Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA-2020)
IEEE Xplore Part Number: CFP20J88-ART; ISBN: 978-1-7281-6387-1

User Location and Collaborative based Recommender System using Naive Bayes Classifier and UIR Matrix

R. Suguna
Assistant Professor
Department of Computer Science and Engineering
Bannari Amman Institute of Technology
Sathyamangalam - 638401
Tamilnadu, India
sugunar@bitsathy.ac.in

P. Sathishkumar
Assistant Professor
Department of Computer Science and Engineering
Bannari Amman Institute of Technology
Sathyamangalam - 638401
Tamilnadu, India
sathishkumarp@bitsathy.ac.in

S. Deepa
Assistant Professor
Department of Computer Technology
Kongu Engineering College
Perundurai
Tamilnadu, India
sdeepakec@gmail.com

Abstract - **The world is filled with information and getting the right information is a challenging task for internet users and online buyers. Recommender system helps internet users to get their information in a short span of time. It acts as an information extraction system that works behind users to perform their search easier. The recommender system comes under user's content or item based search, similar users browsing behavior called collaborative and combination of both known as a hybrid. Here collaborative-based approach is adopted which recommends items to their users based on their past browsing behavior. In this article, the User-Item-Rating matrix is formulated concerning user personal profile, rating of the product, and reviews given by the users during their previous browsing history. In this research, user location is considered as an important attribute to group similar users. It also attempts to suppress the scalability and sparsity problems of the traditional collaborative filtering approach. So, the User-Item-Rating (UIR) matrix has considered the location, ratings and reviews for future recommendation. The Navie Bayes classifier algorithm is used to provide accurate topmost recommendations to internet users. The data set is taken from the MovieLens and IMDb database. The accuracy of the recommender system is measured based on the main metric f-measure. The experimental result has proven the improvement of the recommender system with the mentioned added attributes.**

Keywords - Recommender System, Collaborative Filtering, User-Item-Rating Matrix, Naive Bayes Classifier.

I. INTRODUCTION

The world is surrounded by information that consumes most of the internet user's time to provide the expected and desired information to them. Information overload and its related issues trigger the researchers to find the absolute solution to obtain the desired information by spending a short span of time. The enormous growth of the internet and e-commerce applications has stepped into the introduction of recommender systems. The ultimate intention of developing a recommender system is to solve the trouble of information excess and also reduce information searching time [1][2]. Recommender systems help their customers to search for their products more easily by proving a set of recommendations.

The researchers have proven the success of the recommender system by following the collaborative filtering approach which works on the interest of a similar set of users as well as products. Most of the research has proven the user-based collaborative filtering approach given a better accuracy rate for recommending the product [1]. User interest based recommender approach is the widely successful approach to construct recommendation systems which are exclusively utilized by most of the systems which are used for recommending the products.

Recommender systems are the real solution to information overload and information retrieval. Many of the e-commerce applications namely Amazon, Flipkart and Shop clues are using their own recommender systems to address the issues related to information search and retrieval. The recommender systems help the internet users to find their expected product easier and quickly by providing a set of recommendations based on their past browsing behavior [3].

The recommender systems come under one of the following three techniques namely (i) collaborative filtering approach, (ii) content-based approach, and (iii) hybrid recommendation approach [6]. The customer's browsing behaviors, rating, comments, feedback, and their personal information and preferences are the main sources for predicting the current product recommendation while using the content-based approach.

978-1-7281-6387-1/20/$31.00 ©2020 IEEE

Collaborative filtering approach recommends the products to its customers by considering their preferences and also other customer preferences who have similar interest based on their browsing pattern, rating and reviews [6]. This type of recommendation system is working based on the analogy and resemblance among the web surfers and their preferences and browsing behavior. A recommendation approach falls under content is providing recommendations to the users only by their interest in the product based on their browsing history, rating, comments and feedback. This type of recommendation system is watching the users every clicks and movement, recognize their interest and recommending the products to them. The hybrid recommendation approach works on the combination of the above two approaches [10].

