A Movie Recommendation System using Collaborative Approach and Cuttlefish Optimization

Senbagaraman.M[1] , Senthilkumar.R[2] , Subasankar.S[3], Indira.R[4]

[1,2,3] *U.G Scholar, Department of Computer Science and Engineering,*
[4]*Assistant Professor, Department of Computer Science and Engineering,*
Adhiyamaan College of Engineering (Autonomous),Hosur, Tamilnadu ,India

Abstract**—** *The Internet is a living entity, always changing and evolving. In addition to an evolving Internet, the usage of devices and systems has tremendously increased which leads to the need of recommended systems. The recommended system is one in which the user will receive the intended or appropriate information according to the user choices over the Internet. The proposed method combines the collaborative filtering technique, which carry out recommendations by referring to the public, who has the same taste. The System uses K-means clustering by adopting nature inspired cuttlefish optimization algorithm, which works on two parameters namely reflection and visibility. The system is built with the help of IMDB dataset for movie choices. It is also compared with the existing system which uses cuckoo search optimization and the results have been analyzed.*

Keywords**—***Collaborative, Cuttlefish, Clustering, Recommendation*

I. INTRODUCTION

With boon of Internet, the users of today's system were prone to high amount of choices. The information overload phenomenon have invaded in every fields of lives. Recommendations narrow down what could become a complex decision to just a few recommendations. The research in the field of recommendation system is still growing and not yet implemented full due to some of the practical constraints and limitation. Prediction delivery system or the recommendation system have added the economy of some of the famous ecommerce sites. It is estimated that 2/3rd of movies watched on Netflix are based on the recommendation.[21] When the quality of recommendation gets started to increase, it directly adds up another dimension in the space of recommendation.Web personalized recommendations are one which gains more importance in current real world applications[2]. First known recommendation system is "Usenet", a distributed discussion system made by Duke University in the late 1970s, which operated in client/server format, allowing users input to be categorized into specific newsgroups. The best known example is targeted advertisement on the internet. They are more prone to recommending accurate advertisement based on the searches. A recommendation system suggests a few data points on a large pool of data. There are several classic approaches for generating recommendations: collaborative filtering, content-based, case-based and hybrid methods. This paper suggest a method of combining collaborative filtering along with k-means clustering techniques and uses bio-inspired cuttlefish optimization procedures to carry out the recommendation steps. For example, taking one user's bookshelf and crosschecking it with shelves of other users, finding those with similar books will yield several possible book recommendations for that user. To carry out such an operation, ratings of items must be gathered and stored from a large number of users. This approach is called collaborative filtering. The proposed method is built by taking the taste of users in watching movies and then recommending the best optimized results to the users.

The paper is organized as follows: section 2 describes the related and background work, section 3 describes the proposed solution, section 4 describes the algorithms employed on the proposed system and section 5 describes the evaluation metrics on the designed system.

II. RELATED WORK

Over the decades recommendation system has been increasing. Tapestry[3] is one of the earliest method to create a recommended system by considering collaborative approach consist of workgroup peoples. Later Several rating based collaborative approach techniques have been evaluated. Philip et.al[4] applied content-based approach in research paper recommender system for a digital library. The authors used content-based technique as the recommendation technique. The system provides recommendations to users based on their taste or needs supplied as a query. The users' queries and research papers were represented as vectors of weights using vector space model and

TF-IDF weighing scheme. Cosine similarity was used to determine similar papers to the users' queries (supplied as input). This system did not use the ratings of an active user to provide recommendations.[5] An earlier study by Eyjolfsdottir et.al for the recommendation of movies through MOVIEGEN had certain drawbacks such as , it asks a series of questions to users which was time taking . On the other hand it was not user friendly for the fact that it proved to be stressful to a certain extent. Rendle et.al. [6] applied matrix factorization in a movie dataset with the aim of enhancing the quality and accuracy of the recommendation as well as reducing the complexity of building their model. Cumbreras et.al.[7] applied collaborative filtering algorithms and calculated user's score on a particular movie without having any remark, in which calculated a distance among the user and item when there is no relation at all. The proposed methodology eliminate the use of prompting series of questions to the user and also uses collaborative filtering approach inorder to find the user's personal choices to provide optimal recommendation of movies to the user.

