www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 4 April 2017, Page No. 21149-21153
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i4.61

Personalized Book Recommendation System

Jayanti Rathnavel[1], Kavita Kelkar[2]
[1]Student MTech, Department of Computer Engineering,
K J Somaiya College of Engineering, Mumbai.
jayanti.r@somaiya.edu
[2]Assitant Professor, Department of Computer Engineering,
K J Somaiya College of Engineering, Mumbai.
kavitakelkar@somaiya.edu

Abstract: Recommender systems are found in many ecommerce applications today. Recommender systems usually provide the user with a list of recommendations that they might prefer, or supply predictions on how much the user might prefer each item. Choosing what book to read next has always been a question for many. Even for students, deciding which textbook or reference book to read on a topic unknown to them is a big question. There are two common approaches for providing recommendations, they are collaborative filtering and content based filtering. In this report we try to present a model for a personalized recommendation system for books that uses hybrid recommendation approach which is combination of content based and collaborative filtering. The proposed recommendation system tries to learn the user's preferences and recommends the books to the user based on their preferences. The system also recommends the books to the user based on the user's demographic parameters like age and location. The system also tries to understand the user's favourite author and recommends accordingly.

Keywords: Recommendation System, Collaborative filtering, Content-based filtering, Demographic parameters.

1. Introduction:

Recommender systems have become popular from the last decade. Since the number of products has grown in number, the need for recommender systems has also increased. Recommendation system tries to predict the interest of a user and recommend products that match their interest as accurately as possible. Also, e-commerce business will be profited by the increase of sales which will obviously occur when the user is presented with more items that he/she would likely found to match the interest. It simplifies the users job thus making it beneficial for both the user as well as the dealer. They also help the users to manage their reading list by knowing their preference. Recommendation system tries to predict the interest of a user and recommend products that match their interest as accurately as possible. Also, e-commerce business will be profited by the increase of sales which will obviously occur when the user is presented with more items that he/she would likely found to match the interest. There are basically two approaches for Recommendation System:

Collaborative filtering: This approach builds a model from a user's past behavior as well as similar decisions made by other users to predict items that the user may have an interest in.

Content-based filtering : In content-based filtering approach the characteristics of an item are analyzed to recommend items to the user.

A recommender platform with the combination of all these techniques is also possible. It is called a Hybrid Recommender System. Different techniques have been developed over time to give accurate recommendations. Apart from the regular filtering techniques, other approaches are being adopted. Ontology-based

recommendations, Demographic-based recommendations have gained importance in recent times. Natural Language Processing is also being incorporated nowadays to analyze user feedback. Context aware recommendations too are gaining popularity[3].

Issues or Challenges faced by Recommendation System:[23]

1.Cold-start problem :

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 cold start problem. In some recommender systems this problem is solved with survey when creating a profile. Items can also have a cold-start 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. Trust:

The voices of people with a short history may not be that relevant as the voices of those who have rich history in their profiles. The issue of trust arises towards evaluations of a certain customer. The problem could be solved by distribution of priorities to the users.

3. Scalability:

With the growth of numbers of users and items, the system needs more resources for processing information and forming recommendations. Majority of resources is consumed with the purpose of determining users with similar tastes, and goods with similar descriptions. This problem is also solved by the combination of various types of filters and physical improvement of systems. Parts of numerous computations may also be implemented offline in order to accelerate issuance of recommendations online.

4. 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 its pretty difficult to determine his taste and he/she could be related to

the wrong neighborhood. Sparsity is the problem of lack of information.

5. 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.

2. Literature Review

The literature survey on the existing recommendation system for books as well as the literature survey on the recommendation system for other applications are as follows :

2.1 Recommendation System and its Approaches

Mukta Kohar, Chhavi Rana in their paper have described what is meant by recommendation system, the various approaches like content-based, collaborative filtering and hybrid approaches that are involved for making recommendations[22]
Lipi Shah, Hetal Gaudani, Prem Balani in their work have described different techniques involved in content-based , collaborative filtering and hybrid approach and the issues related in each of the category[20].

2.2 Problems related to Recommendation System

Soanpet.Sree Lakshmi, Dr.T.Adi Lakshmi in their work have described in detail the problems like overspecialization, data sparsity, cold start, scalability, ranking of the recommendation, etc. that are related to the recommendation system[21].

