International Journal of Advanced Computer Technology (IJACT)

ISSN:2319-7900

A SURVEY ON MULTIMEDIA HYBRID AND COLLABORATIVE RECOMMENDATION BASED ON CLOUD

Reshma.M.R#1., Dr.S. Neduncheliyan#2., and Kalaivani K#3.,

ABSTRACT: *Internet user share large number of video clips on video sharing and social networking websites. Online videos are shared mostly in popular websites such as Youtube and other sites. Generally, more than expected time is spent by the users, to search and identify the contents. There are many filtering algorithms has been proposed, among them Collaborative filtering (CF) is most widely used and almost popular. In this paper, we propose comparison of various filtering algorithms with that of a hybrid filtering and collaborative filtering. Although Collaborative Filtering is used commonly, there are also some problems such as cold start and sparsity of data. Empirical studies show that these algorithms outperform much similarity and with hybrid recommender cold start can be overcome to a limit. Thus Hybrid with Collaborative recommendation will provide high precious and reduce delayed search in multimedia recommendations.*

Keywords: Collaborative filtering, Recommendation system, Hybrid Recommendation, cold start

1. INTRODUCTION:

We have seen the rise in technology and how information spread across the network. According to latest generation, internet has become the biggest portal for search and share of information. As the number of users grows day by day, they share large number of video clips on video sharing and social networking. Online videos are shared mostly in popular websites such as Youtube, facebook. Youtube provides one of the leading Service in video search.

The recent survey conducted, shows that most of the consumer internet traffic will be video by the end of year 2020. Videos will continue to be approximately 90% of global consumer traffic. There are always 24 * 7 on-line accessibility and millions of customer access has resulted large variety of multimedia choices, due to this customers are faced with information overload. The contents in web are numerous so there is always chance of duplication, similar, related, or may be quite different. Thus users generally waste a lot of time [2], just by searching and the same goes for people using mobiles to access internet. The user's own arduous efforts are necessary for retrieving the information required. To help users to obtain their desired content lists from billions of web pages in a short time is very challenging. Due to this conditions some video sharing websites provide users with search engine to in order to search their desired videos quickly that match their preference. Even search engines have their limitations; they are not much accurate in providing proper recommendations and consume lot of time.

Thus the recommendation systems have been introduced for providing a way for the users to get their required information. An important technique which is the most likely influenced machine learning method. They are different from search engines in sense, recommender systems only goal is to search and list those recommendations. They recommended data from huge databases and not from countable number of items. Cloud computing, well is a storage hub of much data, so it solve the problem of mass data storage. There are many services or sites, in which recommendation is provided not only to videos, but also that attempt to recommend books, movies, stories applications as such in play stores to user based on their past recorded data. The internet users try to retrieve and identify unknown items and preferences that are of their interest. As we know users waste a lot of time to obtain their interests in downloading and watching those video. For this purpose, there are several techniques in video recommendations and other document filtering has been proposed and most of them are successful. For example, considering one of the global leaders of internet services, i.e Google. It mostly does keyword related search, has adopted content-based filtering (CB) recommender system. It is included in Adwords service but this has been removed by the company due to lack of interest of user. While another leading company Amazon and Taobao have achieved great success in recent years by using collaborative recommender systems into their e-commerce websites to help users find their interested goods.

Filtering is a process or tools to help mine the essential details and unnecessary information are hided to a limit. People search their required information out of vast ocean of data. Filters provide an helping hand for users, so that the their valuable time will not be wasted in reading or viewing the entire sections. Filters as per the name indicates, it generally structures information and later organizes those information. Filters also make use of internet search engine results to immediately get what the group of users would mostly like to be suggested. The gains through using better filtering

ISSN:2319-7900

techniques are enormous. Mostly recommendation is based on search history, access preferences and ratings while their results of searches given are not very accurate and not satisfied by the users. The most important and literally used approaches for recommender systems can be classified into; collaborative filtering techniques, content-based filtering and hybrid filtering. As it is stated that collaborative filtering (CF) is used widely and in almost all recommendation system, that uses known user's preferences to recommend products and kind of popular compared to other techniques. Even though many recommendations have been proposed, one prevalent goal in recommender systems research shown is, to achieve high performance; it is obvious to combine few recommendation techniques.

