International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 03, 2021

A Hybrid Recommender System Using Bio Inspired Clustering Technique For Top N Item Recommendation

R Suguna[1], P Sathishkumar[2], and S Deepa[3]
[1,2]*Assistant Professor, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India*
[3] *Assistant Professor, Computer Technology, Kongu Engineering College Perundurai, Tamilnadu, India*

Email: [1]sugunar@bitsathy.ac.in, [2]sathishkumarp@bitsathy.ac.in, [3]sdeepakec@gmail.com

Abstract. Now, we all are living in the information rich world, people depend on technologies for their everyday activities. The technologies and its massive growth give rich set of benefits and services to the users. The entire world is running towards digitization. The people, company, organizations and institutions are making their everyday activities in digital form. The growth of digitization gives better solution for information search. But, on the other end, people struggle with information overloaded. We are finding difficult to get the desired information from the internet. Sometimes it is a great challenge to get the wished data from the internet. This is the reason behind the research in recommendation and invention of recommender systems. A smart recommender system provides better solution for information search and its related issues. Many recommender systems and tools are widely used by the e-commerce applications, but the accuracy of the tools still needs improvement. The previous research in the area of recommendation comes under collaborative filtering, user content search and by combining the best of the above mentioned. The proposed work focuses on the combination of the first two methods which are mentioned earlier. This research attempts to utilize the possible features of the two clustering techniques to achieve the user interest prediction. The item ratings are the major parameters taken for evaluating the efficiency of the proposed work. The data set are taken from Imdb database (www.imdb.com) and MovieLens dataset to test the accuracy of the proposed system.

Keywords: Recommender System, Hybrid Approach, K-Means, Particle Swarm Optimization.

1. INTRODUCTION

In the current era, everybody is using the World Wide Web (WWW) and its applications for variety of purposeS. E-commerce and its related applications provide suitable way out for the users to economize their time and make their task in easier way. Organizations take effort to make their business transactions online and provide the at most facility to their users. As a result, data are overburden and makes the world as big data [18]. Customers spend most of their time for information searching, but finding the relevant information is very challenging task. There are systems and tools which identify the user's expectation and works on

behalf of them to get the desired result [1]. One such system is web recommender system which helps the customers to get their expected result by recommending the product, items, services or information [19].

Recommender system identifies the item of interest of current user from the bundle of items stored in the repository with respect to the current user interest and preferences [2]. To accomplish the target, the contemporary technique is introduced to synchronize the current user's requests with the repository of items and produce the relevant item recommendation to the users. Most of the e-commerce applications namely Amazon, Snapdeal, Jiomart and other similar type of services are using recommender systems to attract the customers by listing the fine set of item recommendations to make their work ease and comfort. This helps the organization to retain their customers [2][3].

1.1 Recommendation

Recommendation is a sub task of personalization which deals with the recommendation of one or set of items, products, web pages or services by relating the interests and preferences of the user. Three types of recommendation approaches are used in general namely (i) Recommendation based on User's Content search, (ii) Recommendation with respect to the collaborative effort of user and similar users (iii) Combination of (1) and (ii) [3][2][9][16]. In the above three approaches, collaborative and content based approaches are widely used methods for developing recommendation systems. The approaches which are listed above have its own for and against as per its applicability. In our previous approach, we made an attempt to develop collaborative based web recommender system [16]. The proposed research takes an attempt to develop an accurate recommendation system by combing the best features of collaborative and content based approach, known as hybrid approach.

Collaborative Filtering Approach

The collaborative filtering method recommends web pages to the lively users positioned on the preferences of added users from the user's repository [10]. Collaborative filtering approach recommends the web pages from the past experience of the similar type of users. The similar kind of users are identified by (i) users who give similar kind of rating for the item or product (ii) users who have similar browsing patterns. A collaborative filtering approach collects user's browsing patterns from the repository and identifies fair minded users. Users who have homogeneous browsing behavior are formed in one group [1][2][3].

Content Based Filtering Technique

It suggested an item to the users based on the relationship between the web contents [21]. The web pages are suggested with respect to the matters of the web pages in which the users has surfed in their past. This method built a browsing behavior of each user and recommends the web pages based on their own past browsing patterns [1][2][4].

Hybrid Recommender System

This system incorporates the approaches of the above two techniques to produce an accurate recommendation system [20][2][17].

