International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 08 | Aug 2018 www.irjet.net p-ISSN: 2395-0072

# Web Based Hybrid Book Recommender System Using Genetic Algorithm

**Parth Parekh[1], Ishaan Mishra[2], Abhishek Alva[3], Vikas Singh[3]**

*[1]UG Scholar, Computer Science & Engineering Department, Shah and Anchor Kutchhi Engineering College, Mumbai, Maharashtra, India.*
*[2] UG Scholar, Computer Science & Engineering Department, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India.*
*[3] Undergraduate, Computer Science & Engineering Department, Srinivas Institute of technology, Valachil, Mangalore, Bangalore, India.*
*[3] Undergraduate, Electronics Engineering, Ramrao Adik institute of technology, Navi Mumbai. Maharashtra, India.*

---***---

**Abstract-**Recommender Systems have been for more than a decade now. Selecting what book to read after completing the present one has always been a question for everyone. Even for school pupil, deciding which workbook/textbook or reference book to read on a topic is a big question. In this paper, we try to represent a model for a website-based personalized hybrid book recommender system which utilize varied aspects of sending recommendations apart from the regular communal and content-based filtering approaches. A hybrid model is proposed to integrate outputs produced by every recommender at the basis of Genetic Algorithm. The function is designed based on forecasting accuracy, and a selecting strategy is constructed to avoid pre-mature of Genetic Algorithm. Experiments express the proposed model has higher capability than traditional method. 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. Scraping information from the web and using the information obtained from this process can be equally useful in making recommendations.

***Key Words***: Hybrid Recommender, Recommender System, Book recommender, Genetic Algorithm, Web Based Recommender System.

## 1.INTRODUCTION

Through estimating the requirement of customer, proves the suitable product and services for individual, Personalized Recommender System aims to solving the customer's overloading of information, and improving the performance of e-Commerce system. With the development of Electronic Commerce, Personalized Recommender System becomes an important research item [1]. At present, many e-Commerce system employ Personalized Recommender System in different degree, such as eBay, Amazon, DangDang book store, and so on [2]. The recommendation system not only help customer, but also enhance the customer's satisfaction to the commercial activity. In general, the recommendation system has following effect in electronic commerce activity: (1) Help user retrieval useful information, (2) Promotion sales, (3) Personalized service, (4) Enhances the customer loyalty. In general, recommender system is constructed on rating data matrix. At the basis of analyzing products rated by target customer, forecasting the value of unrated items, then items with higher forecasting value are selected and recommended to customer [3]. The rating data can be collected by directed or concealed way, such as analyzing customer's clickstream and behavior [4].
Recommendation system [11] is the type of information filtering which involves prediction of rating and user preferences, which would help user to buy items according to their needs and interest. The books suggestion at Amazon.com is best example of recommendation system .Recommendation System directs [11] the way to find products, Information according to their interest. Recommendation system uses following technologies to recommend products: Content filtering, Collaborative filtering & Association mining. Content filtering recommends item which is based on Users profile, which he has liked in past.
Currently, more and more AI technology has been introduced in e-Commerce to improve the capability of recommendation system. To gain exact and real-time recommendation, some recommending methods have been constructed based an different theory [5], such as Collaborative Filtering algorithm, Bayesian Network, Association Rule Mining, Clustering, Horting Graph, Knowledge-based recommender, etc. According to the basic thought, recommender system can be classified into three kinds: Content-based System, Collaborative System and Hybrid System [6]. Content-based system generates recommendations through analyzing the features and similarities among products, it mainly applied the column data of Table 1. But, there are some limitations in generalization and recommending to new user.

Collaborative system generates recommendations at the basis of similarity among users. Collaborative Filtering (CF) algorithm is a successful method, widely applied in many eCommerce systems, such as recommending movies or news for user. CF algorithm evaluates current customer's near neighbors according to the product rating data of customer. Through neighbors' rating data, the current customer's evaluation for a new product can be forecasted, then, the recommendation for current customer can be gain. This kind of recommender also has defaults, such as data sparsity, new user and new product problem.