All of the known recommendation techniques, which are in use currently, have their respective strengths and weaknesses. This paper describes the similarities of collaborative filtering and hybrid model for multimedia recommendation. To overcome some of the limitations in CF technique, such as data sparsity and cold start and to achieve better recommendation and to reduce the delay in search of the contents required especially in cloud environment. In future, the rise of technology is undeniable, so it is expected recommendation hardware and software go on to next stage. It must include more user-friendly functions in filtering and must be advanced further.

2. BACKGROUND

The recommender systems are classified into the following types, which take account on ratings, user preferences, history and other factors:

- Content-based recommender systems: In this recommendation system prediction is done based on content information which is stored in the system about each item to be recommended. It used to recommend with similarity of items of the user had previously preferred, based on similarity with respect to user preferences and similarity of certain items are to each other. The system pays attention to two type of information on creation of user profile, such as model of user's preferences and history of their interactions. Issues in content-based filtering [1] is learning as a pressurizing, focus on probability of the system, is able to learn user preferences from user's actions and use them across other content types. The user profile is learnt regarding contextual independence, so a constraint satisfaction problem is considered.
- Collaborative filtering systems: This filtering algorithm tries to identify groups of people or users with identical tastes to those of the user and recommend contents that they have liked. This also considered collecting users' behaviors or preference. The prediction is given on what users will like base on their similarity to other users. Depending on the machine analyzable content is one of the demerits of filtering techniques but the collaborative filtering approach does not rely on that information and so it is an important advantage of CF system. Therefore, the accuracy of prediction increases with capability to recommend complex items.
- Hybrid recommender systems: The hybrid is platform is a cross breed. This model is an approach of combining collaborative filtering with content-based filtering. This method can be more effective in some cases where recommendation will be precious. The implementation of this approach can be made either by merging content-based or collaborative based recommendation with hybrid separately or unifying both the approach together with hybrid systems as one model. The gain of performance is the effect of this approach. Netflix a social network is a good example of hybrid systems.

From the above methods, as we know each systems have their own merits and demerits. Most of approaches of recommender system problem have been deployed, using methods from machine learning capabilities, approximation theory, and various heuristics algorithms. The recommendation systems are used based on how the prediction results are obtained, independently of the techniques used. The only focus is to provide a dynamic system, to help the viewers in this competitive environment. These approaches are generally implemented for cloud environment, these are cloud based recommenders. The general recommendation diagram is show below in figure 1.

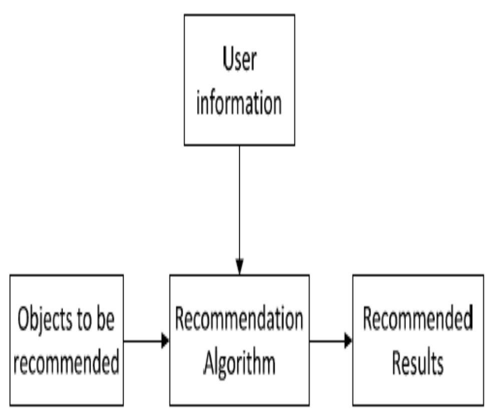

Fig.1 General Recommendation system

The recommendation algorithms used can be collaborative systems that make use of ratings to recommend videos and identify similar tastes of users or it can be content-based recommender systems that rely on content information in order to prediction of similar items else use hybrid model. Generally, on comparing collaborative systems and content-based approaches, a better report of performance is obtained from collaborative. But this successful system relies on few recorded ratings which pay way for drawbacks such as cold-start problem, which is eventually part of data sparsity problem and ramp-up, all are somehow related with each other. A

ISSN:2319-7900

common approach to solve the problems of the above techniques is to combine both content-based and collaborative information into a hybrid recommender system. But we suggest it is evident to combine hybrid and collaborative filtering as a single model, instead of content-based which provide much complexity.