The researchers have proven the success of the collaborative filtering approach but it has faced issues while it is dealing a huge number of users and items or products to predict user preferences [5]. This paper follows the User – Item - Rating (UIR) matrix to filter the number of similar users and products for recommending the product to the current active customers. Fig 1 describes the problem formulation of the proposed methodology.

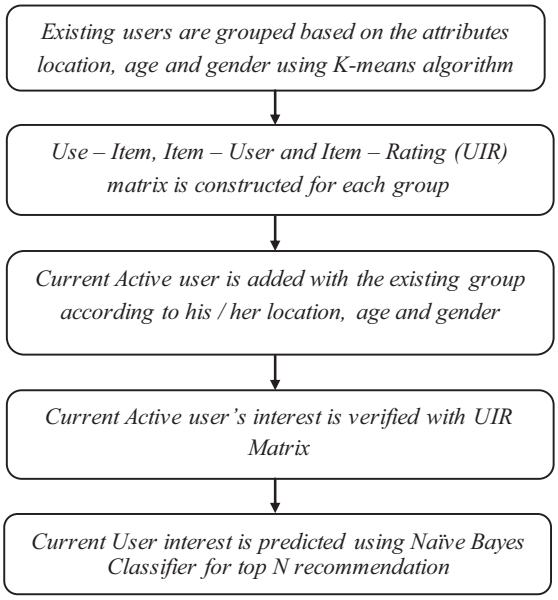

Fig 1. Proposed Methodology

The layout of the article is organized as follows. The existing research of recommendation is discussed in Chapter 2. The adopted new techniques and their features are elaborated in Chapter 3. Chapter 4 has covered the experimental result and its discussion. The conclusion and further research direction are detailed in Chapter 5.

2. RELATIVE WORK

[1] The authors have done an extensive study on the recommendation and its research issues. They found that the accuracy of the recommendation system is mainly affected by the cold start problem. The authors suggested incorporating a hybrid recommendation approach for future research. [2] The authors attempted to optimize the value of k in the clustering algorithm to group similar users together. [2] Association rule mining was implemented with a collaborative filtering approach to improve the accuracy of the recommendation. They compared the findings with seven different algorithms. There is no significant improvement in the result when k is optimized. [6] Association rule mining was incorporated with particle swarm optimization techniques to enhance the accuracy of the author's recommendation method called GRAM. [3]The item rating has played an important role in any recommender system. The users were grouped by referring to the rating given by them. [4][5] Deep learning algorithms were well utilized in healthcare recommendation systems which significantly reduce the error rate of the recommender system. [7][9] A collaborative filtering approach was combined with sentiment analysis algorithms to provide electronic item recommendations to the users. The authors have addressed the scalability and sparsity issues of the traditional collaborative approach. In this research item rating and review comments were taken into consideration. [8] In this research, Naïve Bayes classifier is used for item prediction in association with collaborative filtering approach. [10][11] Researchers attempted to recommend the items to the internet users based on the content which they browse. [12] A personalized food recommendation technique was devised by using neural network techniques for thyroid patients. A food preference matrix was formulated for a recommendation. [13][14][15] Hybrid recommendation systems were widely proposed by the researchers by combining the attributes of a collaborative and content-based approach for item recommendation.

Researchers attempted to improve the accuracy of the recommendation of items by applying the learning abilities of the machine and deep approaches. Also, researchers took an effort to reduce the scalability and sparsity issues of the collaborative filtering approach. In this current era, a fine recommender system is necessary to reduce internet surfing time. This motivates to propose a new methodology in this field of research.