2.3 Existing Recommendation system for books

S. Vinodhini, V. Rajalakshmi, B. Govindarajalu proposed a personalized recommendation system which recommends books to the user based on the keyword extraction[1] using hadoop and the ratings given by user. They also implemented Region Aggregation will recommending the book.

Pijitra Jomsri developed a library book recommendation system based on user profile loaning and apply association rule to create model[2].

Salil Kanetkar, Akshay Nayak, Sridhar Swamy ,Gresha Bhatia proposed web-based personalized hybrid book recommender system[3] which exploits varied aspects of giving recommendations apart from the regular collaborative and content-based filtering approaches. Temporal aspects for the recommendations are incorporated. Also for users of different age, gender and country, personalized recommendations can be made on these demographic parameters. Scraping information from the web and using the information obtained from this process can be equally useful in making recommendations.

Punit Gupta and Ravi Shankar proposed a tagging[6] based evolving recommendation system for digital library system for user and library administration.

Guangqian Zhangl, Wei Sunl they performed a data analysis in order to explore the
importance of different attributes[7] in the user preference to a book and provided the suggestion for book recommendation design.

CaiNicolas Ziegler et al (2005) proposed a recommendation system that considers a concept called topic diversification[8]. According to this concept, the list of top n recommendation will be balanced as the users extended interest will also be taken into account. Thus the user will not be bored upon the similar kind of recommendations often made. The concept of User-based Collaborative filtering and Item-based Collaborative filtering are combined and the recommendations are made.

2.4 Other Recommendation system

1. Recommendation system for recommending courses to students:

Boban Vesin et al (2012) developed a recommendation system termed as PROTUS (PRogramming TUtoring System)[10] that recommended courses to the students. The courses are usually recommended to the students based on their age and domain of study but in this system semantic web technology concepts are used. Navigation patterns are obtained from the past history of the student and from that pattern, future recommendations are made.

2. Personalized online news recommendation system:

Saranya.K.G and G.Sudha Sadhasivam, proposed personalized news recommendation approach based on dynamic updating policy and collaborative filtering[12]. This paper also solves the problem of scalability asscociated with personalized news recommendation system with the help of hadoop framework.

3. Recommendation system for music:

Brian McFee et al(2012) developed a recommendation system for music by learning the content similarity[14]. It used content based similarity method initially and then collaborative similarity method is imposed on the results. It avoided the cold start problem and the overhead of query-to-answer technique.

4. Recommendation for fashion Company Lyst[16]:

Maciej Kula in his work presents a hybrid matrix factorization model and proves that this model outperforms both collaborative and content based models in cold-start or sparse interaction data scenarios. He compares all the hybrid algorithm with lightfm algorithm and shows the lightfm algorithm outperforms the rest of the algorithm.

3. Proposed System

The idea of the system is to develop a recommendation engine that can recommend books to the users with increased accuracy by analyzing the interest of the user and features of the books.. The data set considered is a large set of books which is a big data. The proposed model tries to eliminate the problems like cold start problem by using demographic based recommendation, overspecialization problem by using lightfm model which tries to predict books in such a way that the recommendation list contains book which has not been explored by the user yet.

3.1 Architecture

The overall architecture of the developed system is given below:

