recommendation approaches: Content based and Collaborative Filtering (CF)[2] based and Hybrid Filtering[1] based approaches. The basic idea of the content based approach is to use properties of an item to predict a user's interests towards it. The key idea of collaborative filtering is to use the feedback from each individual user. CF approaches can be further grouped into model-based CF and neighborhood-based CF. Neighborhood based CF approaches use user-item ratings stored in the system to directly predict ratings for new items. In contrast, model-based CF approaches use user-item ratings to learn a predictive model. The hybrid filtering is nothing but the combination of any two approaches.

A. CONTENT BASE FILTERING:

The content base technique [5][6] is a domain depend approach which focuses on analyzing the attribute of the item for recommendation. The content base considers two things First, item profile: feature extracted from the detail description of the item. 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[5].

Vectors $\overrightarrow{w_c}, \overrightarrow{w_s}$ Two items with attributes are compared using cosine similarity function as follows:

$$i(c,s) = \cos\left(\overrightarrow{w_c}, \overrightarrow{w_s}\right)$$

$$= \frac{\overrightarrow{w_c} \cdot \overrightarrow{w_s}}{||\overrightarrow{w_c}|| \times ||\overrightarrow{w_s}||}$$

B. COLLABORATIVE FILTERING:

Collaborative filtering [8][9][11] is the most commonly user approach for recommendation. Collaborative filtering is a domain-independent prediction technique for content that cannot easily and sufficiently described by metadata.CF

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

builds the database of preferences for item by user. Recommendation is made by calculating the neighborhood 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;

User base: In the user-based [10] approach, the users perform the main role. If certain majority of the customers has the same taste then they join into one group. If the item was positively rated by the community, it will be recommended to the user. It forms a group of user having same taste by comparing their rating on same item.

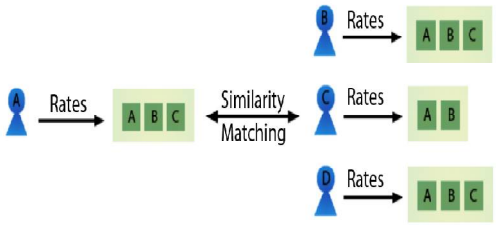


Fig 2.1 User-based collaborative recommender system

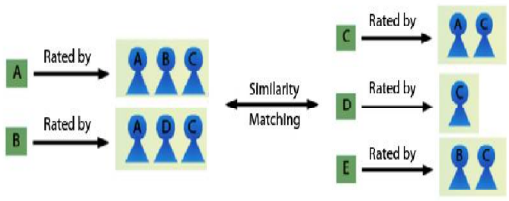


Fig 2.2 Item-based collaborative recommender system

Item Base: Item base [10] it computes prediction using similarity between items. Referring to the fact that the taste of users remains constant or change very slightly similar items build neighborhoods based on appreciations of users. Afterwards the system generates recommendations with items in the neighborhood that a user would prefer

C. HYBRID FILTERING:

For better results some recommender systems combine different techniques of collaborative approaches and content based approaches. Using hybrid approaches [1][4] we can avoid some limitations and problems of pure recommender systems, like the cold-start problem. The combination of approaches can proceed in different ways:

1) Separate implementation of algorithms and joining the results.
2) Utilize some rules of content-based filtering in collaborative approach.
3) Utilize some rules of collaborative filtering in content based approach.
4) Create a unified recommender system that brings together both approaches.

D.TECHNIQUES USED IN TRADITIONAL FILTERING APPROACH FOR RECOMMENDATION

1. Matrix Factorization:

To solve large-scale recommendation problems with sparse data, matrix factorization [5][6] is a state-of-the-art methodology. The matrix factorization based approaches factorize the partially observed user-item matrix into two latent low-rank matrices with the regularization of users' social relations, and then fill the missing data entries by spanning two low-rank matrices.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

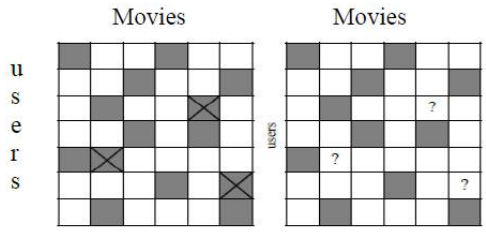
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

2. Probabilistic model based method
Probabilistic model based approaches conclude the probabilistic model from the partially observed user-item matrix and then predict the missing entries based on the probabilistic model as shown in fig. below.



