An Improved Approach for Movie Recommendation System

Shreya Agrawal (ME Student)
CSE
SVITS
Indore, MP, India
shreyaagrawal24@gmail.com

Pooja Jain (Assistant Professor)
CSE
SVITS
Indore, MP, India
poojacs35@gmail.com

***Abstract*- In this hustling world, entertainment is a necessity for each one of us to refresh our mood and energy. Entertainment regains our confidence for work and we can work more enthusiastically. For revitalizing ourselves, we can listen to our preferred music or can watch movies of our choice. For watching favorable movies online we can utilize movie recommendation systems, which are more reliable, since searching of preferred movies will require more and more time which one cannot afford to waste. In this paper, to improve the quality of a movie recommendation system, a Hybrid approach by combining content based filtering and collaborative filtering, using Support Vector Machine as a classifier and genetic algorithm is presented in the proposed methodology and comparative results have been shown which depicts that the proposed approach shows an improvement in the accuracy, quality and scalability of the movie recommendation system than the pure approaches in three different datasets. Hybrid approach helps to get the advantages from both the approaches as well as tries to eliminate the drawbacks of both methods.**

Keywords- Movie Recommendation System, Hybrid Approach, Support Vector Machine, Genetic Algorithm, Scalability, Quality

I. INTRODUCTION

Movies are a part and parcel of life. There are different types of movies like some for entertainment, some for educational purposes, some are animated movies for children, and some are horror movies or action films. Movies can be easily differentiated through their genres like comedy, thriller, animation, action etc. Other way to distinguish among movies can be either by releasing year, language, director or by cast etc.

For watching movies online, there are a number of movies to search in our most liked movies [6]. Movie Recommendation Systems helps us to search our preferred movies among all of these different types of movies and hence reduce the trouble of spending a lot of time searching our favorable movies. So, it requires that the movie recommendation system should be very reliable and should provide us with the recommendation of movies which are exactly same or most matched with our preferences. There are many movie recommendation systems which do this work for us reliably but since there are a lot of movies and huge number of users online and also both are increasing day by day, due to which there is a compromise in the quality of the movie recommendation systems.

In the last few years, with the day by day increase in the social networking mania on the internet, lots and lots of information is getting generated on the internet. This increase in the amount of the data on the internet and also the flow of data on the web at a brisk speed makes it problematic for the users to manage the information adequate tools. To eradicate the overload of the data, recommendation system is used as information filtering [5] tool in social networking sites [1]. Hence, there is a huge scope of exploration in this field for improving scalability, accuracy and quality of movie recommendation systems.

Movie Recommendation system is very powerful and important system. But, due to the problems associated with pure collaborative approach, movie recommendation systems also suffers with poor recommendation quality and scalability issues.

II. RELATED WORK

Hirdesh Shivhare et al. [2015] proposed an integrative method by merging fuzzy c-means clustering method and genetic algorithm based weighted similarity measure to construct a movie recommendation system. The proposed movie recommendation system gives finer similarity metrics and quality than the existing Movie recommendation system but the computation time which is taken by the proposed recommendation system is more than the existing recommendation system. This problem can be fixed by taking the clustered data points as an input dataset [1].

Gaurangi Tilak et al. [2008] introduced MovieGEN, an expert system for movie recommendation. They implemented the system using machine learning and cluster analysis on the basis of hybrid recommendation method. Based on the Support Vector Machine prediction it selects movies from the dataset, clusters the movies and develops questions to the users. Based on the user`s answers, it refines the movie set and finally recommends movies to the users [12].

III. TECHNIQUES USED IN PROPOSED METHODOLOGY

The proposed solution is for improving the scalability and quality of the movie recommendation system .We use a

Hybrid approach [9], by unifying Content-Based Filtering and Collaborative Filtering, so that the approaches can be profited from each other. For computing similarity between the different movies in the given dataset efficiently and in least time and to reduce computation time of the movie recommender engine we used cosine similarity measure.

A. Content-Based Filtering

For the implementation of a content-based filtering system following steps to be done:

- Terms Allocation
- Terms Representation
- Learning Algorithm Selection
- Provide Recommendations

B. Collaborative Filtering

a) Simple Support Vector Machine Algorithm [15]

It is an iterative algorithm. Steps of a simple Support Vector Machine algorithm:

- Setting up the Training Data
- Setting up the SVM`s Parameters
- Training the SVM
- Regions are classified by SVM
- Obtaining Information about the support Vectors

b) Adjusted K-Means Algorithm

In this algorithm, in input the numbers of clusters and items attribute features are given. For this input, we get output as a set of cluster which reduces the squared error. Also, the probabilities of each movie belonging to each cluster center are represented. Then arbitrarily some objects are selected as initial cluster centers. On the basis of mean value of objects each object is reallocated to the most matching cluster. Then cluster means are updated and the possibility between objects and each cluster centre is computed. This process is repeated till all the centroids no longer change [2].

C. Genetic Algorithm

The genetic algorithm is applied for optimization [4]. The steps of Genetic algorithm are as follows:

- Initializing Population
- Fitness Evaluation
- Generating new population
- Selection of two fit parent chromosomes
- Cross over of best parent chromosomes
- Mutation
- Generated new child should be assigned in the new population.
- The newly generated population should be utilized for the furthermore run of the algorithm.
- Repeated the algorithm until it meets the termination condition.