3. PROPOSED METHODOLOGY

This research aims to ameliorate the exactness of the recommendation by reducing the scalability issues of the conventional collaborative filtering approach by grouping the users based on their information provided in the personal profile namely age, gender and location. In this research, location plays a major role while recommending the products to the currently active users. The users in a particular location especially have similar tastes and interests. The algorithm for the proposed methodology is given below

Input : *Users, Items and Ratings*
Output : *Top N Items*
Algorithm

Step 1 : *Construct User Cluster C (Age, Gender and Location)*
Step 2 : *For each C*
form UI, IU, IR Matrix
Step 3 : *Check Current active profile with C.*
Step 4: *Add the current user with suitable cluster C*
Step 5: *Extract frequent items from UI*
Step 6 :*Compare ratings and review commends with IR matrix*
Step 7 *: Apply Naïve Bayes to recommend N items to the current active user*

A. User Grouping

Here, the K-means algorithm is owned to assemble users for their age, gender and location. The clusters are formed by calculating the average of the squared distances from the cluster centers of the respective clusters. Consider x and y are two data points then sim*(x,y)* is calculated using Euclidean distance to the given data set.

$$sim\ (X,Y) = \sqrt{\sum_{i=1}^{n}(xi - yi)(xi - yi)} \qquad (1)$$

B. User – Item, Item - User and Item-Rating Matrix

In this stage, each user group is taken into account for the construction of User-Item matrix (Table I), Item-User matrix (Table II) and Item-Rating matrix (Table III). Consider

$$Cluster = (c1, c2, c3, \ldots cp) \qquad (2)$$

$$User = (u1, u2, u3 \ldots un) \qquad (3)$$

$$Item = (i1, i2, i3, \ldots im) \qquad (4)$$

$$Rating = (r1, r2, r3, \ldots, rk) \qquad (5)$$

Where C – User clusters, U – User group, I – Item set, R – Item rating. For each cluster namely *C1, C2, C3…Cn*, User-Item matrix (*UI),* Item-User matrix (*IU)* and Item-Rating matrix (*IR)* is formulated.

In Item-Rating matrix, the rating and review comments are converted into the range 0 to 5. Finally, the mean rating of each item is computed and maintained in the resultant matrix.

TABLE I. USER – ITEM (UI) MATRIX

User	Item			
U_1	I_1	I_2	I_3	I_m
U_2	I_2	I_4	I_1	I_m
U_3	I_3	I_2	I_1	I_m
U_n	I_3	I_2	I_1	I_m

TABLE II. ITEM - USER (UI) MATRIX

Item	User			
I_1	U_1	U_2	U_3	U_n
I_2	U_2	U_4	U_1	U_n
I_3	U_3	U_2	U_1	U_n
I_n	U_3	U_2	U_1	U_n

TABLE III. ITEM - USER (UI) MATRIX

Item	U_1	U_2	U_3	U_n	Avg
I_1	4	4	5	3	4
I_2	5	4	3	2	3.5
I_3	3	3	4	2	4.3
I_n	5	3	4	5	4.5

C. Item Recommendation using Naïve Bayes Classification Algorithm

A supervised algorithm namely Naive Bayes classifier is a probabilistic classification technique. Let us consider that *P(C/i)* states that that item *i* belongs to class C. The classifier works as follows: *X = {x1, x2, . . . , xn}* where x is independent items and *C = {c1, c2, . . . , cm}* where c is class. This research considers item rating as the major attribute. The current active user profile is matched with the user group and the items are recommended to the active user by selecting the items which have top ratings from the Item-Rating matrix.

4. EXPERIMENTAL RESULT

The proposed system uses the movie data set taken from the imdb and MovieLens. MovieLens data set is public in nature which has 18 categories in rating. The relevant attributes used-id, movie-id and ratings are extracted from the dataset. Totally 7120 users were rated 130642 movies. The movie rating is range from 1 to 5. The imdb dataset is in tsv formatted file which contains title, average rating and vote details. It is proposed to apply two different datasets to assess and prove the accuracy of the recommender system. The proposed methodology is implemented using python programming. The accomplishment of the research approach is evaluated based on the performance metrics recall, precision and f-measure. Table IV represents the data set obtained from two databases.