Here, a contemporary recommender system is presented by adopting hybrid approach. To ameliorate the exactness of the recommender system, k-means clustering technique and PSO algorithm are proposed. Here the core idea of developing recommender system is to give the personalized recommendation to the users based on their interest and wish by referring their own browsing behavior and the behavior of similar type of users. This research work is assembled as follows: The second chapter deals with existing research work done by the researchers on this field. Chapter 3 explains the clustering algorithms. The achievement of the current method is discussed in chapter 4. Finally chapter 5 highlights about the conclusion and future scope of the research in the similar field.

2. LITERATURE SURVEY

The authors of this research [2] used the collaborative filtering approach to produce similarity based recommendations. They incorporated clustering techniques to group the similar kind of users to upgrade the accuracy of the recommendation technique using collaborative methodology. The authors took an effort to develop bio-inspired based clustering methodology which is a combination of swarm intelligence and fuzzy clustering models for recommending an item. The successes rate of their recommendation system was evaluated by using various quality metrics.

Machine Learning techniques were incorporated in the e-commerce applications to recommend the products to the customer [3]. In addition to the recommendation system, personalized web recommendations were implemented. Deep learning techniques were proposed [5] to identify the tempting property of learning feature representations from scratch. The effect of deep learning algorithms was proven in their research.

In this research [6], dynamic recommendation was proposed in combination of hierarchical temporal convolution networks and deep learning architecture. This methodology was proposed to scale up with billions of items and users. The result of the research result has proven the competence of the recommender system with the standard data set. The authors [7] addressed the issues and challenges in enhancing the accuracy of recommender services by proposing a new emotion based recommender system which focused on the hybrid user's features namely user rating, social network data and reviews. The efficiency of the recommender system was analyzed with the standard parameters and proven the higher prediction rating and improved recommendation accuracy.

In our earlier research [16][21], we the authors proposed that, the browsing history of the users are considered as the main part of recommendation. The interest of similar users is played a major role for developing an accurate recommendation system. The combination of collaborative filtering approach and bird flocking algorithm which is a bio inspired clustering algorithm proven the best result in recommendation system. Quality and time based association rule mining [20]is applied as the next step of clustering algorithm to optimize and further enhancement of the accuracy of the system. The obtained results were experimented with standard parameters to prove the accuracy of the system. The web logs of the users were the important considerations in this research.

3. CHALLENGES AND ISSUES OF RECOMMENDATION SYSTEM

Recommender system controls the problem of information excess in the area of information retrieval by suggesting appropriate and relevant items to the users of information search. But,

the designing of such recommender systems have many issues and challenges which requires the attention of the researchers [10][11].

3.1 Cold Start Problem

It [3][2] is actually inherited from the commonly occurring problem in vehicles especially cars. There are possibilities of happening cold start problem in e-commerce applications namely product cold start and user cold starts. When the new product, item or services are introduced in the market, there is no way of giving any recommendation on it, and then there may be a possibility of occurring product cold start [25].

3.2 Data Sparsity

The recommender systems are applied in large datasets [16]. So, the possible pattern comprises of user, item and its rating formed by the collaborative approach is immensely sizeable and sparse, which leads great opposition in the effectiveness of the system [11].

3.3 Scalability

The major challenge faced by collaborative recommendation approach is scalability [16], when the numbers of users and items are increased; the performance is decreased in this approach [11][10].

4. PROPOSED METHODOLOGY

In this research, hybrid recommendation approach is proposed to eliminate the difficulties of the stand-alone approach of the recommender system. The hybrid approach provides more specific and accurate recommendations to the users than the other two recommendation approaches [20][22][26]. The proposed research uses hybrid recommendation approach and clustering technique to provide an efficient recommender system. K-means clustering algorithm is utilized to identify the similarity. The particle swarm optimization (PSO) algorithm is biological nature of grouping the similar items. There are many bio-inspired algorithm namely ant colony, fish schooling, bird flocking etc. [8][16][14]. The proposed algorithm simulates the behavior of birds.

The existing research conveyed that the simple K-means clustering is suffering from premature convergence problem due to its random selection of the initial cluster centroids [7][12][13]. So, in order to avoid such an above mentioned issue, a PSO algorithm is proposed for evolving initial cluster centroid. The proposed system recommends items to the customers based on their own interest and the ratings of the item given by the other users as well who are similar to the current user. The proposed work falls on the below algorithm shown in Fig. 1.

