Movie Recommendation System

Pooja
M.Tech Scholar
Department of Mechanical Engineering
Geeta Engineering College, Naultha, Panipat, India

Mr. Bhupender Sharma
Assistant Professor
Department of Mechanical Engineering
Geeta Engineering College, Naultha, Panipat, India

Abstract

The objective of this work is to assess the utility of personalized recommendation system (PRS) in the field of movie recommendation using a new model based on neural network classification and hybrid optimization algorithm. We have used advantages of both the evolutionary optimization algorithms which are Particle swarm optimization (PSO) and Bacteria foraging optimization (BFO). In its implementation a NN classification model is used to obtain a movie recommendation which predict ratings of movie. Parameters or attributes on which movie ratings are dependent are supplied by user's demographic details and movie content information. The efficiency and accuracy of proposed method is verified by multiple experiments based on the Movie Lens benchmark dataset. Hybrid optimization algorithm selects best attributes from total supplied attributes of recommendation system and gives more accurate rating with less time taken. In present scenario movie database is becoming larger so we need an optimized recommendation system for better performance in terms of time and accuracy.

Keywords: Matlab, Optimization, Particle Swarm Optimization, Movie Recommendation System

---

I. INTRODUCTION

It has been vividly seen that recommendation systems has been a major part of consumerism as majority of people relate recommendation systems as a apart of e-commerce sites and shopping destinations. Therefore, recommendation systems find numerous applications when it comes to purchasing or surfing information about any particular product .To substantiate my view, internet sites, shopping websites, electronics etc. are visited by people on daily basis foe meeting their needs. In addition to this, recommendation systems are conspicuously used in guiding people about variety of products that they enquire or they are interested in buying.

The category of the idea is based upon some general issues that may include previous buying behaviour of the customer and demographics of persons. Paul Re-snick and Hal R. Varian, in the year 1997 introduced the term Recommender System'. There were basically two prominent reasons behind finding this particular name as the first and foremost was recommenders could not work properly as the users were unaware of each other's identity, apart from that the other reason was their recommendation also included the interest of the user. Implicit feedback for recommendations was also being provided in combination with the aggregation techniques [2]. To exemplify, collaborative filtering of internet news was done by a system called Group lens, it worked on Two strategies first one was implicit feedback through reading time and explicit feedback through ratings. Furthermore, taking into consideration the privacy concerns of users a new strategy was also recommended and it was pseudonyms and to combine various recommendations weighted voting was used. Not only the efficiency but the cost was also given attention and it was found that maintenance and progress of these systems was hefty and it was important to identify whether the merits outweigh the demerits.

A. Classification of Recommendation System

Recommendation systems are broadly classified into three categories.
- Content-Based Recommendations

Recommendations are made solely based on the attributes of the products which the user preferred previously.
- Collaborative Recommendations

Recommendations are made based on the similarity between the preferences of users and not the content of the products. The recommended items will depend on what other similar users liked.
- Hybrid Approach

Recommendations are generated based on the resemblances between users and the analogy between products along with their given feedback.

B. Objective

Keeping these points in consideration following will be our objectives*:*
- To use Neural Network (NN) classification to recommend the movies based on query input
- To optimize the parameters of data attributes to achieve higher accuracy using hybrid algorithm which is combination of two powerful optimization algorithms PSO and BFO
- To use the overall accuracy as the objective function

– Movie lens recommendation dataset will be used for training and testing using NN classifier.

II. Proposed work

The data used here is movie lens dataset which has a total of 8 attributes to be optimally selected. Dataset description is detailed in next section of this chapter. Out of these attributes, Hybrid Algorithm will choose all those only which contribute more in the accuracy improvement in recommendation. The number of bacteria in Hybrid Algorithm acts as the number of options available for attributes selection at an iteration and their position is the column choice in the data. The whole data is first divided into testing and training datasets. 80% of data is used for training and rest is used for testing. For the first iteration, each position of agent is chosen randomly which is for number of attributes set for an iteration. The selected attributes for them are used to make trained model using NN classifier and tested for the query data. This accuracy is noted down in a matrix for each Hybrid algorithm bacteria.

A complete step by step Hybrid algorithm is explained below.

1) Step1: Load the movie lens dataset in numeric format and divide that into random 80:20 ratio for training and testing of recommendation engine.
2) Step2: Initialize the hybrid algorithm parameters.
3) Step3: Randomly initialize the bacteria's new positions which must be either 1 or 0 and will choose the attributes out of 8 in total.
4) Step4: Call the objective function to train the model for selected attributes in training data and test the model for testing data to get the recommendation accuracy.
5) Step5: To update the random positions of bacteria, using PSO updating equations for partice swarm position update.
6) Step6: The new updated position is obtained from the equation (3.1) by using velocity update function of PSO
7) Step7: For this new updated position or values of weights and biases, objective function is again called and accuracy is saved.
8) Step8: The attributes' positions for which minimum of accuracy is obtained out of previous two set of values, is further considered for updating.
9) Step9: This process continues till all iterations are not completed.
10) Step10: The final maximum accuracy is obtained and attributes selected for them are used as final set of attributes which gives higher accuracy.