TABLE IV. DATA SET DESCRIPTION

Database	*No. of Users*	*No. of Movies*	*Testing Data in %*	*Training Data in %*
MovieLens	5000	1,25,345	30	70
Imdb	5000	1,67,348	30	70

A. Evaluation Metrics

The resulting stage of the current research is showcased and it is evaluated for the metrics precision, recall and f-measure.

$$Precision = \frac{Number\ of\ Corretcly\ Recommended\ Items}{Total\ number\ of\ Recommended\ Items} \qquad (6)$$

$$Recall = \frac{Number\ of\ Corretcly\ Recommended\ Items}{Total\ number\ of\ Items\ liked\ by\ the\ User} \qquad (7)$$

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (8)$$

B. *Performance of Proposed Algorithm*

The advancement of the current recommendation system is compared with the existing technique [6]. Table V shows the result of the current research methodology.

TABLE V. EVALUATION RESULT OF IMDB AND MOVIELENS DATA SET

Methodology	*Imdb*			*MovieLens*		
	Precision	***Recall***	***F-measure***	***Precision***	***Recall***	***F-measure***
Existing	0.602	0.799	0.687	0.602	0.799	0.687
Proposed	0.852	0.902	0.876	0.863	0.899	0.881

Fig 2 depicts the enhanced result of the current research work when compared to the existing method for the data set obtained from the imdb data set. The performance metrics precision, recall and f-measure values are significantly improved in the proposed method which produces the recommendation accuracy of 87%.

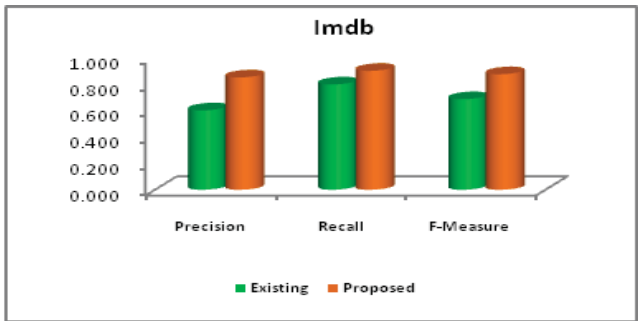

Fig 2 Performance Comparision using Imdb Dataset

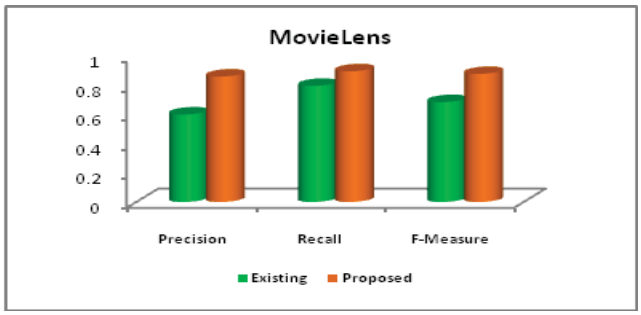

Fig 3 Performance comparison using MovieLens Dataset

Fig 3 depicts the accuracy of the proposed recommendation system which is 87% in the MovieLens database. The f-score of the proposed research clearly indicating the performance improvement when compared with the existing methodology.

5. CONCLUSION AND FUTURE SCOPE

The proposed system is to understand the behavior of internet users is a challenging task. In this research, a contemporary recommender system using the UIR matrix and Naïve Bayes classifier algorithm is used to raise the perfectness of the recommendation. This research focuses on user location to identify the similarity between them. The location of the users is considered as a major parameter to form the initial user clusters. The proposed system is evaluated with the metrics named recall, precision and f-measure. The experimentation of the proposed algorithm is carried out by considering two different data sets. The proposed algorithm gives better performance to recommend the top items to the current live users when compared with the existing methodology. The accuracy of the recommendation system is extended to 87% in Imdb and 88% in the MovieLens data set. This research used the unsupervised learning algorithm to find out the similarity of users. In the future, personalized item-based recommendation by using deep learning techniques may be proposed to improve the recommendation accuracy.

