AMSE Association for the Advancement of Modelling & Simulation Techniques in Enterprises
IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Advances in Modelling and Analysis B
Vol. 61, No. 2, June, 2018, pp. 70-75
Journal homepage: http://iieta.org/Journals/AMA/AMA_B

Recommendations of products based on combination of collaborative, content and pearson filtering

Jyostna D. Bodapati, Naralasetti Veeranjaneyulu*, Malkari M. Rao

Vignan's Foundation for Science, Technology and Research, Vadlamudi 522213, AP, India

Corresponding Author Email: veeru2006n@gmail.com

Received: April 18 2018
Accepted: May 31 2018

Keywords:
content, collaborative and pearson recommendations

ABSTRACT

In today's world the numbers of ecommerce companies are increasing day by day and also huge number of products is coming into the market. When customers want to buy the products they generally just see a numerical rating of the products and then purchase them. Later they come to know that products are not good. 3 kinds of rating systems are implemented in this work namely collaborative rating, content based rating and Pearson rating. In the implementation makes use of latest technology stack namely spring framework for the backend and Ext JS Framework for the front end. Content Based recommendations are based on user past transactions, Collaborative recommendations are based on rating from across the users and Pearson recommendations takes logged in user and other user ratings to provide better quantifying recommendations

1. INTRODUCTION

In the competitive world the number of companies is increasing day by day. Each of the companies will say their product is good but the new buyer do not have knowledge whether the given product is good or not until he/she buys the product. This is a disadvantage for the consumers. Top companies are providing a rating system on the basis of STARS where stars are high means product is good. All these are based on the numerical scale. Text Mining is the most significant subject which can solve the issues.

Text mining is an approach in which a lot of quality information can be extracted by performing the analysis on the data and then identifies patterns or come to conclusions or recommendations. E-Commerce makes use of electronic communications and digital transactions to create a value added relationship between merchant and the consumer. The steps taken by users regarding shopping has changed with the evolving times. The internet has rapidly motivating the life of people and large portion of population moves online. Hence there is a need of intelligent application which can provide easy mode of payments, create new first impressions, makes a lot of user specific recommendations based on the user's alignments towards the product and help user to have the best product.

2. BACK GROUND

There is plenty of ecommerce system [1] and the user performs a normal process of buying like selecting the book, entering the credit/debit/account details and then purchases the book. This paper discusses about the content based and collaborative based approach in which the users will be able to provide a direct rating to the books on the numerical rating. Finally books are ranked based on highest numerical rating and provided to the end user. The content based recommendations algorithm works on the basis of the history of products bought by the user. If the user has purchase a product and number of such purchases exceeds threshold then product is recommended otherwise not recommended. Another form of rating system for collaborative users. The user behavior based on product purchases need to be obtained.

The collaborative rating is based on just a rating on a scale of 1-5 and any user can give the rating. This does not relate to any specific feature of the product. The content based filtering requires the user to set the threshold and also requires the third party transactions for analysis.

As the investment [2] in the markets is increasing and also the big giants like Amazon etc coming to the Middle East and India. They provide in their system the review options for the end user and the users can provide comments on the products. There is huge amount of data that is available but it not based use of product. This paper provides the literature review on the rating systems that are currently in use and benefits and disadvantages of works in the area. Simple rating system based on Starts i.e 1-2 product is bad , 3-4- product is ok above 4 product is awesome. The opinion approach which takes into consideration of reviews and computes the positive and negativeness of products based on sentiments. It does not take into consideration the common features of products which is very much important for the users who want to look into the specific features of the product.

In this paper [3] phishing email concepts have been discussed. First the emails data sets are taken and then text mining is applied on the data sets and emails are classified as attack or non attack. A new technique for determination of fraud has been proposed. No consideration of a specific category or concepts just a plain set of words have been considered and then the email is regarded as harmful or non harmful. The challenges [4] involved in the collaborative based filtering namely data sparsely, scalability, synonymy, gray sheep and shilling attacks. Finally determine the predictive performance.

Recommendations systems [5] provide an effective way by which there can be guidance for an end user which are best

products. Content Based, Collaborative Based and Association Rule mining methods have been given. A content Based recommendation makes use of user behavior. Collaborative based filtering works based on all users rating and finally association rule mining is used based on intersection of content and collaborative based filtering. The work introduces Association Rule Mining which is new way of recommendations.