III. Result

We further optimized the attributes for the accuracy improvement using hybrid optimization and compared the results with GA and hybrid algorithm. Any optimization algorithm works well if it converges early and settle to a maximum value (in our case) with no further changes. We plotted iteration curve for hybrid algorithm to check whether our optimization is selecting the optimal attributes or not. Error given by hybrid algorithm (combination of PSO and BFO) is also very less as comparison to GA algorithm. Other values such as prevalence , specificity, positive likelihood, positive predictive value and error are also indicate that results produced by hybrid algorithm is convincing than GA. We used neural network classifier to recommend the movies similar to user selected movie. The multiclass here because the rating of movie is the criteria on which NN modeling will recommend the movie and this rating is in between 1-5.

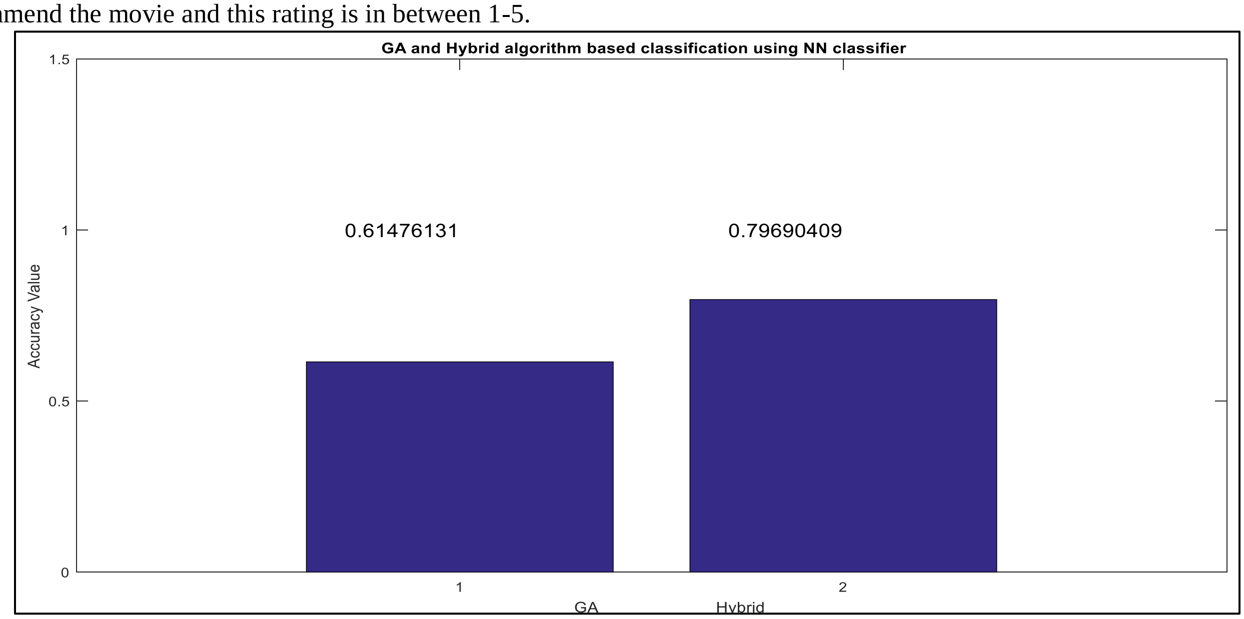

Fig. 3.1 (a): Accuracy Comparison Plot for Optimization Algorithm

Figure 3.1 (a) shows the bar chart of accuracy obtained from both the algorithms. Here it is clearly evident that accuracy of getting correct rating of movie using proposed hybrid algorithm is better 0.7969 than accuracy obtained from GA which is 0.614.