Input data
C // current user searching pattern
U // existing user searching pattern
I // set of items
R // ratings
Output
N // Top N items for recommendation
Process
Begin
Preprocessing using Principle Component Analysis (PCA)
Find initial cluster centroid using PSO clustering
Find the exact cluster using k-means algorithm
Recommend Top N items
End

Fig. 1. Proposed Algorithm

4.1 Data Reduction using PCA

This step is proposed to extract the foremost features from the identified data. It also used to (i) extract only relevant attributes from the data repository set (ii) extract the representation of the data set. In the proposed work, the obtained data set is preprocessed using principle component analysis which follows linear feature extraction method which reduces the natural high dimensional data into low dimensional space. The resultant data of PCA is given as input to the PSO algorithm.

4.2 Clustering

Clustering is an unsupervised learning method of grouping the data items based on its similarity with pre-specified condition.

4.2.1 K-means clustering

It forms the given data items into k clusters, the data items in a group is closest to the centre of that particular group. The initial cluster centroids are randomly chosen by the algorithm. Based on the random initial centroid, all remaining data items are assigned to their closer centre. This process is done in iterative manner till there is no change in the cluster centroid. The proposed work uses Euclidean distance method to find the distance between data points and centroids [12][13][14]. The algorithm is given in Fig. 2.

Input data:
E = {e_1,e_2,e_3 …, e_n} //set of elements
k //number of desired clusters
Expected Output:
k //set of cluster
Process
Assign t_i ->C, where C is a random cluster
Find cluster mean
Iterate
Assign t_i ->C, where C is a cluster with closet mean
Find new cluster mean
Compute square error
Until
Min Square Error

Fig. 2. K-means Clusteing Algorithm

4.2.2 Particle Swarm Optimization (PSO)

The algorithm follows either the habit of bird congregate or fish schooling [15]. The algorithm was first introduced by Zennedy and Eberhart in 1995.The core concept behind the algorithm is, an individual demonstration as indicated by its restricted knowledge just as to the insight of the gathering. The individual watches the conduct of its neighbors and changes its own conduct appropriately [8][16][14].

4.2.3 Bird Flocking Algorithm

The proposed system uses bird flocking algorithm to optimize the clusters. The bird flocking algorithm [16] follows swarm intelligence algorithms which follow the flocking attitude of birds. This algorithm was first proposed by the author C. Reynolds [24] which is shown in Fig. 3.

In the original bird flocking algorithm, the boids are gathered with respect to the closeness value. The proposed approach uses the closeness calculation to group the items based on its rating [23]. The closeness Aij between two items is computed by using the equation,

$$A_{ij} = \sqrt{\sum_{k=0}^{L}(x_{ik} - y_{jk}} \quad (1)$$

where x_i and y_j are the items involved for closeness computation, *L* is the total number of items involved for similarity computation.

Data : Set of Items
Result : Group the Items.
Algorithm
Begin
For each Item do
Separation (Item);Cohesion (Item);Alignment (Item);
End
For each Item do
Find Affinity
End
End

Fig. 3. Bird Flocking Algorithm

4. EXPERIMENTAL RESULT

The proposed research methodology and its performance was evaluated by obtaining data set from two sources imdb and MovieLens [24]. The performance of the system was evaluated by applying the quality parameters namely recall, precision and f-measure. The data set of MovieLens contains 4500 users and 1,25,650 movie ratings. In imdb dataset, 1,60,250 movie ratings are obtained from 4,500 users.

4.1 Evaluation Metrics

Precision – The ratio of the number of the recommended items which are liked by the user over the total number of item recommended by the system.

$$Precision = \frac{TP}{TP+FP} * 1 \quad (2)$$

Recall - The ratio of the number of the recommended items which are not liked by the user over the total number of item recommended by the system. Accurate recommendation can be given by the system if the precision rate if higher and recall rate is lower.

$$Recall = \frac{TP}{TP+FN} * 1 \quad (3)$$

F-measure–This metric is used to find the accuracy of the recommender system

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

where TP – True Positive and FN – False Negative

4.2 Performance of proposed System

Table 1 shows the accuracy of the system obtained from MovieLens and imdb data-base. The data obtained from data repository are divided into testing and training data sets.