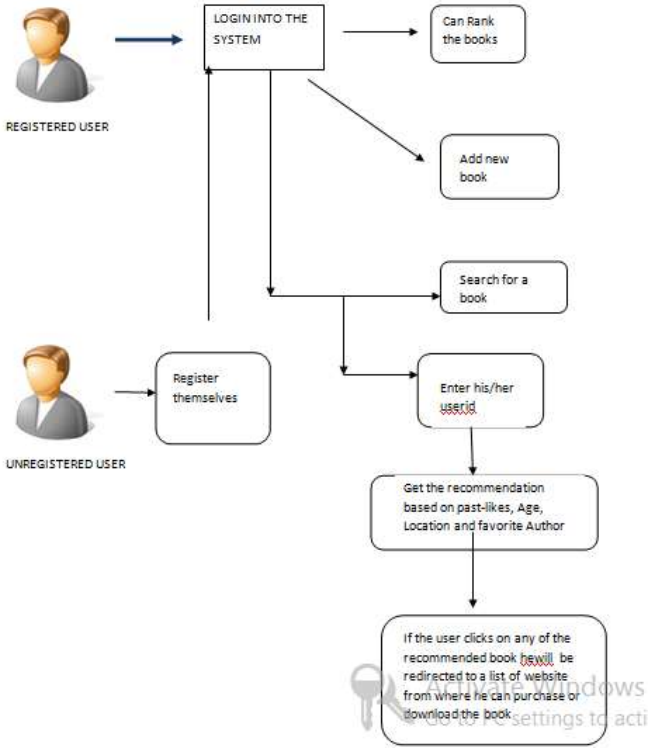

Figure1:Proposed System

A new user has to register himself in the system by clicking the register link present in the home page. He has to fill in the details like e-mail id, password, location, Age, profession in the register page and submit it. On submission the system redirects to the homepage. Then, the user has to login with his registered e-mail id and password to access his account.

After login, three options are seen they are to add a new book, to search for a book and to get the recommendations.

To add a new book the user has to click on the "new book" link where he can add book to the database by filling the details like book-title, author, ISBN, publisher and year of publication.

To search for a book the user has to give the keyword present in the book title and click on search. A list of books will be displayed which will contain that keyword which the user has searched for. On clicking it he/she will be redirected to google page which consist of various website selling that book and the reviews for that book.

To get the recommendation the user has to click the "Recommendation "link. If the user is a new user he can rank some books by clicking the given link, he will be redirected to the initial ranking page with the names of five books where he has to provide ratings to each book with the value of 0 to 10 and submit it.

After rating the book he can enter his user-id and submit it inorder to get the recommendations. He will get the recommendation of two books each for based on past-likes, based on Age, based on location and based on user's favorite Author.

The list of books that are displayed under the recommendation, are links to the google page which contains the information of the website where the user can get the information about the price of the book in different sites. If free e-book copy of the book is available he can download it.

3.2 Dataset Collection[19]

The dataset of books are collected from IIF (Institute For Informatik Freiburg). The dataset Contains 278,858 users (anonymized but with demographic information) providing 1,149,780 ratings (explicit / implicit) about 271,379 books.

The dataset comprises 3 tables. They are as follows:

BX-Users

Contains the users. Note that user IDs (`User-ID`) have been anonymized and map to integers. Demographic data is provided (`Location`, `Age`) if available.

BX-Books

Books are identified by their respective ISBN. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given (`Book-Title`, `Book-Author`, `Year-Of-Publication`, `Publisher`), obtained from Amazon Web Services. Note that in case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours (`Image-URL-S`, `Image-URL-M`, `Image-URL-L`), i.e., small, medium, large. These URLs point to the Amazon web site.

BX-Book-Ratings

Contains the book rating information. Ratings (`Book-Rating`) are either explicit, expressed on a scale from 1-10 (higher values denoting higher appreciation), or implicit, expressed by 0.

3.3 Recommendation Algorithm[16]

To describe the model formally, let U be the set of users, I be the set of items, FU be the set of user features, and FI the set of item features. Each user interacts with a number of items, either in a favorable way (a positive interaction), or in an unfavorable way (a negative interaction). The set of all user-item interaction pairs (u; i) εU * I is the union of both positive S+ and negative interactions S-.Users and items are fully described by their features. Each user u is described by a set of features fu ε FU. The same holds for each item i whose features are given by fi ε FI .

The features are known in advance and represent user and item metadata.

The model is parameterized in terms of d-dimensional user and item feature embedding's eUf

and eIf for each feature f. Each feature is also described by a scalar bias term (bUf for user and bIf for item features). The latent representation of user u is given by the sum of its features' latent vectors:

$$q_u = \sum_{j \in f_u} e_j^U$$

The same holds for item i:

$$p_i = \sum_{j \in f_i} e_j^I$$

The bias term for user u is given by the sum of the features' biases:

$$b_u = \sum_{j \in f_u} b_j^U$$

The same holds for item i:

DOI: 10.18535/ijecs/v6i4.61

$$b_i = \sum_{j \in f_i} b_j$$

The model's prediction for user u and item i is then given by the dot product of user and item representations, adjusted by user and item feature biases:

$$\hat{r}_{ui} = f\left(q_u \cdot p_i + b_u + b_i\right)$$

There is a number of functions suitable for f(.). An identity function would work well for predicting ratings; but interested in predicting binary data so choose the sigmoid function

$$f(x) = \frac{1}{1 + \exp(-x)}.$$

:

The optimization objective for the model consists in maximizing the likelihood of the data conditional on the parameters. The likelihood is given by

$$L\left(e^U, e^I, b^U, b^I\right) = \prod_{(u,i) \in S^+} \hat{r}_{ui} \times \prod_{(u,i) \in S^-} (1 - \hat{r}_{ui})$$

Loss function:

Loss function measure the errors made by predictions on training data. The loss function used for this hybrid recommendation system is Weighted Approximate Rank Pair-wise loss.