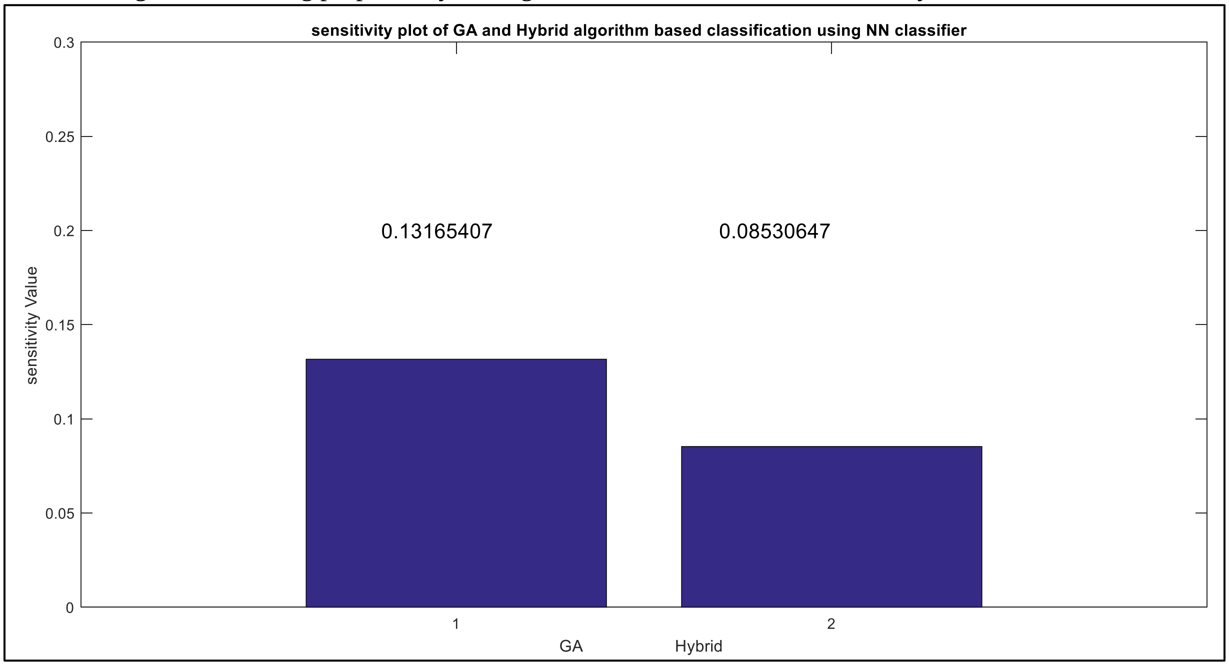

Fig. 3.1 (b): Sensitivity Comparison Plot for Optimization Algorithm

Figure 3.1 (b) shows the bar chart of sensitivity obtained from both the algorithms. Here it is clearly evident that sensitivity using proposed hybrid algorithm is less which is better (0.085) than sensitivity obtained from GA which is 0.1316.

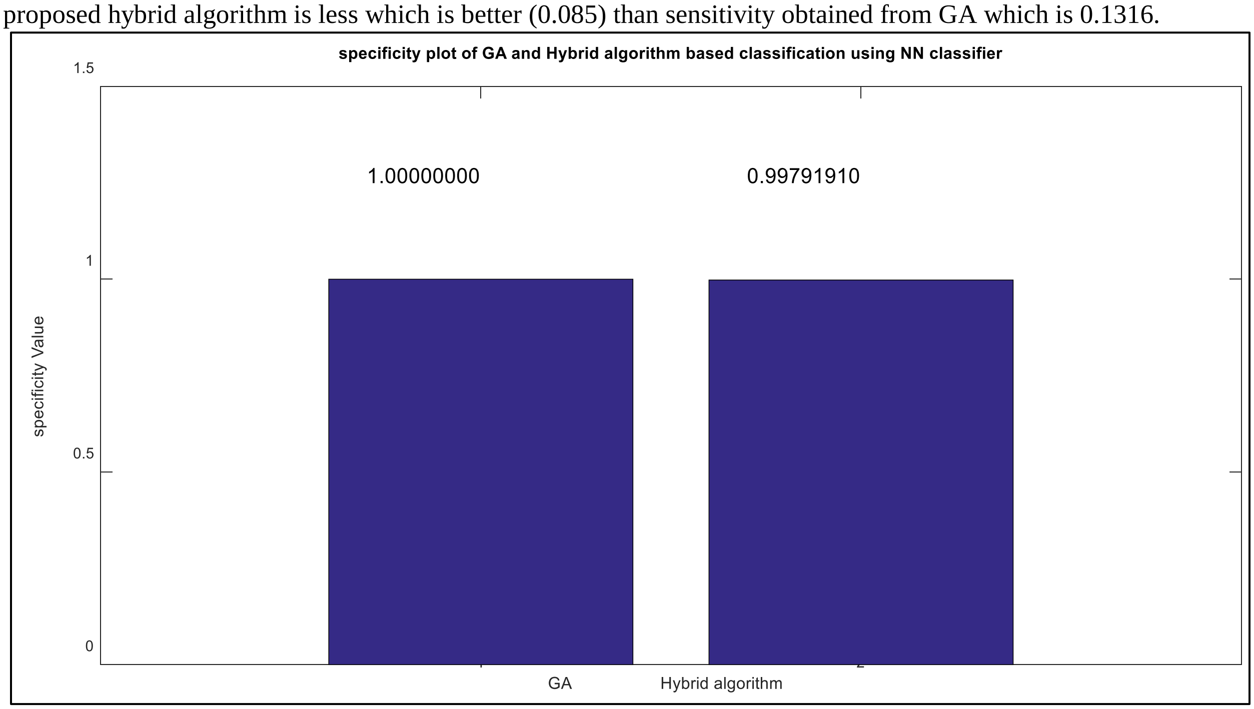

Fig. 3.1 (c): Specificity Comparison Plot for Optimization Algorithm

Figure 3.1 (c) shows the bar chart of specificity obtained from both the algorithms. Here it is clearly evident that specificity using proposed hybrid algorithm is almost same as that of GA.