3. COLLABORATIVE FILTERING

One of most used filtering is collaborative filtering and its recommendation is probably the most known and one of the successful technologies, which is widely implemented. This filtering have been used in many technique has been used in financial sources, and web applications. It is also used in sensing and monitoring process. Basically filtering depends on the following two processes mostly, these as well the parts for completion of filtration process in CF: The two parts of recommendation are prediction and recommendation. Prediction is numerical value, predicted based on likeliness of a video that does not belong to same user, but by another user, the prediction value may be in same scale of this value. While recommendation is basically a list of items that the active user will be provided on search. This recommended list must be on items, which user likes to view and to be recommended so information retrieval is ease by the active user.

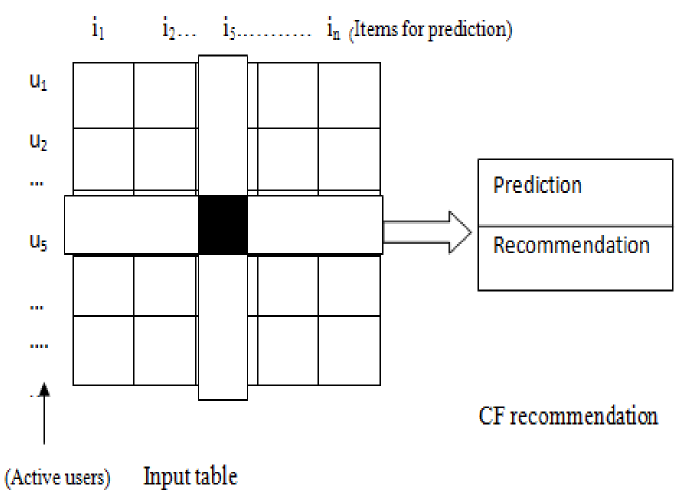

Fig.2 Collaborative filtering approach

As shown in Fig. 2. Given Rating table consists of u_1, u_2,..u_n are user of the system and i_1, i_2,….i_n is a list of items for which prediction is required. This rating table is considered as ratings input table, which given as input. Prediction is numerical value denotes the predicted on item i_5 for an active user. This predicted value is in same scale as opinion values provided by user u_5. So, finally items for the users can be recommended.

The CF algorithms are further classified as memory-based CF and model-based CF. Memory-based algorithms operate over the whole of user database to make predictions. The most common used memory based models are using several distance measures on the notion of nearest neighbors. It uses neighbourhood-based algorithms. In memory-based techniques, there are two phases, they are Pearson correlation and cosine similarity.

Below given equation [7], [11] is used in memory based CF algorithms

$$r_{u,i} = \operatorname{aggr}_{u' \in U} r_{u',i}$$

where $r_{u,i}$ = rating of user 'u' on item 'i',

'U is the set of top users. Therefore,

$$r_{u,i} = \bar{r_u} + k \sum_{u' \in U} \operatorname{simil}(u, u')(r_{u',i} - \bar{r_{u'}})$$

Thus the k stands for (k = Average rating of user u, it is also normalizing factor.)

$$k = 1 / \sum_{u' \in U} |\operatorname{simil}(u, u')|$$

Memory Based Collaborative Filtering have the following types, they are

- User Based Collaborative Filtering: It is where similarity measures are in use and user preference predictions are made in terms of these measures.
- Item Based Collaborative Filtering: It is also called as item-item filtering. Recommendation is generated where a set of items is used as nearest neighbours, if the algorithm computes similarity of different items [5]. It is used in youtube mostly.

Model-based systems are compact model inferred from the informational data. The model-based CF techniques into several categories: content-boosted CF algorithm, Clustering and Bayesian CF algorithms.