Return the optimal solution in current population after the end condition is satisfied and is stopped [4].

D. Cosine Similarity Measure

The cosine similarity for two vectors is a measure that calculates the cosine of the angle between them [16]. This metric measures the orientation and not magnitude. The cosine similarity formula is given as:

$$\vec{a} \cdot \vec{b} = \|\vec{a}\|\|\vec{b}\| \cos\theta$$

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\|\vec{b}\|}$$

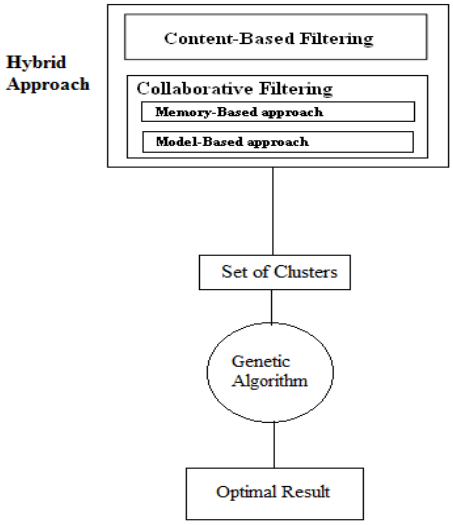


Fig 1. Hybrid Approach with Genetic Algorithm

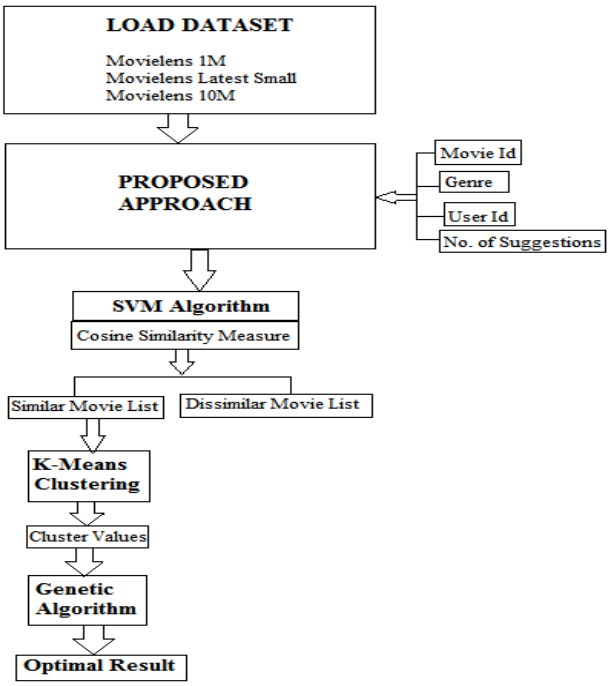


Fig 2. Proposed Approach

IV. PROPOSED METHODOLOGY

A. Dataset

We have used three different data sets available in Movie Lens, which is generated by the group lens research team for the research work in the field of recommender system, to help developers to evaluate their recommendation systems. These are:

1. Movie Lens 1M dataset
2. Movie Lens Latest Small dataset
3. Movie Lens 10M dataset

The Movie Lens datasets [3] has the following features:

1. Ratings are assigned from 1 to 5 (1 means very bad, 5 means very good)
2. Each user has rated at least 20 movies
3. Simple demographic information of the users (age, gender, occupation, zip) is provided.

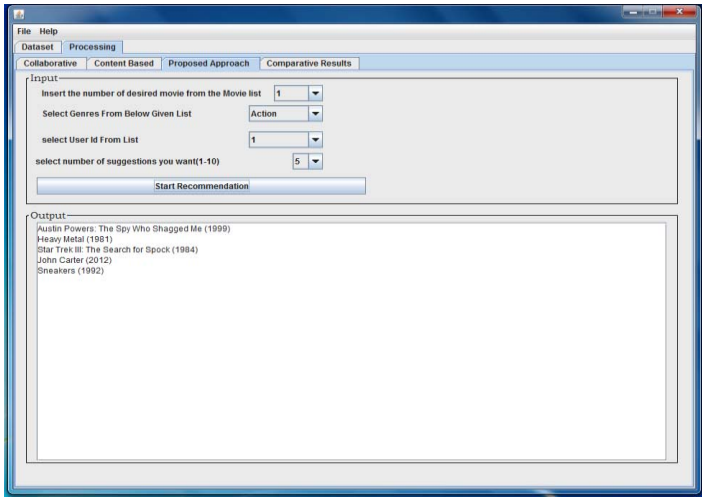
The reason for considering three different Movie lens Datasets of different sizes is to check the scalability parameter, that is, when our hybrid approach is applied to three different movie lens datasets of different sizes the system should continue to work well and give good performance, if the system is scalable. Results depicts that the movie recommendation system with our hybrid approach is scalable.