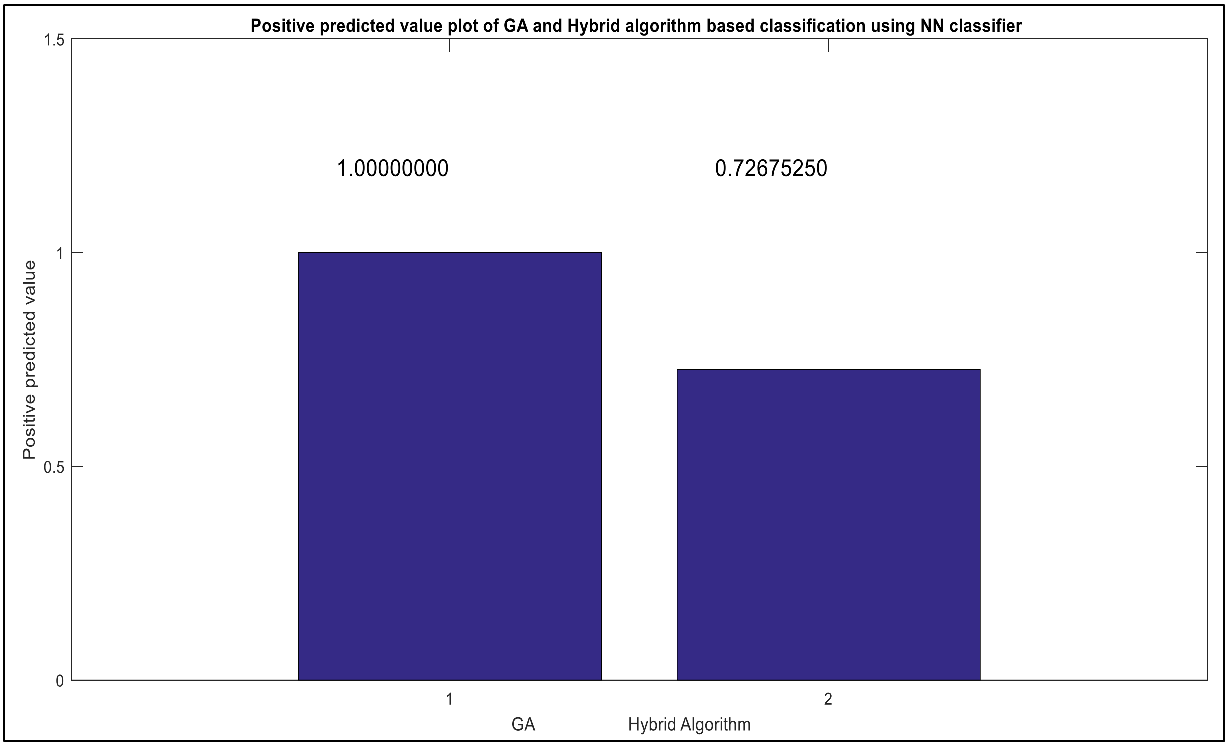

Fig. 3.1 (d): Positive Likelihood Comparison Plot for Optimization Algorithm

Figure 3.1(d) shows the bar chart of positive likelihood obtained from both the algorithms. Here it is clearly evident that positive likelihood using proposed hybrid algorithm is less which is better (0.7267) than positive likelihood obtained from GA which is 1.

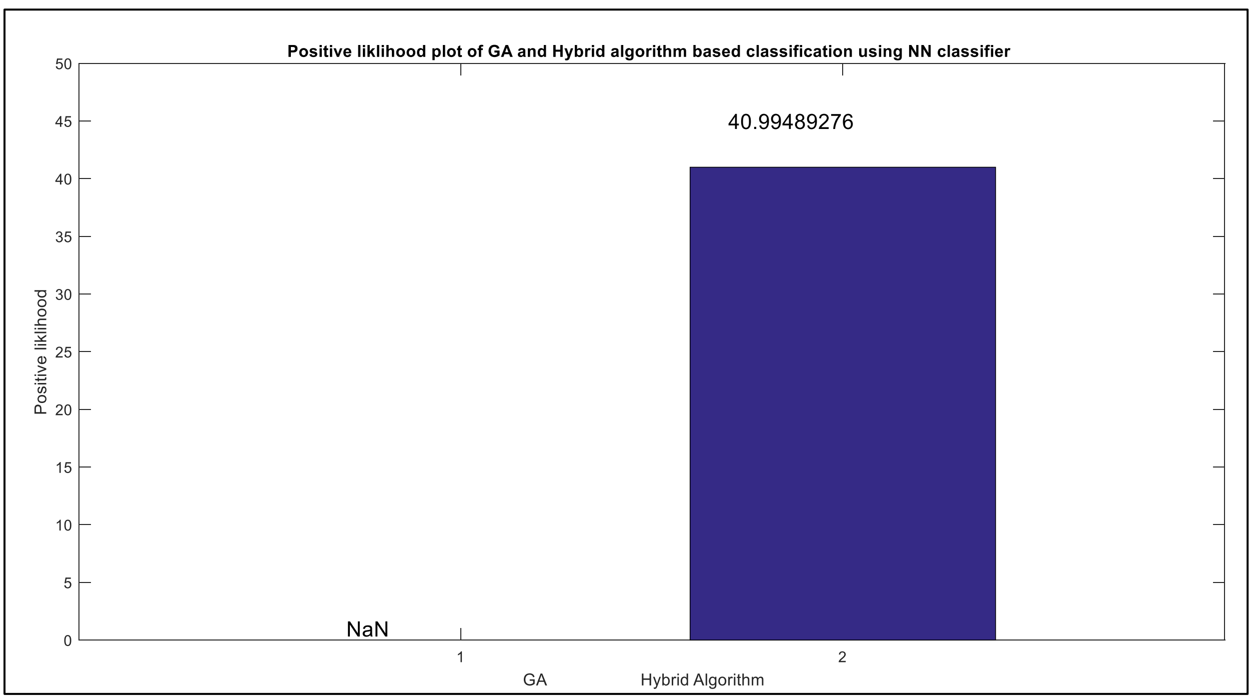

Fig. 3.1 (e): Positive Predictive Value Comparison Plot for Optimization Algorithm

Figure 3.1 (e) shows the bar chart of positive predictive value obtained from both the algorithms. Here it is clearly evident that positive predictive value of getting correct rating of movie using proposed hybrid algorithm is better 40.999 whereas positive predictive value not obtained in GA.