REFERENCES

[1] Jyoti Shokeen and Chhavi Rana, "A Study on Features of Social Recommender Systems", Artificial Intelligence Review, Vol. 53, pp. 965–988, January 2019.

[2] Debby Cintia Ganesha Putri, Jenq-Shiou Leu and Pavel Seda, "Design of an Unsupervised Machine Learning based Movie Recommender System", Symmetry, Vol. 12, 185, 2020.

[3] Minjae Kim, SungHwan Jeon and Heeseong Shin, "Movie Recommendation based on User Similarity of Consumption Pattern Change", IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering, pp. 317-319, 2019.

[4] Abhaya Kumar Sahoo, Chittaranjan Pradhan, Rabindra Kumar Barik and Harishchandra Dubey, "DeepReco: Deep Learning based Health Recommender System using Collaborative Filtering", Computation, Vol. 7(2), pp. 1-18, 2019.

[5] Xiaoyi Deng and Feifei Huangfu, "Collaborative Variation Deep Learning for Healthcare Recommendation", IEEE Access, Vol. 7, pp. 55679 - 55688, 2019.

[6] Behzad Soleimani Neysiani, Nasim Soltani, Reza Mofidi and Mohammad H. Nadimi-Shahraki, "Improve Performance of Association Rule based Collaborative Filtering Recommendation Systems using Genetic Algorithm", International Journal of Information Technology and Computer Science, Vol. 2, pp.48-55, 2019.

[7] N. A. Osman, S. A. M. Noah, and M. Darwich, "Contextual Sentiment based Recommender System to provide Recommendation in the Electronic Products Domain", International Journal of Machine Learning and Computing, Vol. 9 (4), pp. 425-431, August 2019.

[8] Priscila Valdiviezo Diaz, Fernando Ortega, Eduardo Cobos, Raúl Lara Cabrera A, "Collaborative Filtering Approach based on Naïve Bayes Classifier", IEEE Access, Vol. 7, pp. 108581 – 108592, 2019.

[9] Sudhanshu Kumar, Kanjar De and Partha Pratim Roy, "Movie Recommendation System using Sentiment Analysis from

Microblogging Data", IEEE Transactions on Computational Social Systems, Vol. 7(4), pp. 915-923, August 2020.

[10] Ibukun Tolulope Afolabi, Opeyemi Samuel Makinde and Olufunke Oyejoke Oladipupo, "Semantic Web mining for Content based Online Shopping Recommender Systems", International Journal of Intelligent Information Technologies, Vol. 15(4), 2019.

[11] SRS Reddy, Sravani Nalluri, Subramanyam Kunisetti, S. Ashok and B. Venkatesh, "Content based Movie Recommendation System using Genre Correlation", Smart Intelligent Computing and Applications, Vol. 109, pp 391-397, 2020.

[12] Vaishali S. Vairale and Samiksha Shukla, "Recommendation of Food Items for Thyroid Patients using Content based KNN Method", Data Science and Security, pp 71-77, 2020.

[13] Jason Smith, Dillon Weeks and Mikhail Jacob, "Towards a Hybrid Recommendation System for a Sound Library", IUI Workshops, March 20, 2019, Los Angeles, USA.

[14] Xiaomei Bai, Mengyang Wang, Ivan Lee, Zhuo Yang, Xiangjie Kong and Feng Xia, "Scientific Paper Recommendation: A Survey", IEEE Access, Vol. 7, pp. 9324-9339, 2019.

[15] Tianyu Zuo, Shenxin Zhu and Jian Lu, "A Hybrid Recommender System Combing Singular Value Decomposition and Linear Mixed Model", Intelligent Computing, Vol. 1228, pp 347-362, 2019

978-1-7281-6387-1/20/$31.00 ©2020 IEEE