B. Recommendation system quality measures

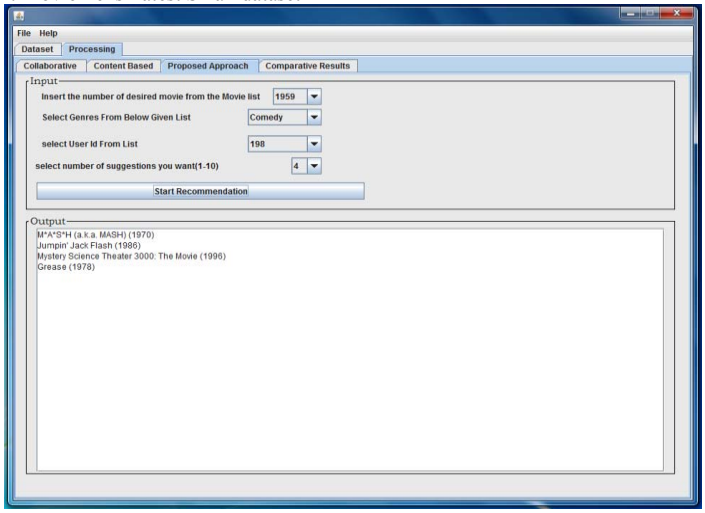
The typical quality measures such as Mean Absolute Error, precision, recall, f-measure and coverage are used for analyzing the quality [1] of our proposed movie recommendation system and the results will be compared with the results of the existing pure content based or collaborative filtering movie recommendation system which are also have been implemented by us.

V. SCREEN SHOTS

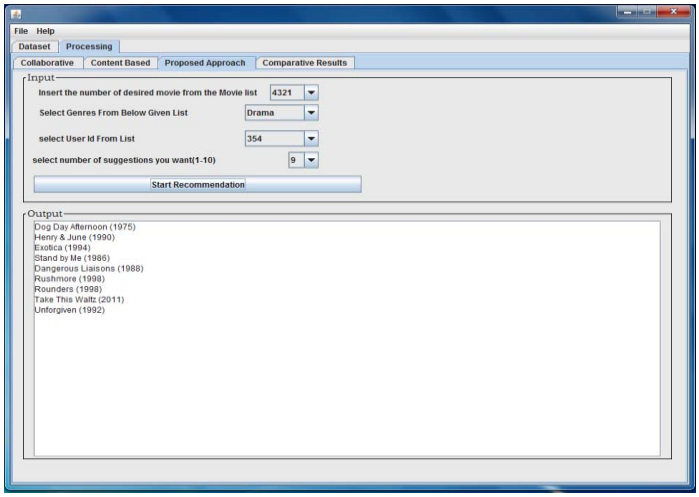
1. Movie Lens 1M dataset



2. Movie Lens Latest Small dataset



3. Movie Lens 10M dataset



VI. EXPERIMENT AND RESULT ANALYSIS

In this section the different experiments performed will be shown to prove that our approach of developing a movie recommendation system using Hybrid approach and genetic algorithm is better and provides better performance than the existing pure content based or collaborative filtering movie recommendation system in terms of accuracy, quality, scalability and computing time.

Accuracy: Degree to which the result of a calculation or measurement fulfills to the precise value or a standard.

Scalability: System should continue to work efficiently, when there is an increase in the number of movies.

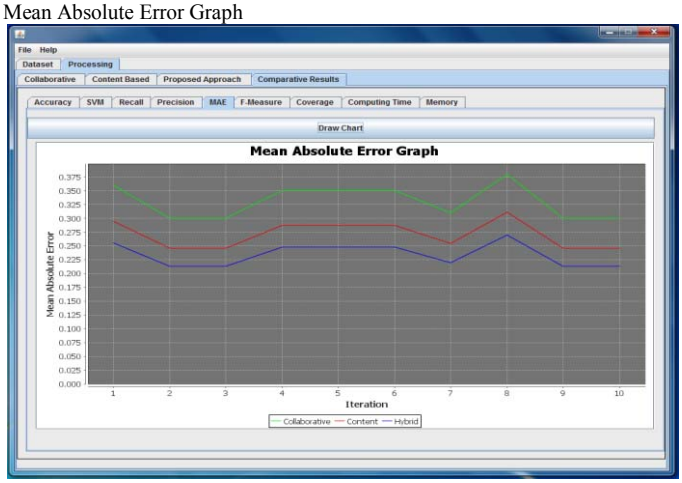
Computing Time: Time required by a particular algorithm to provide movie recommendations.

Following parameters have been considered for quality analysis:

1. Mean Absolute Error: Average variation in the forecasted rating versus the true rating
2. Coverage: Domain of movies over which the system can make recommendations
3. Recall: Proportion of good movie recommendations that appear in top recommendations
4. Precision: Proportion of movie recommendations that are good recommendations
5. F-measure: Harmonic mean of precision and recall