For the sake of discussion, consider the problem of recommending items i to users u, where a scoring function fu(i)gives the score of item i for user u, and the item with the highest score is recommended.

WARP considers each observed user-item interaction (u,i) in turn, choses another “negative” item i' that the model believed was more appropriate to the user, and performs gradient updates to the model parameters associated to u, i and i' such that the models beliefs are corrected. WARP weights the gradient updates using (a function of) the estimated rank of item i for user u. Thus the updates are amplified if the model did not believe that the interaction (u,i) could ever occur, and are dampened if, on the other hand, if the interaction is not surprising to the model. Conveniently, the rank of i for u can be estimated by counting the number of sample items i' that had to be considered before one was found that the model (erroneously) believed more appropriate for user u.

4. Implementation

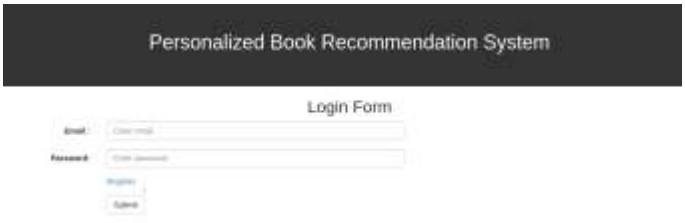

Figure2:Login

In fig.2 The registered user can login into the system by using their e-mail-id and password. If the user is not a registered user then he/she has to register into the system by the filling the registration form.

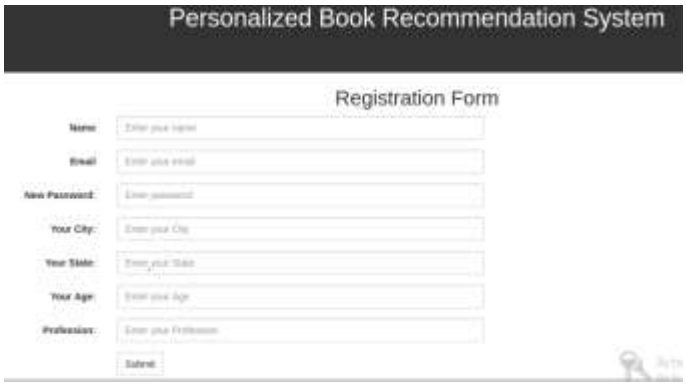

Figure 3:Registration Page

After providing the user-id the system will recommend the books to the user based on four parameters. They are based on past-likes, based on Age, based on location and based on the user's favorite Author.

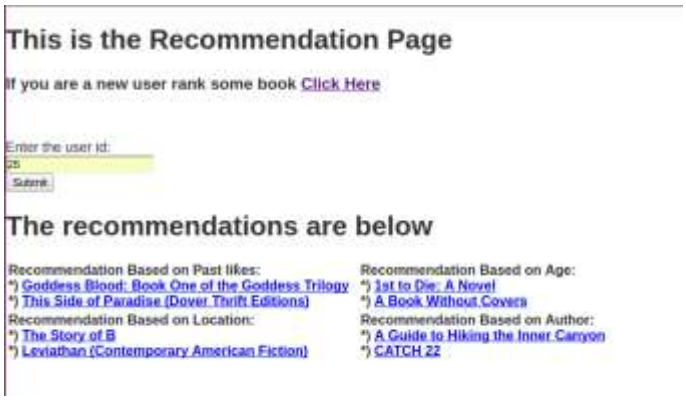

Figure 4: Recommendation List

5. Conclusion

This system aims to provide personalized recommendation of books to the user. This system considers big data of books. The system makes use of both content-based and collaborative filtering algorithm in order reduce the cold start problem and provides the user with recommendation list. The system tries to predict the ranking by considering the item's similarity as well as user's similarity so that a user can get recommendations of new books.