## 2. Proposed Work

### 2.1 HYBRID RECOMMENDER MODEL BASED ON GA

Suppose C is set of customers, P is set of products, R is set of rating. Rating values describe attitudes or preferences of customers about products. In fact, different recommender method constructs different forecasting function E, to predict those rating values of un-rating items. It is,

$$E : C \times P \dashrightarrow R \quad (1)$$

Then, the product with the highest predicting rating value is recommended to customer c.

A. Description of Hybrid recommender:-
For K recommenders, the forecasting result of recommender k is vector $EK=\{e_{1k}, e_{2k}, \ldots, e_{Nk}\}$, $e_{ik}$ is attitude of current customer about product i. N is the number of un-rating items. In traditional model, because there is one recommender, the recommending result can be obtained. But in hybrid model, different recommender has different result, the integrating process is necessary to gain finial result. Various recommenders have different capability because of their different principle. For current customer, the disparity embodies in forecasting feature for different product. On the other hand, for a product, the forecasting results are also different according to different customers. Because recommender model services to customer, it is not the sale system. So, supposing $wk=\{w_{1k}, w_{2k}, \ldots, w_{Nk}\}$ is weight vector of recommender k, wik express its forecasting ability about product i. If the value of wik is higher, the recommender k more specialize in forecasting product i. The following condition must be satisfied.

$W=\{w_1, w_2, \ldots, w_K\}$ is combination of every weight vector to form a weight matrix. In order to gain the finial result oi of customer x about product i, the following formula is used to synthesize each recommender.

$$o_i (W,x)={}^{k}\textstyle\sum_{k=1} W_{ik}\, e_{ik}$$

Then, hybrid recommender problem can be described by a 2-level network,. In the input level, there are K × N nodes, including result vector obtained from recommenders. Output level has N nodes, respectively represent N kinds of products, to describe integrating forecasting results. So, when recommenders have stable ability to forecast rating values, connective weights become the key affecting the final result. From this thought, hybrid recommender problem is transferred to the optimizing problem of weight vectors.

B. Learning process of weights:- In GA, problem is coded as a string, called chromosome, each solution candidate is called individual, the set of individual is called population. The population evaluates, and a selecting strategy is applied to choose better solution [8].
In our problem, the weight matrix
$W=\{w_{11}, w_{21}, \ldots, w_{N1}, w_{12}, w_{22}, \ldots, w_{N2}, \ldots, w_{1K}, w_{2K}, \ldots, w_{NK}\}$
becomes the optimizing target. Its initial solution could be some random values, or some temporary values gained by an off-line process. At the basis of initial population, new individuals produced in each iteration are evaluated by fitness function, a new population is gained through crossover and mutation operation in individuals with higher fitness. Fitness is an evaluated function to analyze feature of individual and judge its forecasting capability. When individual W is applied to combine every recommender's result, the similarity between finial forecasting result with actual rating value can express it's forecasting accuracy. It's the basis of fitness. So, the accuracy function is defined as following:

C. Individual selecting strategy:- At the beginning, difference among individuals is huge, simple random method can be used. So, individual with larger fitness has more offspring. In the later of GA running, fitness of different individual tend to consistence, the evolution of population nearly stop. Then, a selecting strategy is needed to avoid population's pre-mature. Simulated Annealing algorithm is an often-used optimizing method to search better solution. Through adjusting control parameters, the pre-mature problem can be solved.
Individual selecting strategy is constructed as following .
In this, Generation is set of individual, namely population, T is temperature, T0 is the larger initial temperature, t is the number of generation. Along the increasing of generation, T decrease step by step, the fitness of individual W is enlarged
and difference among individuals is expanded. So, the advantage of better individual become more outstanding and has more chance to be selected. The pre-mature problem can be avoided.