Table 1. Performance Result

Testing Data Set %	MovieLens			Imdb		
	Precision	Recall	F-measure	Precision	Recall	F-measure
40	0.867	0.826	0.85	0.862	0.836	0.85
45	0.878	0.838	0.86	0.868	0.839	0.85
50	0.879	0.847	0.86	0.875	0.843	0.86
55	0.869	0.849	0.86	0.865	0.841	0.85
60	0.878	0.846	0.86	0.872	0.836	0.85
65	0.889	0.848	0.87	0.879	0.842	0.86

Fig. 4. Performance of MovieLens dataset

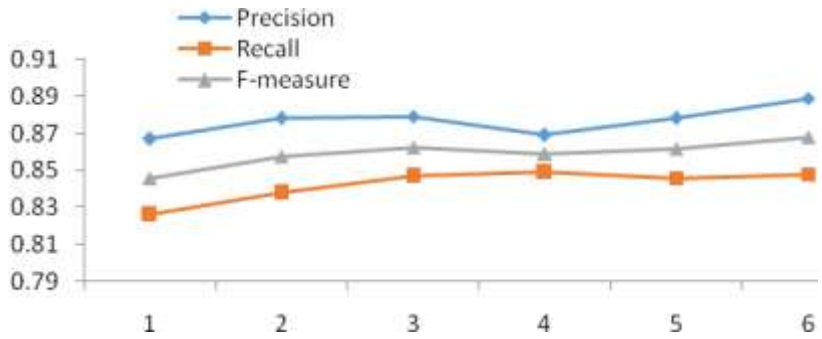

Fig. 4. Performance of the system in MovieLens dataset

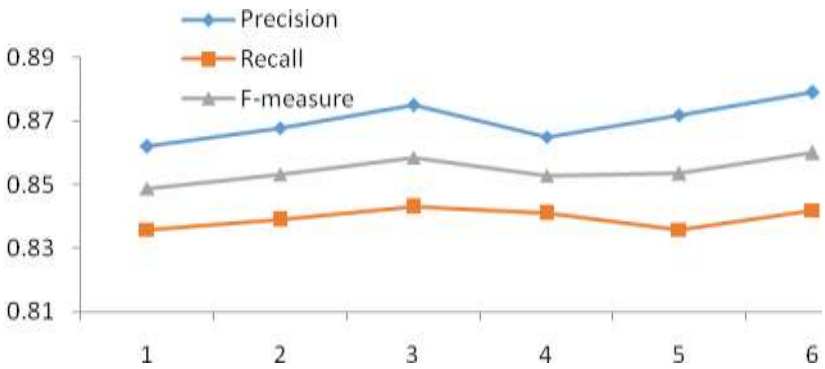

Fig. 5. Performance of the system in imdb dataset

The performance of the proposed work is tested by giving different testing data set ranges from 40% to 65%. Fig. 4 and Fig. 5 show the result of proposed work based on the parameter precision, recall and f-measure. The proposed algorithm gives improved performance in all the metrics. From the figure it is found that the proposed work gives the accuracy rate of 87% in MovieLens data set and 86% in imdb data set in f-score values.

5. CONCLUSION

Knowing the behaviour of internet users is a challenging task. In this research, an accurate recommendation system was introduced by utilizing the searching pattern and content of the internet surfers. The k-means and particle swarm optimization algorithm were effectively used to form the similar users and for further recommendations. The accuracy of the recommendation system was proven by using the database MovieLens and imdb. The research has given 87% accuracy for recommending the items to the live users. In future, personalized recommendation system may be proposed by incorporating user personal characteristics.