Recommendation algorithm should be effective and efficient. In collaborative filter major problems faced during result generations are sparseness of the data and since large number of users and content increases over speed, this lead to scalability issues. In traditional Pearson correlation-based algorithm which is a pure memory-based algorithm, affects to scalability issues [7]. Implementation of this type of algorithms is simple and effective, but recommendations are less accurate or precious due to sparsity of data. Model-based CF algorithms are also affected by sparse data and cold start problem.

ISSN:2319-7900

As we provide the survey for cloud environment, the collaborative filtering works with context aware to form a cloud based algorithms [2]. These are the following steps;

1) User behavior collecting: User behavior and interest changes with respect to time, location and network. The entire search is done to satisfy the users, so there is a need to consider their preferences. User preferences are based on the activity of the users such as access patterns, keywords used by the users and tags. With increase in context as the number of user increases, resources are consumed quickly and high dimensional contexts are very challenging for recommendation algorithms. In these criteria, clusters are used in cloud based recommendations. Every time when a new search is being made, clusters are restarted periodically at regular intervals.

2) Content clustering: From the above behavior collecting, it is clear that user content clusters are formed. User's social connections exhibit their profile in each social websites. With the help of six-attribute tuples, which involve tags, resolutions are all mapped into it., then the user content clustering algorithm is executed on the tuple to retrieve user interest clusters and user content similarity.

3) Recommendation rules: A rule is format or order need to be followed. In a large wide spread growth of videos, a certain pattern must be followed in order to avoid to collapse. Since large number of users involved, the space also increases respectively. The system becomes unscalable and brings latency to the search recommendation lists. So it is necessary to extract certain rules for resolving the above issue.

4) Real-time recommendation: It is recommendations, in which user request are converted into rules [2]. This is done based on keywords, implicit contents and search for their favorite contents or items. This translation guarantees to provide real-time response.

4. CHALLENGES OF COLLABORATIVE FILTERING:

A. Variable user's interest: Traditionally most of the algorithms give equal importance for all ratings. As trends changes and different users have different interest. The shift in user's interest with time is known as concept drift. Users have different ratings for each media, they came across. The item-based Collaborative filtering [4] algorithm doesn't take into consideration of the changing user interest with respect to time and importance must be given to their current interest more than old ratings.

B. Data Sparsity in Collaborative Filtering: One of the major impacts on the working of a collaborative filtering approach is data sparsity. This issue makes it is almost impossible to define similarity between two users, this renders collaborative filtering useless. There are wide numbers of websites, which consists of mass data storage of contents. The items in databases comparatively vast and user rate just few items. Even group of individuals that are very active may show high intend to view the contents than rate it, sometimes even very popular content have been rated less. This is a serious issue, which affects the entire process of recommendation.

C. Cold-start: Generally cold start is also a major issue in collaborative filtering approaches that depends mostly ratings. It is a problem, which emphasizes the importance of sparsity problem. Cold-start refers to the situation in which an item cannot be recommended to a user unless it has been rated by a sufficient number of users [3]. This problem applies to new and obscure item. It particularly refers to user's different tastes of requirements. Sometimes, a new user has to rate a competent number of items before the recommendation algorithm be able to provide reliable and almost accurate recommendations.

5. HYBRID FILTERING:

Hybrid is a platform of different combinations. As the name implies, it is an approach of combining multiple filtering together, to improve performance to some aspects. Researchers are done many studies to show that a hybrid approach is nothing but a combined filtering approach, which include collaborative filtering and content-based filtering, to gain performance by avoiding defects of other filters and attempt to reduce ramp-up problems. All learning-based techniques such as collaborative and content-based suffer from ramp-up problem (a troublesome experience while changing user preferences, after setting profile of that particular user). Several studies are conducted to provide results that hybrid model also helps to reduce cold start and data sparsity. Therefore the performance of the hybrid compared with the other methods(pure CF and Content-based) and demonstrate that the hybrid methods can provide more accuracy in recommendations than pure filtering approaches, in order to generate list of contents such as text or multimedia. The seven hybridization techniques [11] are as follows but some are still unexploited;

- Weighted: It is based on score. A Weighted hybrid recommender is one in which the score or vote of several recommendation techniques are taken and combined numerically, in a sense to produce a single recommendation.
- Switching: A switching hybrid uses switching method, between recommendations based on certain criteria.It chooses recommendation components and also includes additional complexity.
- Mixed: The technique in which recommendations from various recommenders are given at same time. It used to make large number of recommendations simultaneously. As name implies, it is a mixture of one or more techniques.