The Content Based Recommendations and Collaborative Based rating are very much required to find out association rule mining set.

3. ALGORITHMS

3.1 Collaborative filtering algorithm

In Collaborative based recommendations module the users which are unregistered or register each of the users will be able to rate the product based on the numerical rating. After that the aggregated sum is performed. Finally the ranking is performed based on maximum aggregated sum of rating.

3.1.1 Drawbacks of collaborative rating

[1] The entire rating is based on the numerical scale and it does not take into consideration the emotions or whether the products are quality products and lot of fraudulent ratings can also be given which will increase the advantage for a wrong product.

3.1.2 Description of the method

Consider a Set of Products $\{p_1,p_2,p_3,.....p_n\}$ where p_i = i^{th} product and n = number of products

Consider a set of N Users $\{u_1,u_2,u_3,u_4,.....,u_m\}$ *Where* $u_i = i^{th}$ *user*

$m = number\ of\ users$

After each of the users rate the products then a n*m matrix is generated with the ratings across the users. For example consider the following set

Table 1. Collaborative rating computation, an example

Product	U_1	U_2	U_3	TR
P1	**5**	**2**	**3**	10
P2	**4**	**2**	**1**	7
P3	**3**	**5**	**5**	13
P4	**2**	**5**	**4**	11
P5	**1**	**1**	**2**	4

In the above matrix there are 5 products namely P1, P2, P3, P4 and P5. U1, U2 and U3 are the users. A value 5 is the rating given by the user U1 for Product P1. Like this there are many values namely indicated from other users and products. The last column is the total aggregated rating. The TR value in the first row is the aggregated rating from all 3 users. The total rating is computed using the following equation:

$$TR_p = \sum_{i=1}^{m} R_i$$

Where,

$n = Number\ of\ products$

$R_i = rating\ of\ i^{th}\ user\ for\ the\ product$

$1 \le P_i \le m$

$P = \Pr oduct$

$m = Number\ of\ users$

After the matrix n*m is computed the products are recommended based on order of total aggregated rating. Based on the data provided in table1, P_3 is highly recommended and P_5 is least recommended.

3.2 Content based recommendations

Content Based Recommendations is user specific data in order to provide recommendations. In this module the user will select a product and enter the credit card details and then completes the transactions. Behind the scenes the merchant maintains the transactions and then finds the best products suited for the user.

3.2.1 Advantages

[1] The history profile of the transactions are used in order to recommend the products for the user.

3.2.2 Drawbacks

[1] In this approach appropriate threshold has to be selected.

[2] The new users do not get any recommendations because there is no buying history for the user.

This Module analyzes the transaction history of the buyers through the transaction logs. Customized threshold is set for each user. If any of the products exceeds the specific threshold then recommendations of the product are suggested.

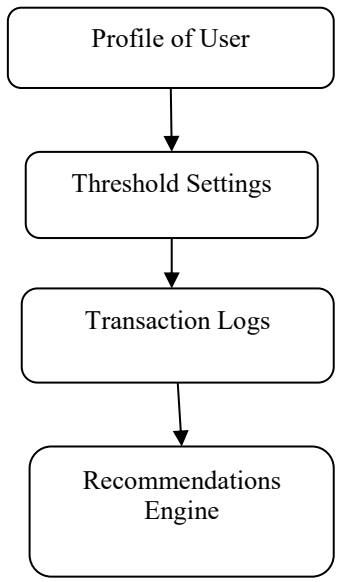

Figure 1. Content based recommendations

Fig. 1 shows that the profile of the user and his/her threshold settings along with transaction logs acts as an input for recommendations engine and recommendation engine filters the transaction based on threshold criteria.