III. Proposed Methodology

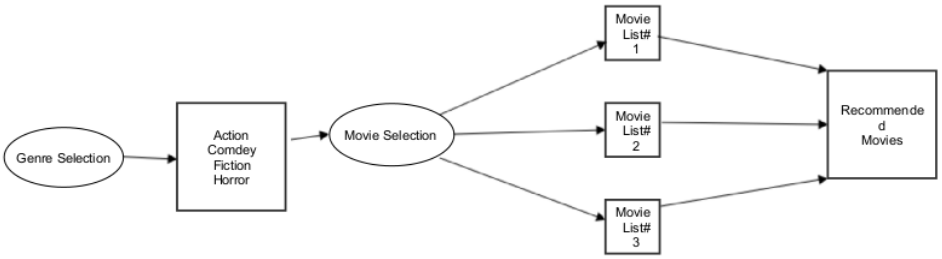
An alternate way of recommending movie to the user with the help of collaborative filtering along with k-means clusterization is proposed. The well-known Movie Dataset [10] is used to analyse the behaviour of the proposed system. In proposed method, users are prompted to select the genre of the movie, which they are likely to watch. A count of three genres are taken from the user in-order to recommend the movie list to the user. Upon the selection of the movie genre, a list of movies based on the selected genres will be display to the user for further selection. With the combination of above selections and by the use of K-means which are used to group the users of similar taste and cuttlefish algorithm to propose a optimal recommendations from the result of k-means , a list of movies which are worth watching for the users taste will be recommended. The flow of work can be seen.

IV. K-means –Cuttlefish Based Collaborative Filtering Framework

Clustering based techniques are broadly used in movie recommendation systems to reduce the problem of scalability.[8,9] Various researchers applied clustering-based methods on recommender systems that delivered expert recommendations. The objective of clustering is to partition objects into groups known as clusters in such a way that two objects within the same cluster have a minimum distance between them to identify similar objects. For the part of optimization, Cuttlefish Optimization algorithm is employed, since it is more efficient than some of the famous nature inspired algorithm such as Genetic Algorithm, Particle Swarm Optimization(PSO),Bees Algorithm..The original K-means algorithm was proposed by MacQueen[11] .The ISODATA algorithm by Ball and Hall[12] was a nearly but sophisticated version of k-means. Clustering divides the objects into meaningful groups. The clusters for evaluation are selected randomly with users at any given space and they are inspected one by one by calculating the inter distance relationship between them. So each user's distance is compared to its cluster mean and relocated according to the smallest distance from any cluster's mean. The process is said to continue until no more relocations are possible. The tradition K means pseudo code can be seen in the Fig 4.1.

For the process of optimization the results of K-means algorithm Cuttlefish Optimization is employed. The cluster is prepared with a fitness function that helps in improving the user's centroid distances, whereas fitness function changes previous centroids for a limited number of iteration. Then classify the users again by calculating the minimum centroid differences or applying k-means again. Cuttlefish [13, 14] is a type of cephalopods which is

well-known for its abilities to change its colour to either seemingly disappear into its environment or to produce stunning reflections. Cuttlefish is made up of different layers such as Chromatophores(which are responsible for pigments)[15],Iridophores(which are responsible for reflecting the lights) [16] and



Leucophores (which are responsible for the white spots)[17].

The cuttlefish algorithm computes the solution by the sum of reflection and visibility property of the fish, as described in (1)

newSolution=Visibility+ Reflection (1)

The algorithm works by initializing the users from the dataset and evaluating the fitness function. Divide the entire set of population into four groups and find the average and best cases in each group based on the equations on reflection and visibility property, the process is said to run until the best solution is found. The fitness function is responsible for performing this evaluation and returning a positive number called as fitness value. This value reflects how optimal the solution is, the higher the number, the better and the solution. The working of K-means algorithm and cuttlefish optimization is seen in the Fig 4.2 and Fig 4.3 respectively.

Step 1: Accept the number of clusters to group data from the dataset.

Step 2: Initialize the first K clusters. Take Random sampling of k elements

Step 3: Calculate the arithmetic means of each cluster formed in the dataset.

Step 4: K-means assigns each record in the dataset to only one of the initial clusters. Each record is assigned to the nearest cluster using a measure of distance.

Step 5: K-means re-assigns each record in the dataset to the most similar cluster and re-calculates the arithmetic mean of all the clusters in the dataset.