6. Future Work

This system can be extended to study the accuracy for the recommendation list. The given recommendation list can be enhanced by analyzing the user reviews and working on the security issues. The recommendations can be provided to the user based on users's profession.

References

[1] S. Vinodhini, V. Rajalakshmi, B. Govindarajalu," Building Personalised Recommendation System With Big Data and Hadoop Mapreduce ", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181.

[2] Pijitra Jomsri," Book Recommendation System for Digital Library Based on User Profiles by Using Association Rule", IEEE,pp. 978-1-4799-4233-6/14,2014.
[3] Salil Kanetkar, Akshay Nayak, Sridhar Swamy ,Gresha Bhatia," Web-based Personalized Hybrid Book Recommendation System ",IEEE, pp. 978-1-4799-6393-5/14,2014.
[4] Madden, S.: From Databases to big data. In: IEEE Internet Computing, vol. 16, no. 3, pp. 4-6. IEEE Computer Society (2012).
[5]. C.L. Philip Chen, and Chun-Yang Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Informatics and Computer Science Intelligent Systems Applications, Volume 275, 2014.
[6] Punit Gupta,Ravi Shankar,"Tagging based evolving Recommendation System for digital library system",2015 4th International Symposium on emerging Trends and Technologies in libraries and information services.
[7] Guangqian Zhangl, Wei Sunl, "User Preferences to Attributes of Books for Personalized Recommendation", IEEE pp 978-1-4673-2008-5/12,2012.
[8] CaiNicolas Ziegle.R, Sean M. McNee., Joseph A. Konstan and Georg Lausen (2005), „Improving Recommendation Lists Through Topic Diversification", International World Wide Web Conference Committee (IW3C2), ACM, pp. 5959-30-469.
[9] Fatos Xhafa, and Leonard Barolli, "Semantics, intelligent processing and services for big data", The International Journal of Grid Computing and eScience, Volume 37, 2014, pp. 201-202.
[10] Boban Vesin., Mirjana Ivanovic., Aleksandra Klasnja-Milic and Zoran Budimac (2012), „Ontology-based semantic recommendation in programming tutoring system", Journal on expert systems with applications, Vol. 39, pp 1229-12246.
[11] April Reeve, "Big Data Integration Data Integration, Best Practice Techniques and Technologies", Morgan Kaufmann, 2013, pp. 141–156.
[12] Saranya.K.G, G.Sudha Sadhasivam," A Personalized Online News Recommendation System ", International Journal of Computer Applications (0975 – 8887) Volume 57– No.18, November 2012.
[13] Xi Chen and Liqing Zhou," Design and Implementation of an Intelligent System for Tourist Routes Recommendation Based on Hadoop ",IEEE, pp. 978 -1-4799-8353-7 /15 , 2015.
[14] Brian McFee, Luke Barrington and Gert Lanckriet, "Learning Content Similarity for Music Recommendation" IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, No. 8, 2012.
[15] Berger, A et al (2000). Bridging the Lexical Chasm: Statistical Approaches to Answer Finding. In Proc. Int. Conf. Research and Development in Information Retrieval, 192-199.
[16] Metadata Embeddings for User and Item Cold-start Recommendations, Maciej Kula, arXiv:1507.08439v1 [cs.IR] 30 Jul 2015.
[17] https://github.com/lyst/lightfm
[18] Message Clustering based Matrix Factorization Model for Retweeting Behavior Prediction, Bo Jiang, Jiguang Liang, Ying Sha, Lihong Wang, ACM 978-1-4503-3794-6/15/10 ...$15.00.
[19] http://www2.informatik.uni-freiburg.de/~cziegler/BX/
[20] Survey on Recommendation System, Lipi Shah, Hetal Gaudani, Prem Balani, International Journal of Computer Applications (0975 – 8887) Volume 137 – No.7, March 2016
[21] Recommendation Systems:Issues and challenges, Soanpet .Sree Lakshmi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5771-5772
[22] Survey Paper on Recommendation System, Mukta kohar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3460-3462
[23] Cover, T.,and Hart, P., Nearest neighbor pattern classification. Information Theory, IEEE Transactions on, 13(1):21–27, 1967

Author Profile

Jayanti Rathnavel has received the B.E. degree in Computer Engineering from Yadavrao Tasgaonkar Institute of Technology in 2014. And currently pursuing M.Tech from K J Somaiya College of Engineering.