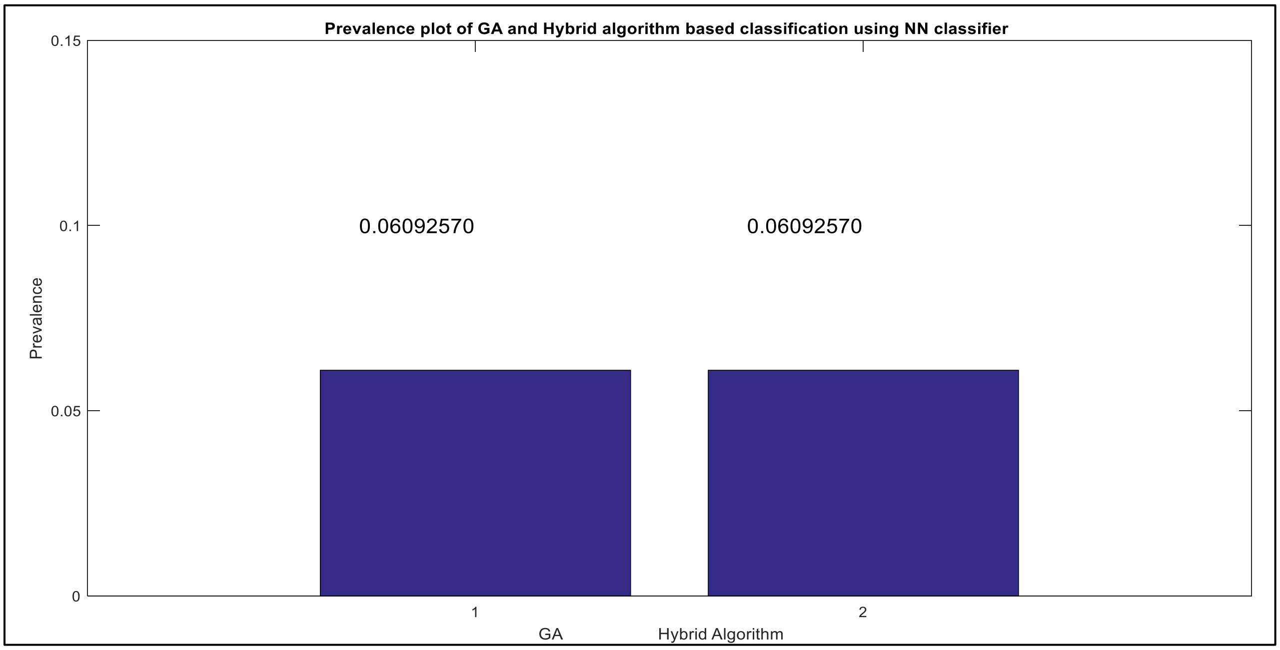

Figure 3.1 (f): Prevalence Comparison plot for optimization algorithm

Figure 3.1(f) shows the bar chart of prevalence obtained from both the algorithms. Here it is clearly evident that prevalence using proposed hybrid algorithm is same as that of GA.

For better comparison results class-wise, confusion matrix is generally used to plot. So, we used confusion matrix plots to analyze the accuracy performance by each class. Confusion matrix will tell us how many samples were picked for each class and exact number of samples detected in each class after model testing. Figure 4.2 shows the confusion matrix plot for these two optimizations. The yellow highlighted box in confusion map represents the accuracy % for that class.

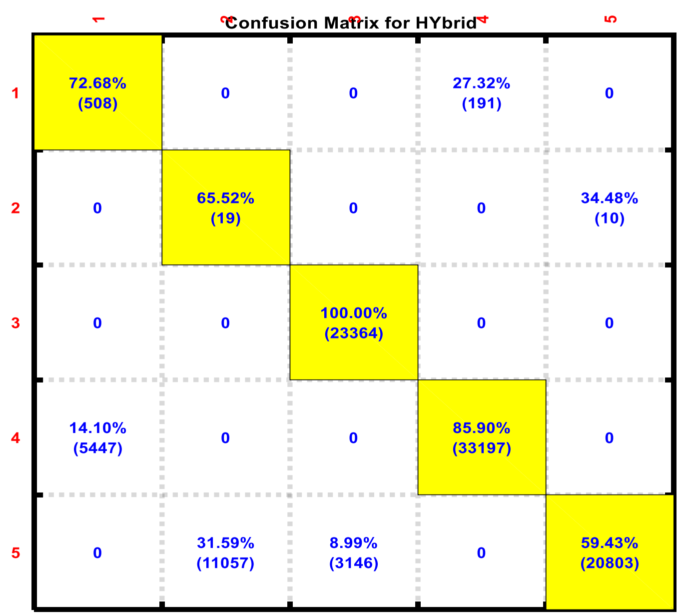

Figure 3.2(a): Confusion Matrix Plot for Hybrid Algorithm Selected Attributes

Figure 3.2 (a) shows the confusion matrix chart of predicted value vs original values obtained from proposed hybrid algorithm. Yellow diagonal elements shows truly predicted output labels by hybrid algorithm. It is observed that for class 1,2,3,4,5 truly predicted ratio are 72.6 %,65.52 %,100 %,85.9 %,59.43 % respectively. It means average predicted values are 77% which is very good.