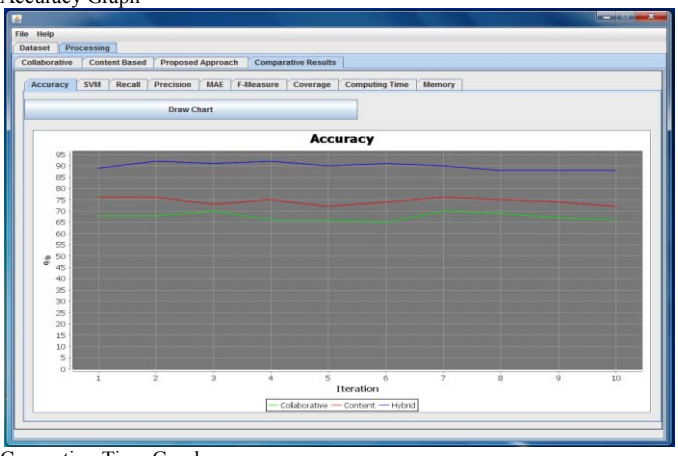
1. Movie Lens 1M dataset

Mean Absolute Error Graph

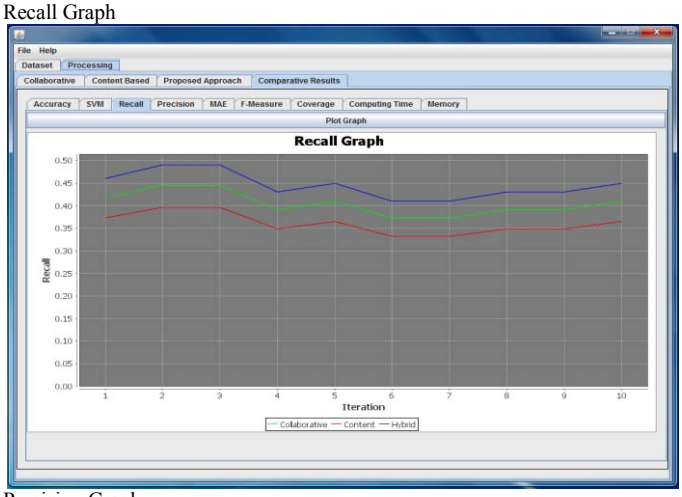


Coverage Graph

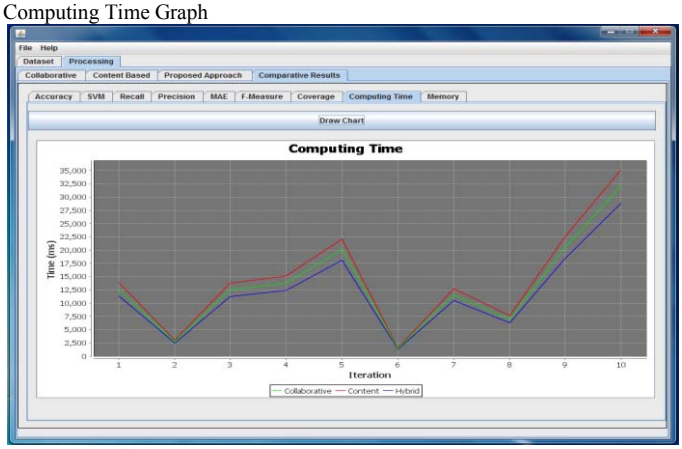
Accuracy Graph



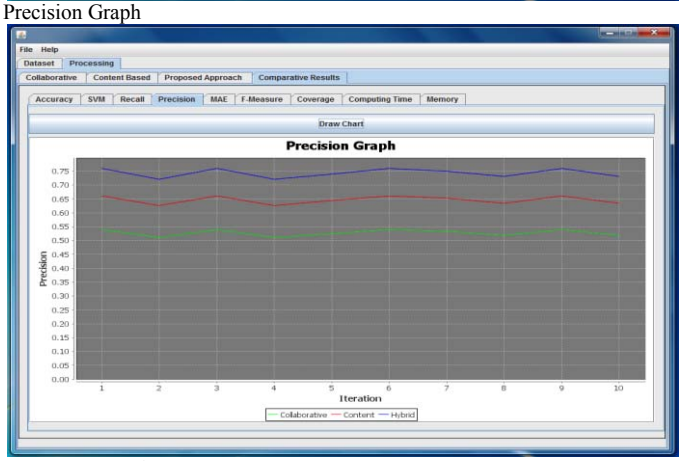
Recall Graph



Computing Time Graph

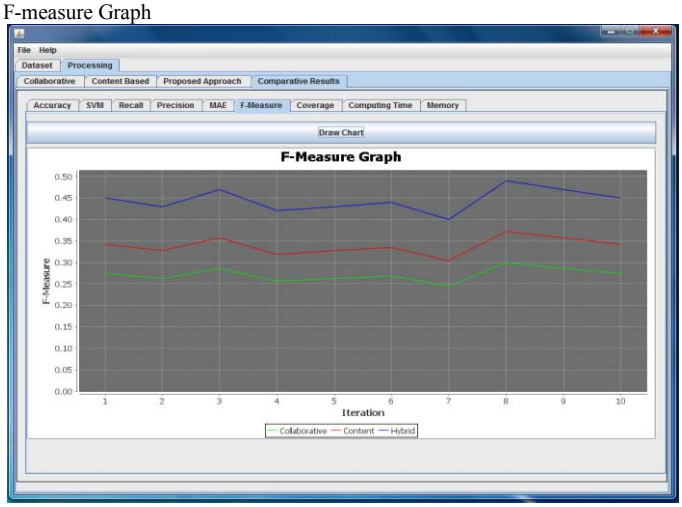


Precision Graph



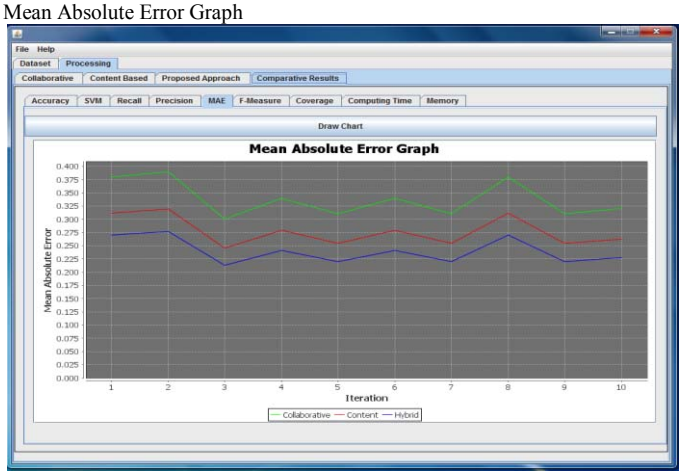
Memory Graph