6. REFERENCES

[1]. Timur Osadchiya, Ivan Poliakova, Patrick Olivier, Maisie Rowland, Emma Foster.: Recommender System Based on Pairwise Association Rules, Expert Systems with Applications, 15, 535-542, (2019).
[2]. Logesh, R., Subramaniyaswamy, V., Malathi, D., Sivaramakrishnan, N., Vijayakumar, V.: Enhancing Recommendation Stability of Collaborative Filtering Recommender System through Bio-Inspired Clustering Ensemble Method, Springer Verlag London, 32, 2141–2164, (2020).
[3]. Jyoti Shokeen, Chhavi Rana.: A Study on Features of Social Recommender Systems, Artificial Intelligence Review, 53, 965–988, (2020).
[4]. Rana, C., Jain, S.K.: A Study of the Dynamic Features of Recommender Systems, Artificial Intelligence Review, 43(1), 141– 153, (2015).
[5]. Shuai Zhang, Lina Yao, Aixin Sun, Yi Tay.: Deep Learning Based Recommender System: A Survey and New Perspectives, ACM Computing Survey, 1-30, (2018).
[6]. Jiaxuan You, Yichen Wang, Aditya Pal.: Hierarchical, Temporal Convolution Networks for Dynamic Recommender Systems, ACM, 2236-2246, (2019).
[7]. Ahmadyfard, A., Modares, H.: Combining PSO and K-Means to Enhance Data Clustering, International symposium on telecommunications, 688–691, (2008).
[8]. Van Merwe, D. W., Engelbrecht, A P.: Data Clustering using Particle Swarm Optimization, Congress on evolutionary computation, 1, 215–220, (2003).
[9]. Cristina Vaz, Z P., Matos, D M.: Improving a Hybrid Literary Book Recommendation System through Author Ranking, ACM, 387-388, (2012).
[10]. Mehrbakhsh Nilashi.: An Overview of Data Mining Techniques in Recommender Systems, Journal of Soft Computing and Decision Support Systems, 3(6), 2289-8603,(2016).
[11]. Kennedy, J., Eberhart, R.: Particle Swarm Optimization, Proceedings of the IEEE International Conference on Neural Networks, 1942–1948, (1995).
[12]. Abdul Nazeer, K. A., Sebastian, M. P.: Improving the Accuracy and Efficiency of the K-Means Clustering Algorithm, International Conference on Data Mining and Knowledge Engineering (ICDMKE), Proceedings of the World Congress on Engineering,1, (2009).
[13]. Chen Zhang, Shixiong Xia.: K-Means Clustering Algorithm with Improved Initial Center, Second International Workshop on Knowledge Discovery and Data Mining, 790-792, (2009).

[14]. Yuan, F., Meng, Z. H., Zhangz, H. X., Dong, C. R.: A New Algorithm to get the Initial Centroids, Proc. of the 3rd International Conference on Machine Learning and Cybernetics, 26-29, (2004).
[15]. Yan Jiang, Tiesong Hu, Chong Chao Huang, Xianing Wu.: An Improved Particle Swarm Optimization Algorithm, Applied Mathematics and Computation, 193(1), 231-239, (2007).
[16]. Suguna, R., Sharmila, D.: An Efficient Web Recommendation System using Collaborative Filtering and Pattern Discovery Algorithms, International Journal of Computer Applications, 70(3), 37-44, 2013.
[17]. Nadi, S., Saraee, M H., Bagheri, A.: A Hybrid Recommender System for Dynamic Web Users, International Journal Multimedia and Image Processing, 1(1), 3-8, (2011).
[18]. Philip Chen, L., Zhang, Y.: Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Informatics and Computer Science Intelligent Systems Applications, 275, 314-347, (2014).
[19]. Jia Li Osmar, Zaiane, R.: Combining Usage, Content, and Structure Data to Improve Web Site Recommendation. International Conference on Electronic Commerce and Web Technologies, Springer Verlag Berlin Heidelberg, 305–315, (2004).
[20]. Burke, R.: Hybrid Recommender Systems: Survey and Experiments. User Model User Adapted Interact, 12(4) 331–370, (2002).
[21]. Michael Pazzani, J., Daniel Billsus.: Content-Based Recommendation Systems, Springer Verlag Berlin Heidelberg, 4312, 325 – 341, (2007).
[22]. Manogaran, G., Varatharajan, R., Priyan, M. K.: Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro Fuzzy Inference System, Multimedia Tools and Applications, 77, 4379–4399, (2018).
[23]. Kyung Yong, Jung Hyun.: User Preference Mining through Hybrid Collaborative Filtering and Content Based Filtering in Recommendation System. IEEE Multimedia Tools and Applications, 77, 4379–4399, (2018).
[24]. Craig W. Reynolds.: Flocks, Herds, and Schools:A Distributed Behavioral Model, Computer Graphics, 21(4), 25-34, (1987).
[25]. 25. Viji, C., Rajkumar, N., Suganthi, S.T. et al. An improved approach for auto
[26]. matic spine canal segmentation using probabilistic boosting tree (PBT) with
[27]. fuzzy support vector machine. J Ambient Intell Human Comput (2020).
[28]. 26. Sujatha, K & Shalini Punithavathani, D 'Extraction of well Exposed Pixels
[29]. for Image Fusion with sub banding Technique for High Dynamic Range imag
[30]. es.International journal of Image and Data Fusion', Taylor & Francis,
[31]. DOI: 10.1080/ 19479832.2016.1226967.