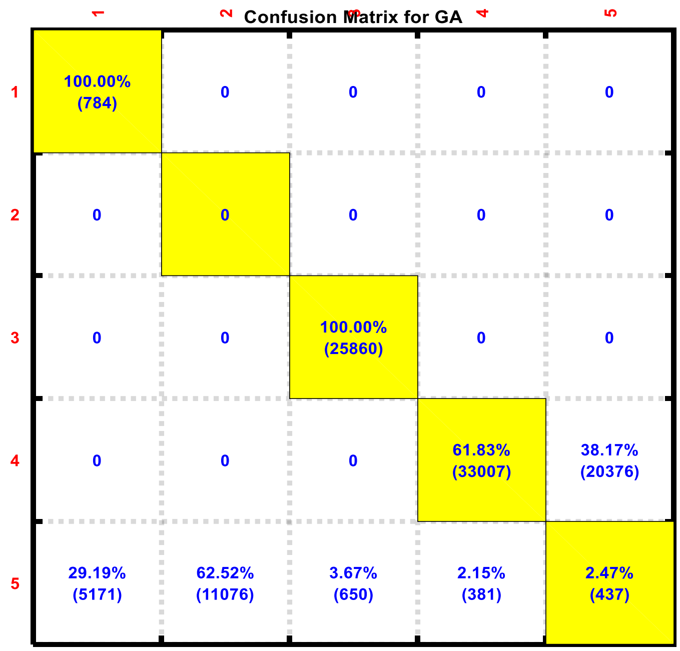

Fig. 3.2(b): Confusion Matrix Plot for GA Selected Attributes

Figure 3.2 (b) shows the confusion matrix chart of predicted value vs original values obtained from GA algorithm. Yellow diagonal elements shows truly predicted output labels by hybrid algorithm. It is observed that for class 1,2,3,4,5 truly predicted ratio are 100 %, 0 %, 100 %, 61.83 %, 2 % respectively. It means average predicted values are 52% which is less in comparison to our proposed algorithm. Also there is very low value in prediction of two class labels which are class 2 and class 5.

IV. Conclusion

In this work a comprehensive study of personalized recommendation system (PRS) in the field of movies recommendation has been carried out. In order to perform the above-mentioned investigation, we have used standard Movie Lens dataset downloaded from web link to create training model using new regression based model. The data is in qualitative form which is converted to quantitative to use in recommendation model. We have proposed a new regression model which is based on neural network (NN) classifier and hybrid optimization algorithm which is combination of BFO and PSO algorithm. We have 1 million user's rating for 1682 movies which is used to make a training model but firstly we have retained important attributes using hybrid algorithm optimization and these reduced features are then utilized to create training model using NN classifier and this model is tested using priory separated testing data and applied to trained model to find out output label accuracy.

References

[1] S. Agrawal, P. Jain, "Performance Evaluation of the Movie Recommendation Systems", International Journal of Computer Applications (0975 – 8887) Volume 158, Issue 2, January 2017.
[2] R. Katarya, O. Prakash Verma, "An effective collaborative movie recommender system with cuckoo search", Egyptian Informatics Journal 18, volume 145, Pages 105–112, March 2017.
[3] Wang X, Luo F, Qian Y, Ranzi G, "A Personalized Electronic Movie Recommendation System Based on Support Vector Machine and Improved Particle Swarm Optimization".PLoS ONE 11(11): e0165868. doi:10.1371 /journal, april 2016.
[4] K. Wegba, A. Lu 1, Yuemeng Li 1, and Wencheng Wang," Interactive Movie Recommendation Through Latent Semantic Analysis and Storytelling", volume 55, issue 4, January 2017.
[5] L. Zhao, Z. Lu, S. Jialin Pan, Q. Yang," Matrix Factorization+ for Movie Recommendation", Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), volume 32, march 2016.
[6] R. Hande, A. Gutti, K. Shah, J. Gandhi, V.Kamtikar, "Movie-mender- A Movie Recommender System", IJESRT, volume 64, November, 2016.
[7] T. Arsan, E. Köksal, Z.Bozkuş, "Comparison Of Collaborative Filtering Algorithms With Various Similarity Measures For Movie Recommendation", International Journal of Computer Science, Engineering and Applications (IJCSEA) Volume 6, Edition3, June 2016.
[8] K. Mohit, B. Rajdavinder Singh and S. Gurpreet,"K-Means Demographic based Crowd Aware Movie Recommendation System", Indian Journal of Science and Technology, Volume 9, Edition 23, June 2016.
[9] P. Yi, C. Yang, X. Zhou and C. Li, "A movie cold-start recommendation method optimized similarity measure," 2016 16th International Symposium on Communications and Information Technologies (ISCIT), Qingdao, may 2016.
[10] O. Bendre, M. Mule and P. Nimbolka, "Movie Recommendation System Considering Multiple scenarios", International Journal of Advance Engineering and Research Development Special Issue for ICPCDECT, Volume 3, Issue 1, 2016.
[11] J. Dodge, A. Gane , X. Zhang , A. Bordes, S. Chopra, A. H. Miller, A. Szlam& J. Weston, "Evaluating Prerequisite Qualities For Learning End -To-End Dialog Systems, volume 32, April 2016.
[12] O. Bendre, M. Mule and P. Nimbolka, "Movie Recommendation System Considering Multiple Scenarios", International Journal of Advance Engineering and Research Development Special Issue for ICPCDECT, Volume 3, Issue 1, 2016.