F-measure Graph

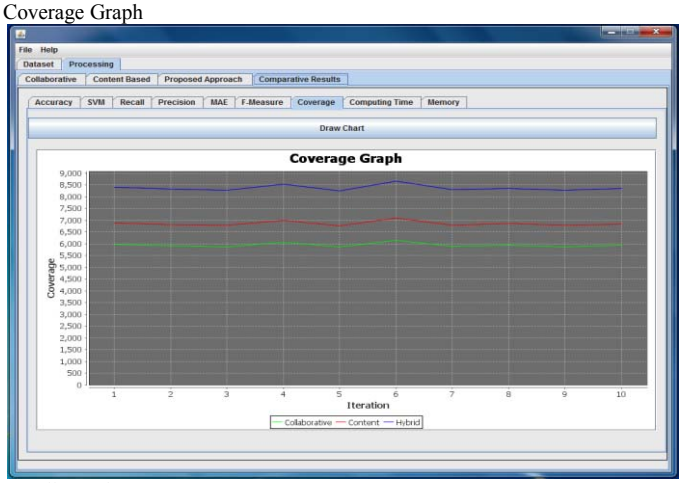


2. Movie Lens Latest Small dataset

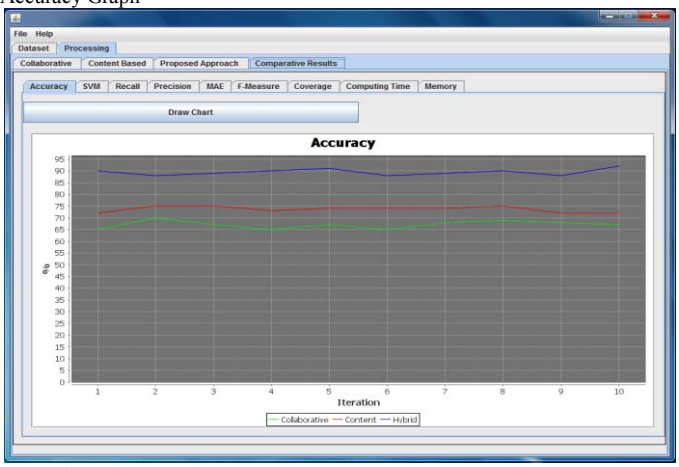
Mean Absolute Error Graph



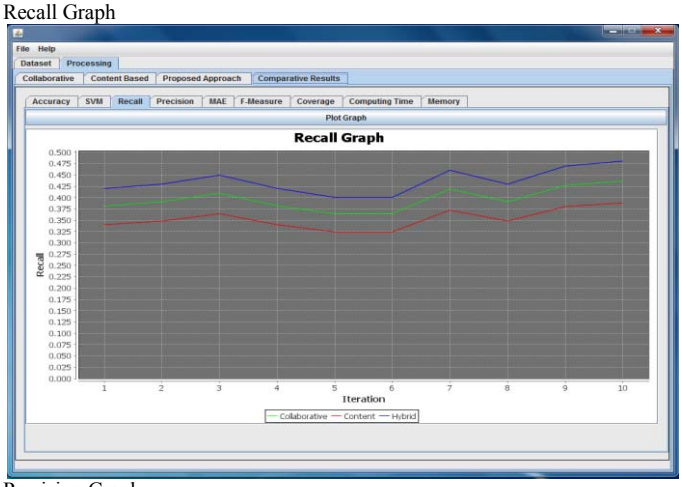
Coverage Graph



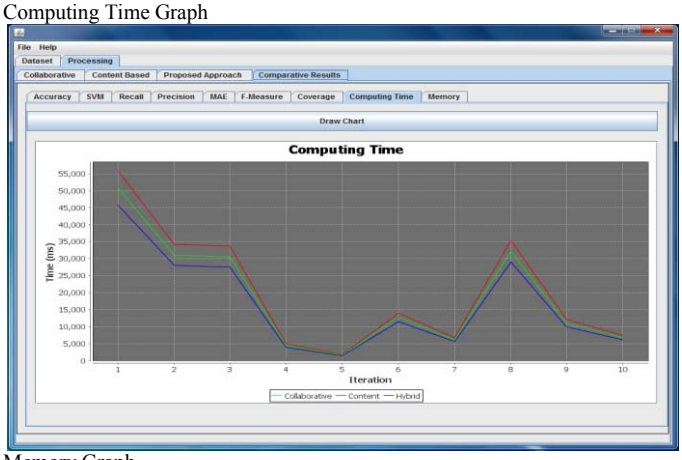
Accuracy Graph



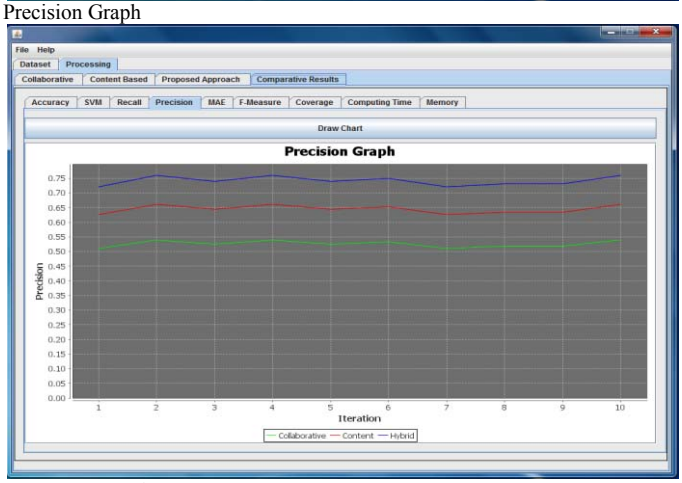
Recall Graph



Computing Time Graph



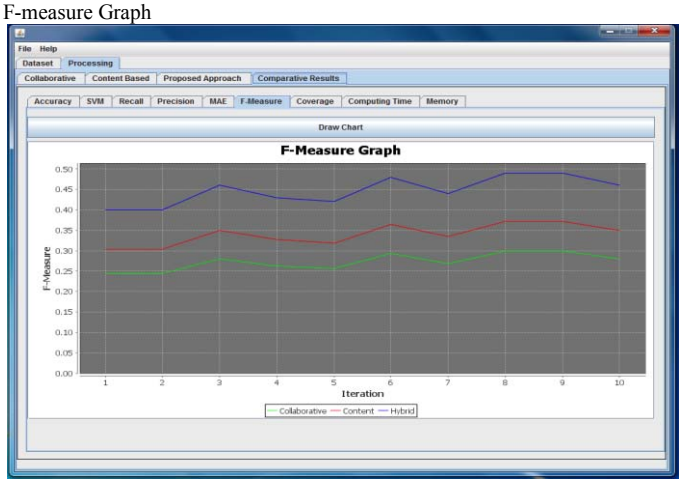
Precision Graph