Description of content based recommendations

1. The user sets a threshold T known as personal settings
2. After the user purchase the product then the tracking is performed for each of the user. The order details and order information is maintained in the following way

The order information is performed using the following parameters

#	Name
1	ORDERID
2	PRODUCTID
3	QUANTITY

ORDERID- Order Id of the product
PRODUCTID- Id of the product purchased by user
QUANTITY- Number of Such products purchased during each transaction

#	Name
1	ORDERID
2	LOGINID
3	ORDERDATE
4	TOTALAMOUNT
5	EMAIL

ORDERID- Order Id of the product
LOGINID- User Id of the user read from the session
ORDERDATE- Date of Purchase for the product
TOTALAMOUNT- The total amount incurred for the product purchase
EMAIL- Email Id of the user

3. After a set of product purchases are performed
4. If the product purchases exceeds the threshold T then the product falls under the content based recommendations list

3.3 Pearson recommendations

This kind of recommendations is given for the authenticated user. The ratings are taken from the user as well as other users and then based on prediction rating the products are recommended. The recommendations are computed by using the following equation:

$$s(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - \overline{r_u})^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - \overline{r_v})^2}}$$

where

$r_u = rating \ \ from \ user \ u$

$\bar{r}_u = average \ of \ \ all \ ratings$

$I_u = set \ of \ \ items \ rated \ by \ \ user$

The prediction is substituted using the following formula

$$p_{u,i} = \overline{r_u} + \frac{\sum_{u' \in N} s(u,u')(r_{u',i} - \overline{r_{u'}})}{\sum_{u' \in N} |s(u,u')|}$$

Description of pearson rating

Consider a Set of Products

$\{p1, p2, p3, ..., pn\}$ *where* $pi = i^{th}$ *product* *Consider*

$n = number \ of \ \ products$

a set of m Users

$\{u1, u2, u3, u4, ..., um\}$ *Where* $u_i = i^{th} \ \ user$

$m = number \ of \ \ users$

[1] The rating provided by the logged in user and it is in the format similar to table1. It is an n*m matrix is generated with the ratings across the users.

[2] The rating for the other users are also considered in the ranking of the products during prediction of the product rating.

4. IMPLEMENTATION DETAILS

The architecture for the implementation follows the best practices of the industry with loose coupling between various layers of the code. The implementation makes use of Spring Framework [12] specifically 3 major blocks namely Spring IOC, Spring JDBC and Spring MVC. Spring IOC creates objects automatically which increases the performance of the application. Spring JDBC is responsible for any kind of Create, Retrive, Update and Delete operations of the database. Spring MVC is responsible for interaction between front end and backend. The User Interface has been implemented using Ext JS and Java Server Pages.

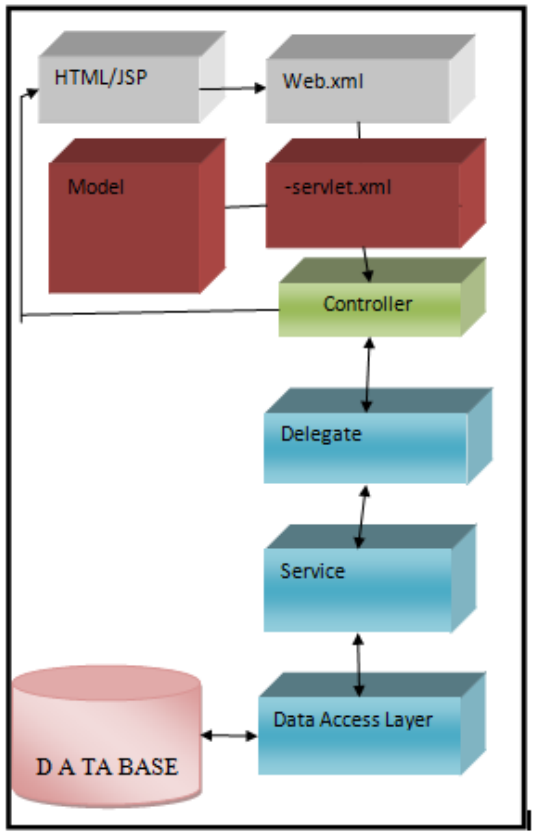

Figure 2. Architectureof the proposed system

Figure 2 gives description about the system architecture which is used in the implementation.

Figure 2 shows that the user interface designed in the HTML/JSP pages and then the request goes to the web

container and web container verifies the request in the web.xml file by looking first into the url pattern and then it goes to the servlet name and then it searches for the corresponding servlet name in the servlet tag and looks into the servlet class and creates an object of Action Servlet and then the action servlet will delegate its job to Request Processor. The request processor will look for the action to which must be called in looked up in the stucts-config.xml and corresponding action form is called and then the action is called. The action class will then call the delegate , then the delegate calls the service and service calls the Data Access layer and results goes exactly in the opposite way and the resultant JSP page is loaded

Model: This is the Plain Old Java Object which will have the getters and setters and setters gets automatically called and data the user has entered will be available.