### 2.2. EXPERIMENTS AND ITS ANALYSIS

The goal of experiments is improving recommendation. So, the validating basis of model is the quality of recommendation. Use data set provided in the MovieLens (http://movielens.umn.edu/) to design the experiment. From rating data database, randomly select 6,000 terms as the experimental data set, contains 118 users and 220 movies. 80% of data is used as training set, and others as test set. Mean Absolute Error (MAE) is applied to evaluate recommendation quality. Let forecasted rating set is {p1, p2, … , pn}, the real rating set is {q1, q2, … , qn}, then,

In experiments, three benchmarks are applied, including, Recommender 1: content-based recommender with related similarity analysis. Recommender 2: collaborative filtering recommender with related similarity analysis. Recommender 3: collaborative filtering recommender with cosine similarity analysis. And, two hybrid methods are used, including the popular hybrid model based on voting theory and proposed hybrid model. Number of individual is 50, crossover rate is 0.45, mutation rate is 0.001, and initial temperature T0 is 100.

### 2.3 Building trust in Hybrid Recommender System

It is of utmost importance that the user should trust the recommendation system. Having confidence in the system would help the user get a healthy recommendation experience. The developers should prioritize this task. One way ensuring trust is explaining the process of recommendation to the user. On the website, it is better to have a webpage explaining how recommendations are made. A video of the same would have a larger impact. This would not only help the user understand the entire mechanism, but also develop a sense of trust with the system. Another way of building trust is via Mirror Behavior. To mirror someone is to have a profile similar to another user. In case of a profile match, we can allow the user to view the profile of another user with similar traits. However, care should be taken not to intrude into the other users' privacy. Recommender systems need to have a proper balance in this case. Mirror Behavior feature should be permitted only in the case that the user allows his profile to be shared with others, else not. Security, is another important aspect of every recommender. A web based recommendation system is subject to attack by malevolent users who try to influence the behavior of the system by inserting fake user ratings [9]. Creating fake accounts and profiles is the simplest strategy adopted by these users. They could have two different goals-increase or decrease the rating of an item. Such attacks are called push and nuke attacks respectively [10]. A book recommender system is often subjected to two specialized from of attacks-bandwagon and segment attack. These are similar to the push and nuke attack and mainly attack the blockbuster items. Prevention of such attacks should be necessary, as the recommendations given to the user should not be influenced by the injection of dubious ratings.

One technique is to use hybrid algorithms for giving recommendations. A system solely based on collaborative filtering technique is more susceptible to the kind of attacks mentioned above. Hence, using a hybrid system or multiple layers of filtering techniques can help reduce the effect of these attacks. Our system which mainly deals with three filtering techniques-Collaborative, Content and Demographic would be better resistant to the attacks. Nuke and push attacks are based on profile injection methods. A method to make these attacks more difficult is to increase the profile injection cost. Use of captchas makes this possible [13]. Hence during every session, captchas could be used before making recommendations. IP Address could be yet another tool for preventing attacks of malevolent users. If too many registrations are happening from the same IP address, then any activity from that address can be simply blocked to prevent any further attacks.

## 3.Conclusion

Apart from just the traditional Collaborative and Contentbased filtering techniques, many modern techniques are being exploited nowadays. The hybrid algorithms are a mixture of many techniques. A hybrid model is proposed to integrate outputs produced by every recommender at the basis of Genetic Algorithm. Demographic filtering helps give more personalized recommendations. With the advancement of the web, its use in the process of recommendation can help improve the efficiency. In this paper, we have explored the use of Web Scraping, which is a form of web content mining. The inculcation of Web into the process of recommendation, can help solve many limitations related to the filtering methods. Through constructing weight vectors that represent different forecasting performances of each recommender, the hybrid problem is translated into optimizing problem about weight vectors. The booming growth of Ontological aspects, semantic and context aware recommendations is sure to improve the quality of recommendations made. Recommendations in ubiquitous domains like that of mobile phones is enhancing the applicability of filtering. A combination of the models presented in the paper and many more can lead to a much greater personalized experience for the users and will also enhance the accuracy of the recommender systems over time. Experiments express the proposed model has higher capability than traditional method.