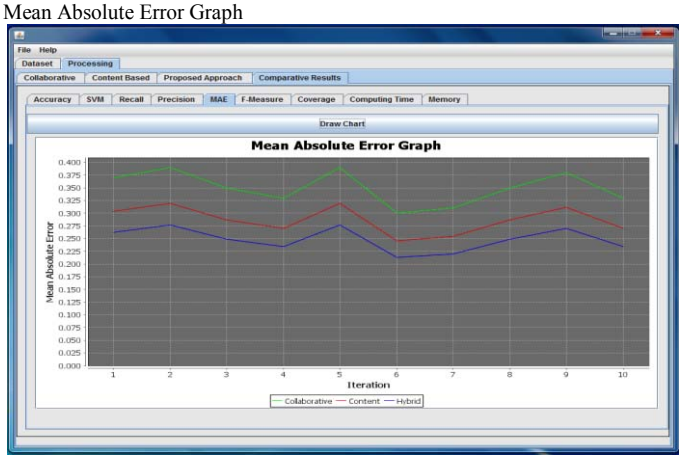
Memory Graph

3. Movie Lens 10M dataset

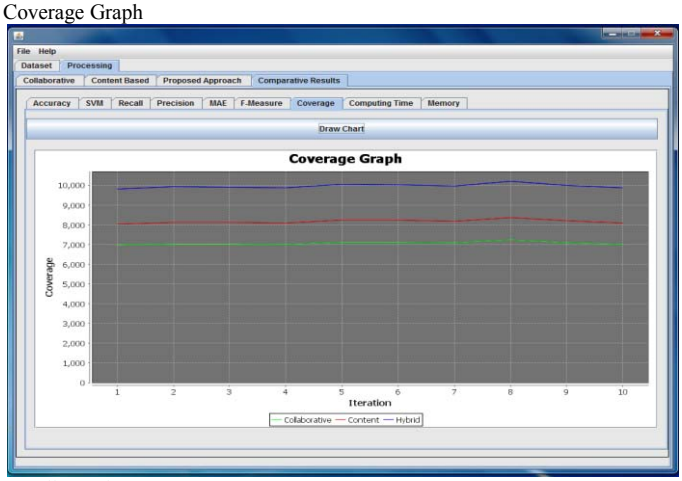
F-measure Graph



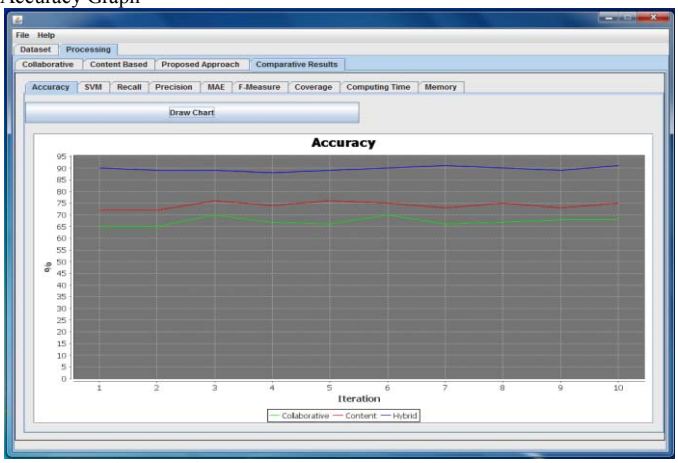
Mean Absolute Error Graph



Coverage Graph

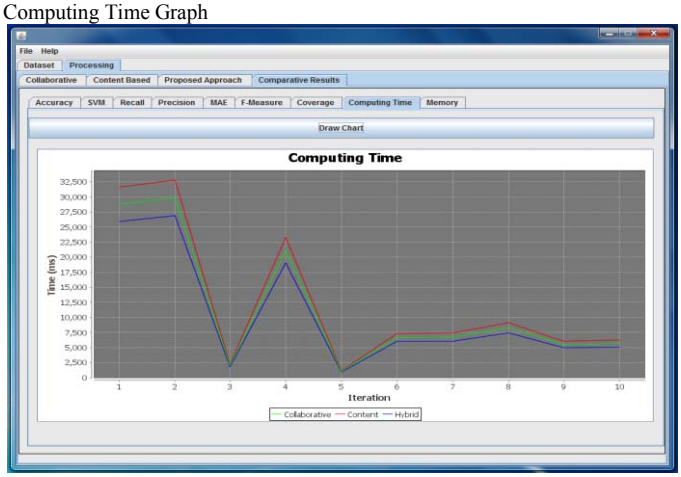


Accuracy Graph

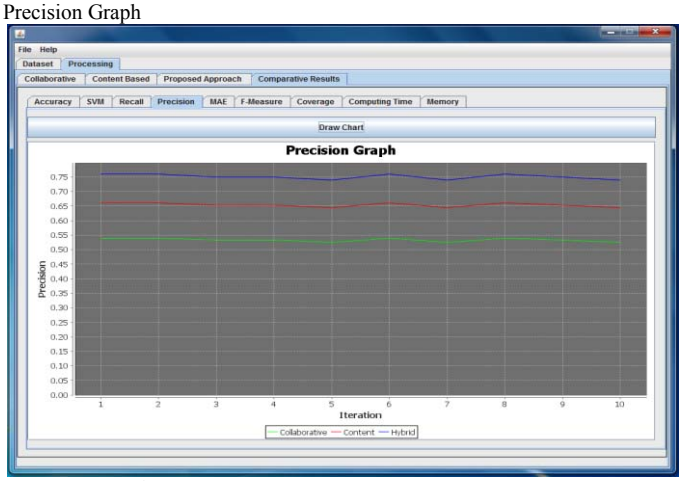


Recall Graph

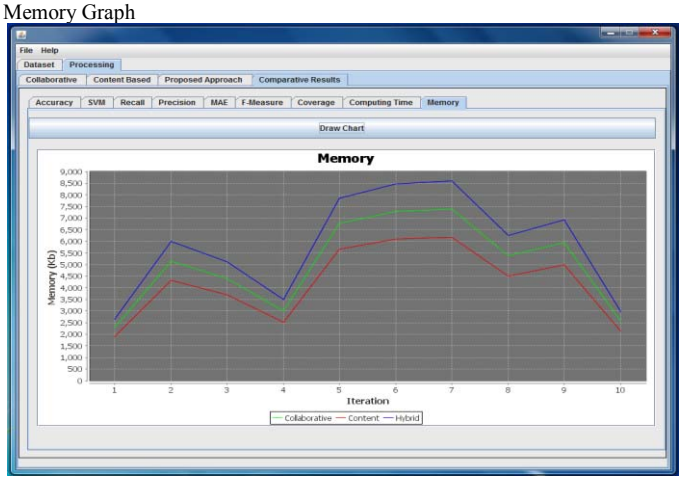
Computing Time Graph