3. nearest neighbor:
The kNN classifier finds the k closest points (Nearest neighbors) [3] from the training records. It then assigns the class label according to the class labels of its nearest-neighbors. The underlying idea is that if a record falls in a particular neighborhood where a class label is predominant it is because the record is likely to belong to that very same class. The most challenging issue in kNN is how to choose the value of k. If k is too small, the classifier will be sensitive to noise points. But if k is too large, the neighborhood might include too many points from other classes
Advantage: the kNN classifier a lazy learner, it does not require to learn and maintain a given model.
Disadvantage: The most challenging issue in kNN is how to choose the value of k. If k is too small, the classifier will be sensitive to noise points. But if k is too large, the neighborhood might include too many points from other classes

E.CHALLENGES AND ISSUES IN RECOMMENDATION SYSTEM:

1. Cold-start
It's difficult to give recommendations to new users as his profile is almost empty and he hasn't rated any items yet so his taste is unknown to the system. This is called the colds tart problem. In some recommender systems this problem is solved with survey when creating a profile. Items can also have a cold-start [8] when they are new in the system and haven't been rated before. Both of these problems can be also solved with hybrid approaches.

2. Sparsity
In online shops that have a huge amount of users and items there are almost always users that have rated just a few items. Using collaborative and other approaches recommender systems generally create neighborhoods of users using their profiles. If a user has evaluated just few items then it's pretty difficult to determine his taste and he/she could be related to the wrong neighborhood. Sparsity is the problem of lack of information.

3. Privacy
Privacy has been the most important problem. In order to receive the most accurate and correct recommendation, the system must acquire the most amount of information possible about the user, including demographic data, and data about the location of a particular user. Naturally, the question of reliability, security and confidentiality of the given information arises. Many online shops offer effective protection of privacy of the users by utilizing specialized algorithms and programs.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

Tab1: Comparison of Recommendation approaches

Sr. No	Approach	Description	Pros/Cons
1	Content Based	Is an Domain dependant. It focuses more on attribute of items to generate prediction. Recommendation made is based on the user profiles and features extracted from the contents.	1. It needed to have in-depth knowledge and description of the features of items in profile.
2	Collaborative	Is a Domain-independent prediction Technique. It calculate similarities with relevant interest and similarities between item's profiles	1.Cold-start problem 2. Data sparsity 3. Scalability
3	Hybrid	The idea is to combine different recommendation techniques in order to gain better system optimization. And creating unified recommendation system.	1. Provide more accurate and effective recommendations than a single algorithm.

III. PROPOSED SYSTEM

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

The proposed system consists of three Steps:
1. Data Collection
2. Preprocessing of Data
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 preprocessing.

1. Data collection:

Data collection phase consist of collecting 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. Preprocessing of Data:

Preprocessing is the technique important for discovering the knowledge from the data collected in the form of text. The task of data preprocessing includes the task of extracting the product feature from the textual review using LDA (Latent Dirichlet Allocation) [1], [5] and calculating the sentiment polarity for the Textual review expressed by user.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