Controller: This is the class which is used to fetch the user entered data and then processes it and calls the delegate layer and obtains the results.

Delegate: Delegate is the layer which contains nothing but call to an appropriate service.

Service: This is the layer which is responsible for entire algorithmic implementation. This is the layer which contains the heavy weight implementation of entire algorithms. Future the service would require the help of Data Access Layer for some operations and many other helper classes.

Data Access Layer: This is the layer which deals with only the CRUD operations namely Create, Retrieve, Update and Delete. It has no other usage. This layer has been used in order to fetch the data from the routing tables.

Database: This is the place where all the tables would have been placed.

5. RESULTS

In this section the results have been discussed regarding the 3 kinds of recommendations system.

5.1 Registration module UI

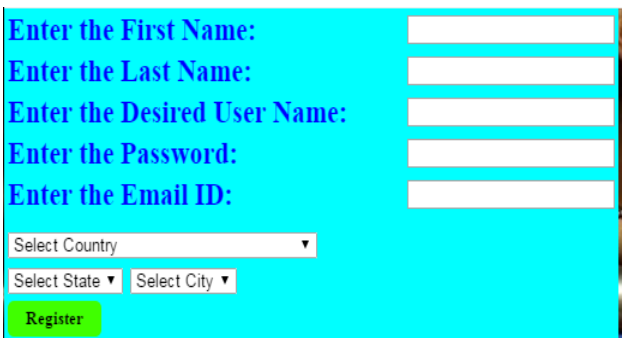

Figure 3. Registration module

Fig. 3 shows the registration page in which all the attributes namely First Name, Last Name, Desired User Name, Password, Email, Country, State and City.

5.2 Login module

Fig. 4 shows the login module for the user. if valid username and password are given then user is allowed to login otherwise not.

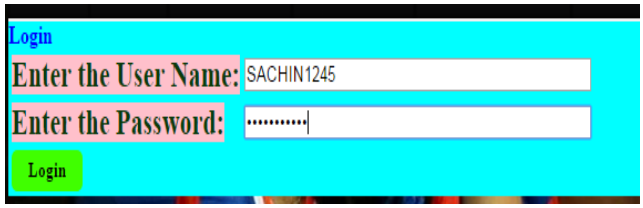

Figure 4. Login module

5.3 Collaborative filtering

Fig. 5 shows the direct rating as shown in the fig the product selected is Samsung Galaxy S3 and then rating 4 is applied.

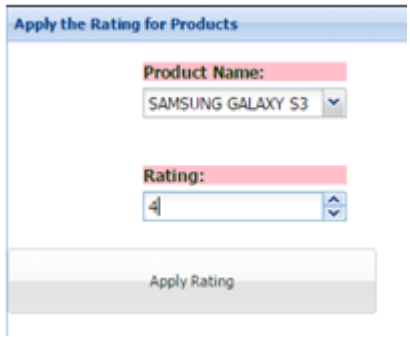

Figure 5. Direct rating

Figure 6. View direct rating

Fig. 6 shows the view direct rating module. As shown in the fig the product name and then aggregated rating for the product.

Figure 7. Collaborative rating graphs

Fig. 7 shows the collaborative rating graphs as shown in the fig nokia Lumina has the highest rating as compared to Samsung galaxy S1 and Samsung galaxy s3.

Product Name	Rating
NOKIA LUMINA	9
SAMSUNG GALAXY S1	4
SAMSUNG GALAXY S3	4

Figure 8. Collabor
ative based ranking of products

Fig. 8 shows the collaborative based recommendations for the product based on the order of aggregated rating.

5.4 Content based recommendations

Fig. 9 shows the Product Purchase page in which the user is able to view Product Name, Author, Product Details and its image and a text box where the user can enter number of purchases the user wants to purchase at a time.

Figure 9. Product purchase page

Figure 10. Banking page

Fig. 10 shows the banking page where the user can enter the account number and password which helps in completing the purchase of the product.

Figure 11. Personal settings

Fig. 11 shows the personal settings. As shown in the fig a value of 2 is specified as the threshold by the user.