Fig 4.1: Pseudo code for K-means Algorithm

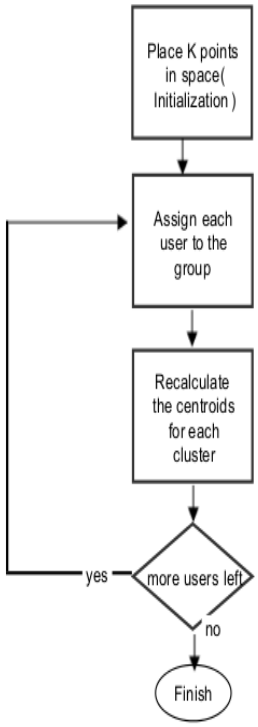


Fig 4.2: Working of K-means Algorithm

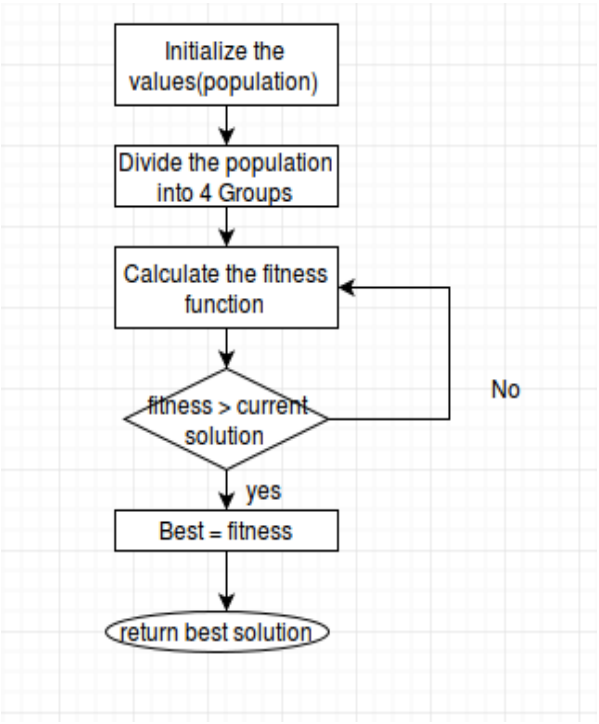


Fig 4.3: Working of Cuttlefish optimization

V. Evaluation Results

The Publicly available Movielens Dataset is taken into account which comprises of 40,000 movies by 260000 users and updated on October 2016[18].Movies are classified into 19 types viz. action, animation, horror, comedy, etc. Evaluation Metrics are key to understand the effective working process of the model designed for recommendation. In recommender systems, for the final the user the most important result is to receive an ordered list of recommendations, from best to worst. In fact, in some cases the user doesn't care much about the exact ordering of the list - a set of few good recommendations is fine.

A. Cross Validations

In order to estimate the accuracy of the system in the real world, the Cross-validation feature is used. In k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k−1 subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged to produce a single estimation. The cross validation result for the proposed model is found in Table 5.1. It is found from the above table, higher the folding (k values), higher the cross-validation

Sl.no	K	Cross-validation
1	2	210
2	3	280
3	4	315
4	5	336
5	6	350
6	7	360
7	8	364
8	9	368
9	10	378
10	11	380

Table 5.1:Cross Validation values based on k

B. Root Mean Square Error

The Root Mean Square Error (RMSE) (also called the root mean square deviation, RMSD) is a frequently used measure of the difference between values predicted by a model and the values actually observed from the environment that is being modelled. These individual differences are also called residuals, and the RMSE serves to aggregate them into a single measure of predictive power.

$$\mathrm{RMSE} = \sqrt{\frac{\sum(X_{\mathrm{obs},i} - X_{\mathrm{model},i})^2}{n}}$$

where X_{obs} is observed values and X_{model} is modelled values at time/place i.

C. Precision and Recall

Precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, while recall (also known as sensitivity) is the fraction of relevant instances that are retrieved. For the context of classification, the following four terms are employed, true positive(tp), true negative(tn) , false positive(fp), false negative(fn).

The precision and recall can be taken as[19]

$$Precision = \frac{tp}{tp + fp} \tag{2}$$

$$Recall = \frac{tp}{tp + fn} \tag{3}$$

The precision and recall curve can be plotted as follows:

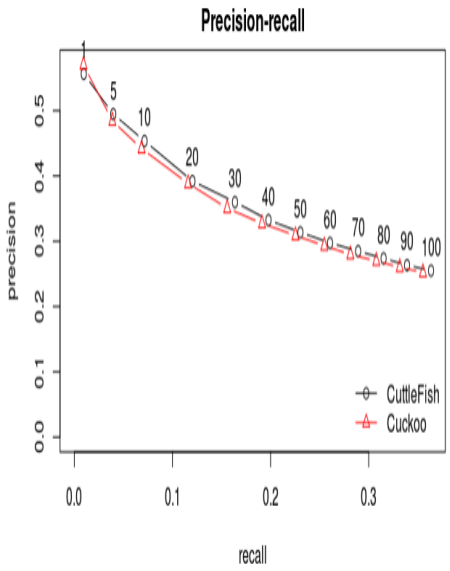


Fig 5.2:Precision Recall Curve

VI. Conclusion and Future Work

The developed system works more efficiently than the predecessors and has high efficiency in the cross validations and are capable of recommending highly reliable outcomes. Privacy protection consideration is also a challenge for recommender algorithms. The recommendation systems should also take care of individual privacy. The future work can be extended by implementing an hybrid variety of cuttlefish algorithm or any other nature inspired algorithm to increase the current accuracy and speed of the recommendation system.