[13] W. X, L. F, Q. Y, R. G (2016), "A Personalized Electronic Movie Recommendation System Based on Support Vector Machine and Improved Particle Swarm Optimization." PLOSONE 11(11): e0165868.doi:10.1371/journal, November 29, 2016.
[14] K. Wakil, K. Ali, R. Bakhtyar, "Improving Web Movie Recommender System Based on Emotions", International Journal of Advanced Computer Science and Applications, Volume 6, Issue 2, September 2015.
[15] B. Bhatt, Prof. P. J Patel, Prof. H. Gaudani, "A Review Paper on Machine Learning Based Recommendation System", IJEDR, ISSN: 2321-9939, Volume 2, Issue 4, October 2015.
[16] R. Kim, Y. J. Kwak, H. Mo, M. Kim, S. Rho1, K. L. Man, W. K Chong, "Trustworthy Movie Recommender System with Correct Assessment and Emotion Evaluation", International Multi Conference of Engineers and Computer Scientists 2015 Volume 2, IMECS 2015, Hong Kong, March 18 - 20, 2015.
[17] M. Kumar, D.K. Yadav, A. Singh, V. Kr. Gupta, "A Movie Recommender System: MOVREC", International Journal of Computer Applications (0975 – 8887) Volume 124, Issue 3, August 2015.
[18] A.Saranya1 , A.Hussain," User Genre Movie Recommendation System Using NB Tree", International Journal of Innovative Research in Science, Engineering and Technology, Volume. 4, Issue 7, July 2015.
[19] R. Kim, Y. J. Kwak, H. Mo, M. Kim, S. Rho1, K. L. Man, W. K Chong, "Trustworthy Movie Recommender System with Correct Assessment and Emotion Evaluation", International Multi Conference of Engineers and Computer Scientists 2015 Volume 2, IMECS 2015, Hong Kong, March 18 - 20, 2015.
[20] G. Arora, A. Kumar, G. S. Devre, Prof. Amit Ghumare, "MOVIE RECOMMENDATION SYSTEM BASED ON USERS' SIMILARITY", International Journal of Computer Science and Mobile Computing, IJCSMC, Volume 3, Issue 4, Pages765 – 770, April 2014.
[21] M. Kumar, D.K. Yadav, A. Singh, V. Kr. Gupta, "A Movie Recommender System: MOVREC", International Journal of Computer Applications (0975 – 8887) Volume 124, Issue 3, August 2015.
[22] Z. Wang, X. Yu*, N. Feng, Z. Wang, "An Improved Collaborative Movie Recommendation System using Computational Intelligence" ,Journal of Visual Languages & Computing, Volume 25, Issue 6, Pages 667–675,December 2014.
[23] G. Arora, A. Kumar, G. S.Devre, Prof. Amit Ghumare, "MOVIE RECOMMENDATION SYSTEM BASED ON USERS' SIMILARITY", International Journal of Computer Science and Mobile Computing, IJCSMC, Volume 3, Issue 4, Pages765 – 770,April 2014..
[24] V. Adi Lakshmi, J. KanakaPriya, "A Hybrid Approach to Movie Recommender System", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9, September 2013.
[25] D. Roy, A. Kundu, "Design of Movie Recommendation System by Means of Collaborative Filtering", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013.
[26] A. Kundu, "Design of Movie Recommendation System by Means of Collaborative Filtering", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013.
[27] S. Halder, A. M. J. Sarkar and Y. K. Lee,"Movie Recommendation System Based on Movie Swarm," *2012Second International Conference on Cloud and Green Computing,* Xiangtan, Pages 804-809, November 2012.
[28] Mirizzi, Roberto, Tommaso Di Noia, A. Ragone, V. C.Ostuni and Eugenio Di Sciascio. "Movie Recommendation with DBpedia." IIR, volume 5, pages 405-410, December 2012.
[29] M. Choi and Y. S. Han, "A Content Recommendation System Based on Category Correlations," 2010 Fifth International Multi-conference on Computing in the Global Information Technology, Pages 66-70, Valencia, June 2010.
[30] B. Bhatt, Prof. P. J Patel, Prof. H. Gaudani, "A Review Paper on Machine Learning Based Recommendation System", IJEDR, ISSN: 2321-9939, Volume 2, Issue 4, august 2010.
[31] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Reidl "Item-based collaborative filtering recommendation algorithm", citeseer.ist.psu.edu/sarwar01itembased.html, 2001.
[32] J. L. Herlocker, "Understanding and Improving Automated Collaborative Filtering Systems", PhD Thesis, 2000.
[33] Rennock, E. Horvitz, S. Lawrence, and C. L. Giles."Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-based Approach". Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, UAI, citeseer.ist.psu.edu/pennock00collaborative.html, 2000.
[34] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. "Combining Content-Based and Collaborative Filters in an Online Newspaper". cite- seer.ist.psu.edu/article/claypool99combining.html, 1999.
[35] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. "Combining Content-Based and Collaborative Filters in an Online Newspaper". citeseer.ist.psu.edu/article/claypool99combining.html, 1999.
[36] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl. "GroupLens," An Open Architecture for Collaborative Filtering of Netnews". Proceedings of ACM Conference on Computer Supported Cooperative Work, 1994.