Precision Graph



Memory Graph



F-measure Graph

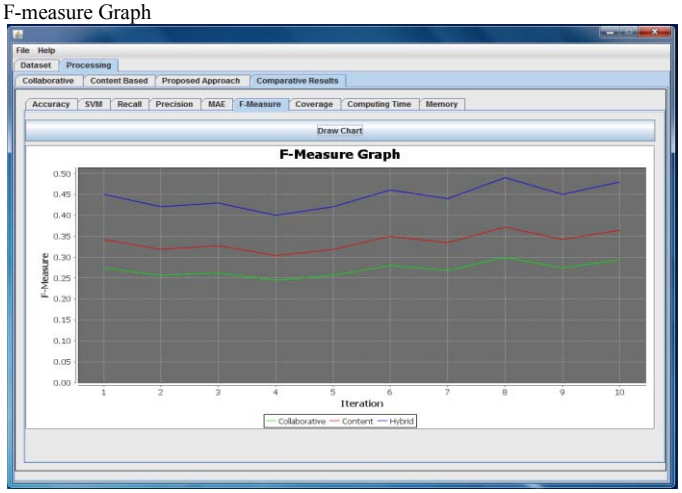


TABLE I. Results

PARAMETERS	COLLABORATIVE APPROACH	CONTENT BASED APPROACH	PROPOSED APPROACH
Accuracy	Low	Average	High
Quality	Low	Average	High
Scalability	Less	Average	High
Computing Time	Average	High	Low
Memory	Average	Low	High

Memory: Total amount of computer memory required by an algorithm to complete its execution.