References

[1] Katarya R, Verma OP. An effective collaborative movie recommender system with cuckoo search. Egyptian Informatics Journal 2016 http://dx.doi.org/10.1016/j.eij.2016.10.002

[2] Brusilovsky, P., Kobsa, A., Nejdl, W.: The adaptive web: methods and strategies of web personalization. Springer-Verlag New York Inc (2007).

[3] GoldBerg,D.,Nichols, D.,Oki, B.M., and Terry,D.(1992).”Using Collaborative Filtering to Weave an Information Tapestry”. Communications of the ACM.December.

[4] S. Philip, P. B. Shola, and A. O. John; “Application of Content-Based Approach in a Research Paper Recommendation System for a Digital Library”, International Journal of Advanced Computer Science and Application, Volume 5, Issue 10, 2014

[5] Eyrun A. Eyjolfsdottir, Gaurangi Tilak, Nan Li; MovieGEN: A Movie Recommendation System,Computer Science Department, University of California Santa Barbara.

[6] Rendle, S., Gantner, Z., Freudenthaler, C., & Schmidt-Thieme,L.(2011, July).“Fast context-aware recommendations with factorization machines”.In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 635-644). ACM.

[7] M.Á. García-Cumbreras, A. Montejo-Ráez, M.C. Díaz-Galiano, “Pessimists and optimists: Improving collaborative filtering through sentiment analysis”, Expert Syst. Appl. 40 (2013) 6758–6765.

[8] Georgiou O, Tsapatsoulis N. “Improving the scalability of recommender systems by clustering using genetic algorithms”. Lect Notes Comput Sci 2010;6352:442–9

[9] Xue G-R, Lin C, Yang Q, Xi W, Zeng H-J, Yu Y, et al. “Scalable collaborative filtering using cluster-based smoothing”. Scalable Collab Filter Using Clust Smoothing 2005:114.

[10] Movie lens dataset available at http://grouplens.org/datasets/movielens/

[11] MacQueen J.. “Some methods for classification and analysis of multivariate observations.”In Procedings Of the 5 th Berkeley Symposium on Mathematical Statistics and Probability, pages 281-297. University of California Press, 1967.

[12] .Ball G. and Hall D.. A Clustering Technique for Summarizing Multivariate Data. Behavior Science,12:153-155, March 1967. Bowman, M., Debray, S. K.,and Peterson, L. L. 1993. Reasoning about naming systems.

[13] .M. M. Lydia, J. D. Eric, and T. H. Roger, N. M. Justin,Mechanisms and behavioural functions of structural coloration in cephalopods, J. R. Soc. Interface, 2008.

[14] E. Florey, Ultrastructure and function of cephalopod chromatophores, Am. Zool. 1969

[15] R. A. Cloney, and S. L. Brocco, Chromatophore organs,reector cells, iridocytes and leucophores in cephalopods, Am.Zool. 1983.

[16] R. T. Hanlon, K. M. Cooper, B. U. Budelmann. and T. C.Pappas, Physiological color change in squid iridophores I,Behavior, morphology and pharmacology in Lolliguncula brevis, Cell and Tissue Research. 259, 1990.

[17] K. M. Cooper, R. T. Hanlon, and B.U. Budelmann,Physiological color change in squid iridophores II,Ultrastructural mechanisms in Lolliguncula brevis, Cell and Tissue Research. 259, 1990.

[18] Movie lens dataset descritption at http://grouplens.org/datasets/movielens/latest/

[19] Olson, David L.; and Delen, Dursun (2008); Advanced Data Mining Techniques, Springer, 1st edition (February 1, 2008), page 138.

[20] The Newyork Times,Date:Feb 16,2012, ”How Companies Learn on your Secrets”.The article found online at http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp

[21] Manoj Kumar,D.K. Yadav,Ankur Singh,Vijay Kr. Gupta,”A Movie Recommender System: MOVREC”,International Journal of Computer Applications (0975 – 8887) Volume 124 – No.3, August 2015.