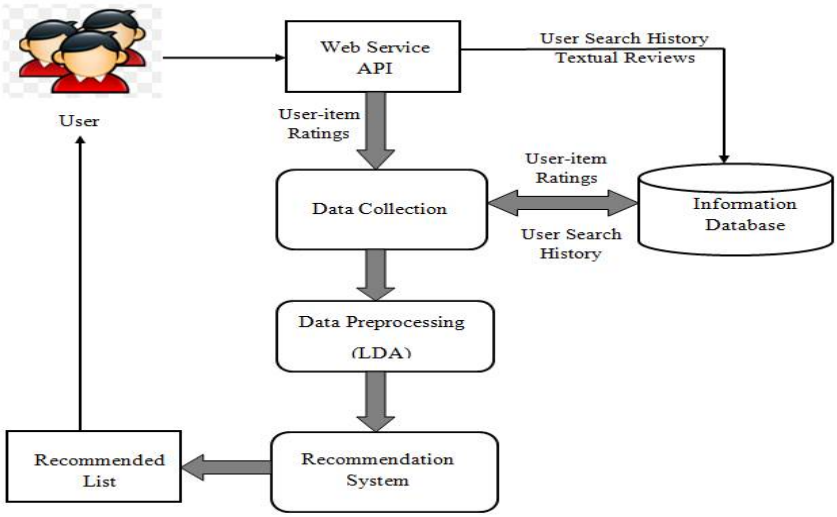


Fig 3.1: System Architecture

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

Tab2: Terminology used in LDA

Term	**Description**
W	Wordbook, consist of N unique words and word is equated using label{1,2,…,N}
$w_i \in \{1,2,\ldots,N\}$	Word, by using character matching each word plot to W having size N
R_m	Review of user or the document Document D = $\{R_1,R_2,\ldots.R_m\}$
T_p	Number of Topics
M_D	Is the multinomial distribution of the topics specific to m
C_k	Component for each k topic
$A_{m,n}$	Topic association, topic associated with n-th token in document m
x,y	Dirichlet priors.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

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, stop words. 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 $w_i \in \{1, 2,\ldots,N\}$.

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 R_m.
2. M_D ~ Dirichlet (x).
3. For each topic T_p the conclusion base upon the observation

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

4. 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 T_R expressed 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.
N_c is for number of sections.
V_w Assign 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
The output of CB is item matrix I_M.
The output of CF is user matrix U_M.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

The combine result of $I_M * U_M$ is used to produce the hybrid matrix as UI_M matrix and this used as an recommendation result.

$$UI_M = I_M * U_M$$

A. General outline for matrix factorization algorithm:
Input: Set of user U, item I
Matrix of user-item ranking R,
Number of Features F
Number of Items to be recommended *N*
Output: Top-N Recommendation R (U) ∈ I
For i← 1 to numIterations do
foreach user ∈ U and item ∈ I with rating R[u,i] do
predicted rating = M[u] * M[i]
return N item with highest predicted ranking

Matrix of interaction is divided into two small matrixes one for user and one for item. The (u,i) matrix is obtain by multiplying these two matrix.

Each item I and user U is having d- dimensional word
Collection of item is given by:

$$I = [I_{1,\dots}, I_x] \in R^{d \times x}$$

Where, x is the total number of items.

Collection of user is given by:

$$U = [U_{1,\dots}, U_y] \in R^{d \times y}$$

Where, y is the total number of user.

User-item rating matrix is given by:

$$R \in R^{x \times y}$$

The value of the R is used by the Recommender System.

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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

$\hat{R}_{u,i}$: is the recommended rating value.

R_{test}: Represents the test data set.

IV. CONCLUSIONS

The Proposed system is hybrid recommendation system combines the collaborative and content base approaches of filtering and uses the matrix factorization method for recommending the product, which decrease the cost of data retraining and try to solve the problem of new item by incorporating the content base approach and cold-star problem in the context of collaborative filtering approach. The propose approach increase the performance by analyzing the user Textual review and description of the item using the LDA and sentiment measurement and providing the most relevant k number of Recommendation to the user of the system.

REFERANCES

[1] 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.
[2] Hongzhi Yin, Xiaofang Zhou, Bin Cui, Hao Wang, Kai Zheng and Quoc Viet Hung Nguyen," Adapting to User Interest Drift for POI Recommendation",IEEE Trans,2016.
[3] Alastair A. Abbott and Ian Watson," Ontology-Aided Product Classification: A Nearest Neighbour Approach", Department of Computer Science, University of Auckland,Auckland, New Zealand.
[4] Kaustubh Kulkarni, Keshav Wagh, Swapnil Badgujar, Jijnasa Patil,"A Study Of Recommender Systems With Hybrid Collaborative Filtering", (IRJET) Volume: 03 Issue: 04 | Apr-2016
[5] Mohammad Aamir,Mamta Bhusry," Recommendation System: State of the Art Approach",International Journal of Computer Applications (0975 – 8887) Volume 120 – No.12, June 2015
[6] Xavier Amatriain, Alejandro Jaimes, Nuria Oliver, and Josep M. Pujol," Data Mining Methods for Recommender Systems", ICREA,Chapter 2,page no.41-49.
[7] Yining Teng, Lanshan Zhang, Ye Tian, Xiang Li," A Novel FAHP Based Book Recommendation Method by Fusing Apriori Rule Mining", International Conference, 2015
[8] 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.
[9] 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.
[10] 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.
[11] 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.