## 4.REFERENCES

[1] Introduction to Recommender Systems Handbook, Francesco Ricci, Lior Rokach, Bracha Shapira, F. Ricci et al. (eds.), Recommender Systems Handbook.

[2] Chhavi Rana, Sanjay Kumar Jain, "Building a Book Recommender system using time based content filtering", WSEAS Transactions on Computers, E-ISSN: 2224-2872, Issue 2, Volume 11, February 2012. [3] Eloisa Vargiu, Mirko Urru, "Exploiting web scraping in a collaborative filtering based approach to web advertising", Artificial Intelligence Research, 2013, Vol. 2, No. 1.

[4] Laila Safoury, Akram SalahExploiting, "User Demographic Attributes for Solving Cold-Start Problem in

Recommender System", LNSE.2013.V1.66, Lecture Notes on Software Engineering, Vol. 1, No. 3, August 2013.

[5] Sean Owen, Robin Anil, Ted Dunning, Ellen Friedman, "Mahout in Action", 2012, Manning Publications, ISBN 9781935182689.

[6] Miguel Ram´ırez J´avega, "Content-based Music Recommender System", 29th June 2005.

[7] Magnus Mortensen, "Design and Evaluation of a Recommender System", INF-3981 Master's Thesis in Computer Science, February 5, 2007.

[8] Parth Parekh, Pournima Awate, Kajal Parmar "Spam URL Detection and Image Spam Filtering using Machine Learning" ,Computer Engineering, 2018 https://www.irjet.net/archives/V5/i7/IRJET-V5I7420.pdf

[9] Mehmet Kayaalp, Tansel Özyer, Sibel Tariyan Özyer, "A Collaborative and Content Based Event Recommendation System Integrated with Data Collection Scrapers and Services at a Social Networking Site", 2009 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2009, 20-22 July 2009, Athens, Greece.

[10] Andrei Toma, Radu Constantinescu, Floarea Nastase, "Recommendation system based on the clustering of frequent sets", WSEAS Transactions on Information Science and Applications, ISSN: 1790-0832, Issue 5, Volume 6, May 2009.

[11] Neil Hurley, "Tutorial on Robustness of Recommender Systems", RecSys 2011: Tutorial on Recommender Robustness, October 2011.

[12] Judy C.R. Tseng, Gwo-Jen Hwang. "Development of an Automatic Customer Service System on the Interne," Electronic Commerce Research and Application, 2007, (6): 19—28.

[13] B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan and J. Riedl. "MovieLens Unplugged: Experience with an Occasionally Connected Recommender System," Proc. Int'l Conf. Intelligent User Interfaces, 2003.

[14] Robin Burke. "Hybrid recommender systems: survey and experiments," User Modeling and User-Adapted Interaction, 2002, (12): 331-370

[15] Linqi Gao. "Production Recommender Method Based on Customer's Behavior," Computer Engineering and Application, 2005, (3): 188-190.

[16] S. E. Middleton, N.R. Shadbolt and D.C. de Roure. "Ontological User Profiling in Recommender Systems," ACM Trans. Information Systems, 2004, 22(1): 54-88.

[17] Gediminas Adomavicius, Alexander Tuzhilin. "Toward the Next Generation of Recommender Syetem,". IEEE Transaction on Knowledge and Data Engineering, 2005, 17(6): 734-749.

[18] Robin Burke. "Hybrid recommender systems: survey and experiments,". User Modeling and User-Adapted Interaction, 2002, (12):331-370.

[19] Liu Juan, Hitoshi Iba. "Selecting informative genes with parallel GA in tissue classification," Genome Informatics, 2001, 3(12): 14-23.