The result is that when the proposed methodology is individually implemented on three different Movie lens datasets it increases accuracy, scalability and quality of the movie recommendation system as compared to the implementation of pure content based approach or pure collaborative approach individually on three different Movie lens datasets. The computing time of the proposed approach is lesser but the memory requirement of the proposed approach is more than the existing pure content based approach or pure collaborative approach.

VII. CONCLUSION

In this paper, to improve the accuracy, quality and scalability of movie recommendation system, a Hybrid approach by unifying content based filtering and collaborative filtering; using Support Vector Machine as a classifier and genetic algorithm is presented in the proposed methodology. Existing pure approaches and proposed hybrid approach is implemented on three different Movie lens datasets and the results are compared among them. Comparative results depicts that the proposed approach shows an improvement in the accuracy, quality and scalability of the movie recommendation system than the pure approaches. Also, computing time of the proposed approach is lesser than the other two pure approaches.

VIII. FUTURE WORK

The research work was pretty successful and satisfactorily completed. Following future work can be done:

In this hybrid approach, we have considered Genres of movies but, in future we can also consider age of user as according to the age movie preferences also changes, like for example, during our childhood we like animated movies more as compared to other movies. There is a need to work on the memory requirements of the proposed approach in the future. The proposed approach has been implemented here on different movie lens datasets only. It can also be implemented on the Film Affinity and Netflix datasets and the performance can be computed in the future.

REFERENCES

[1] Hirdesh Shivhare, Anshul Gupta and Shalki Sharma (2015), "Recommender system using fuzzy c-means clustering and genetic algorithm based weighted similarity measure", IEEE International Conference on Computer, Communication and Control.

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

[3] RyuRi Kim, Ye Jeong Kwak, HyeonJeong Mo, Mucheol Kim, Seungmin Rho,Ka Lok Man, Woon Kian Chong (2015),"Trustworthy Movie Recommender System with Correct Assessment and Emotion Evaluation", Proceedings of the International MultiConference of Engineers and Computer Scientists Vol II.

[4] Zan Wang, Xue Yu*, Nan Feng, Zhenhua Wang (2014), "An Improved Collaborative Movie Recommendation System using Computational Intelligence",Journal of Visual Languages & Computing,Volume 25, Issue 6.

[5] Debadrita Roy, Arnab Kundu, (2013), "Design of Movie Recommendation System by Means of Collaborative Filtering", International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 4.

[6] Sajal Halder, A. M. Jehad Sarkar, Young-Koo Lee (2012), "Movie Recommendation System Based on Movie Swarm", Second International Conference on Cloud and Green Computing.

[7] Birtolo, C., Ronca, D., Armenise, R., &Ascione, M. (2011), "Personalized suggestions by means of collaborative filtering: A comparison of two different model-based techniques", In NaBIC, IEEE (pp. 444–450).

[8] Jesus Bobadilla, Fernando Ortega, Antonio Hernando and Javier Alcala (2011) "Improving collaborative filtering recommender system results and performance using genetic algorithms", ELSEVIER Knowledge-Based Systems, Vol. 24, issue 8, pp. 1310–1316.

[9] Mustansar Ali Ghazanfar and Adam Prugel-Bennett (2010)," Building Switching Hybrid Recommender System Using Machine Learning Classifiers and Collaborative Filtering", IAENG International Journal of Computer Science, 37:3, IJCS_37_3_09.

[10] J. Bobadilla, F. Serradilla and J. Bernal (2010) "A new collaborative filtering metric that improves the behavior of recommender systems", ELSEVIER Knowledge-Based Systems, Vol. 23, Issue 6, pp. 520- 528.

[11] J.L. Sanchez, F. Serradilla, E. Martinez, J. Bobadilla, (2008) "Choice of metrics used in collaborative filtering and their impact on recommender systems", in: Proceedings of the IEEE International Conference on Digital Ecosystems and Technologies DEST, pp. 432–436.

[12] Eyrun A. Eyjolfsdottir, Gaurangi Tilak, Nan Li (2008), "MovieGEN: A Movie Recommendation System", 2008 Conference Proceedings.

[13] J. Ben Schafer, ShiladSen, Dan Frankowski, Jon Herlocker, (2007) "Collaborative filtering recommender systems" In The Adaptive Web, pp. 291–324. Springer Berlin / Heidelberg.

[14] F. Kong, X. Sun, S. ye, (2005), "A comparison of several algorithms for collaborative filtering in startup stage", in: Proceedings of the IEEE Networking, Sensing and Control, pp. 25–28.

[15] S.V.N. Vishwanathan, M. Narasimha Murty (2002), "SSVM: A Simple SVM Algorithm", Neural Networks, IJCNN '02. Proceedings of the International Joint Conference on (Volume: 3)

[16] Joseph Konstan, George Karypis, BadrulSarwar, and John Riedl, (2001) "Item-Based Collaborative Filtering Recommendation Algorithms", ACM.