Serial no.	Technology Used	Author's Name	Advantages	Disadvantages
1.	Quantitative evaluation	Shreya Agrawal, et al.[1]	Summarise the performance evaluation parameters for movie recommendation system	Doesn't make the comparative importance of parameters
2.	Cuckoo search optimization algorithm	Rahul Katarya, et al.[2]	K means clustering is used and modified by cuckoo search optimisation	Again unsupervised approach which is less accurate then supervised
3.	Personalized recommendation system (PRS) and Support vector machine (SVM) classification	Xibin Wang, et al.[3]	Extracted the un-important features of the data	Use PSO based local optimisation for this purpose which has problem of premature termination
4.	Support vector machine (SVM) classification and an improved PSO (IPSO)	Wencheng Wang, et al.[4]	Presented an improved PSO algorithm with the evolution speed factor and the aggregation degree factor (IPSO).	Local optimization used
5.	Particle Swarm Optimization (PSO)and Gravitational Search Algorithm (HYBRID).	Lili Zhao, et al.[5]	Considers the visual features like form posters or movie scene	These features are not revealed from rating data
6.	Hybrid approach of Collaborative Filtering and Content-based methods	Rupali Hande, et al.[6]	It considers the preference of users and similarity of users	NA

7.	Statistical accuracy metrics	Taner Arsan, et al.[7]	Compared the item and user based prediction on different similarity matrix	Very old approach
8.	K-Means based crowd-aware recommender system	Kathpal Mohit, et al. [8]	Collaborative filtering is used	Unsupervised method is less accurate
9.	Using movie labels sets to identify similar sets	Peng Yi, et al.[9]	Resolved the cold start problem	Computation time is high in model training
10.	Movie Dialog dataset	Jesse Dodge, et al. [10]	Evaluate the performance of various methods on the new data created	NA
11.	Collaborative filtering using Mahout	Omkar Bendre, et al. [11]	Multiple test cases are used on the hadoop platform	No new method is proposed
12.	Hybrid approach of emotions detection algorithm, Content Based Filtering (CBF) and Collaborative Filtering (CF)	Karzan Wakil, et al.[12]	Better understanding of the relation between emotional states and the recommended movies.	NA
13.	Collaborative Filtering, Content based, hybrid based Approach, Memory based and Model based Recommendation	Bhumika Bhatt, et al.[13]	A review of various presently available techniques	NIL
14.	Quantitative evaluation	RyuRi Kim, et al. [14]	Provides trustworthy recommendation with combination of unfair ratings evaluation	Based on opinions
15.	K-means algorithm	Manoj Kumar, et al.[15]	Collaborative filtering is used	Unsupervised method is less accurate
16.	Prediction rating method and naïve Bayesian algorithm	A.Saranya, et al. [16]	User's interest and his social circle was used to analyse the movie recommendation	A lot of users data has to be used
17.	Improved K-means clustering coupled with genetic algorithms (GA)	Zan Wang, et al. [17]	Uses data reduction by PCA which exclude the unimportant features	Unsupervised method is used which is not accurate as supervised model
18.	Hybrid approach of Collaborative Filtering and Content-based methods	Gaurav Arora, et al. [18]	Employed users behaviour	Doesn't consider the out of box entries in the database.
19.	Hybrid approach of Collaborative Filtering and Content-based methods	V. Adi Lakshmi, et al.[19]	Increased accuracy levels	Even after 5000 iterations the graph was not converging
20.	Collaborative filtering approach	Debadrita Roy, et al. [20]	Include users behaviour along with movies	Addition of new data bias the trained system
21.	Personalized recommendation system (PRS) and Support vector machine (SVM) classification	Amarjit kundu,et al.[21]	Extracted the un-important features of the data	Use PSO based local optimisation for this purpose which has problem of premature termination
22.	Swarm mining	Sajal Halder, et al. [22]	Help producers to plan a movie	Unable to find the group of a user who likes multiple movies depending on the movie genre.
23.	Classical Vector Space Model (CVSM)	Roberto Mirizzi, et al. [23]	Developed face-book application which link movie recommendation with user's preference	No validation was available
24.	Content correlations	Sang-Min Choi, et al. [24]	Focused on the cold start problem	NA