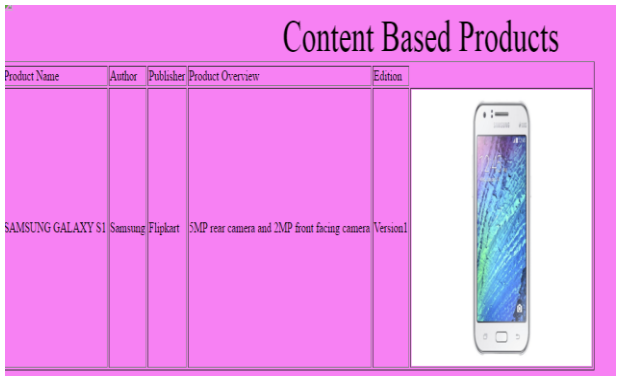

Figure 12. Content based recommendations

Fig 12 shows the content based recommendations as shown in the fig Samsung Galaxy S5 is recommended for the user based on transaction history as it exceeds the personal settings.

5.5 Pearson based recommendations

Product Name	Product ID	Predicted Ratings
NOKIA LUMINA	3	0.933333333333333
Apple I Phone6	5	0.166666666666667
SAMSUNG GALAXY S1	1	-0.233333333333333
SAMSUNG GALAXY S3	2	-0.666666666666667
LG	4	-1.4

Figure 13. Pearson recommendations

Fig. 13 shows the Pearson based recommendations for the products namely Nokia Lumina, Apple I Phone, Samsung Galaxy S1, Samsung Galaxy S3 and LG as shown in the fig products are ranked based on descending order of the predicted ratings. The predicted ratings are computed by using the Pearson rating formula

6. CONCLUSIONS

In this work 3 kinds of data mining recommendation systems have been implemented namely Content Based, Collaborative Recommendations and Pearson Rating Recommendations. Each of the recommendation systems have their own way of determining for the user which products are

the best based on rating from other users, specific user as well as transaction history.

The current work does not take into consideration the sentiments expressed by the users who have already used the products it considers only numerical sentiments. There are many sentiment analysis algorithms available but those algorithms does not consider feature and demographic based information while calculating the positive and negative sentiments of the product and recommendation it to the user.

REFERENCES

[1] Praveena M, Bincy K, Vinayak H. (2016). Book recommendation system through content based and collaborative filtering method. 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE).

[2] Yee W.Lo; Vidyasagar P. (2009). A review of opinion mining and sentiment classification framework in social networks. 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies.

[3] L'Huillier G, Hevia A, Weber R, Ríos S, (2010). Latent semantic analysis and keyword extraction for phishing classification. 2010 IEEE International Conference on Intelligence and Security Informatics.

[4] Sachan A, Richariya V. (2013). Survey on Recommender System based on Collaborative Technique. Department of Computer Science And Engineering intemational journal of innovations in engineering and technology(IJIET) 2(2): 1-7.

[5] Tewari AS, Kumar A, Barman AG. (2014). Book recommendation system based on combine features of content based filtering and association rule mining. IEEE International I Advance Computing Conference(IACC), pp. 502.

[6] Sie SH, Yeh JH. (2014). Library book recommendations based on latent topic aggregation. International Publishing Switzerland, 411-416.

[7] Pranav B, Navinkumar A. (2015). Book recommendation system using opinion mining technique. International Journal of Research in Engineering andTechnology(IJRET) 4(1): 333.

[8] Sun D, Luo Z, Zhang F. (2011). A novel approach for collaborative filtering to alleviate the new item cold start problem. IEEE 11th International Symposium on Communications and Information Technologies (ISCIT) 402-406.

[9] Xie F, Xu M, Chen Z. (2012). RBRA: A simple and efficient rating-based recommender algorithm to cope with sparsity in recommender systems. IEEE 26th International Conference on Advanced Information Networking and Applications Workshops.

[10] Chikhaoui B, Chiazzaro M, Wang S. (2011). An Improved Hybrid Recommender System By Combining Predictions. IEEE Workshops of International Conference on Advanced Information Networking and Applications, pp. 644-649.

[11] Deng XY, Fei F, Huang F. (2017). An effective recommendation model using user access sequence and context entropy. AMSE Journals, AMSE IIETA Publication, Series: Advances B 60(1): 57-73.

[12] Moghaddam SG, Selamat A. (2011). A scalable collaborative recommender algorithm based on user density-based clustering. 3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMiA), pp. 246-249.