ISSN:2319-7900

- Feature Combination: In this, features obtained from different knowledge sources are thrown together and given to a single recommendation algorithm. The feature combination hybrid considers collaborative data but does not rely on it, so the sensitivity of the system can be reduced.
- Feature Augmentation: The technique is employed to produce a rating of an item and also classify it; with this information, the next step is to process the recommendation.
- Cascade: It is a staged process. Cascade involves two steps, first one recommendation technique is used to generate a coarse ranking and a second technique purifies the recommendation from the user set. In cascade model poorly-rated or low priority items will never be recommended.
- Meta-level: This hybrid shows that one recommender model can be used as input to next model. It is differs from feature augmentation model, as it use the entire model as input.

6. COMPARISON OF HYBRID AND COLLABORATIVE FILTERING:

The techniques of collaborative filtering have emerged not only in commercial fields but also in many research areas. This method as we have seen, it recommend based on ratings obtained as feedback from many individuals and use those ratings to make recommendations to a given user. A number of growing companies, including Amazon.com, Levis.com, etc, provide recommender system solutions. Amazon [6] is a major company makes use CF. Though this filtering has been very successful, it usually depends on history alone and reject information of data retrieved form contents. User rating is an important factor which determines the quality of recommendation, instead of contents. Taking consideration of multimedia recommendations, which lead to cold start problem, which mostly lead to data sparsity. While some users won’t rate the particular content after watching it, due to their lack of interest in rendering feedbacks. This led to low results in online video recommendation to the user. The recommender system may not be able to provide recommendations, when user’s rating is very less than that of the large number of items in the system. Due to this coverage problem occurs that is chance of missing valuable information.

There are also problems such as Neighbour Transitivity and Synonymy. Neighbour transtivity is a problem that occur with sparse databases, it refers users with similar taste of content may not be recognized as such if they have not rated any of the same items. The next problem Synonymy, it the matter of name used to refer the items, it is the likelihood of a number of the same or very similar items to have different names or entries. This will mostly reduce the computation capability of a recommendation system, which relies on generating predictions by comparing users in pairs. Video sharing sites, which thousands of videos are shared day by days and broadcasted over millions of channels. CF algorithms seriously suffer from scalability issues [7]. These limitations that are stated above must be overcome, in order to process further results. Therefore hybrid recommender systems have been provided, which can make use of both user preferences and contents.

In Hybrid filtering a number of hybrids remains less explored, while collaborative filtering is the most fully explored technique. The Content-based/collaborative feature augmentation hybrid was considered earlier a content-based and while Collaborative/content-based meta-level hybrid, in which collaborative generate information in the form of presentation of entire users rating and this representations later used in comparison of items across. Therefore, it is evident to know that meta-level hybrid techniques helps in avoid the problem of data sparsity by compressing ratings into models over many examples, which can be compared across users more easily than before.

In a typical recommendation system, the user coverage should be 100%, which is very hard to be acquired, as CF recommender system faces problem of data sparsity, it may be the case few users cannot be recommended any content. In a cloud environment coverage must be high. We observe that there is loss of coverage in CF recommenders and content based recommender [8], while gaining performance. However, most probably in large datacenters, there is no clear relation between performance and coverage variables. The coverage and accuracy must be combined, in order to eliminate unusual behaviors in recommendation. It is better to propose that hybrid filtering can reduce the sparse data and by including the features of collaborative recommender. It improves the performance gain by it unique hybridization techniques and provide user coverage. This shown by an assumption graph in fig. 3 where performance gain is plotted based on coverage.

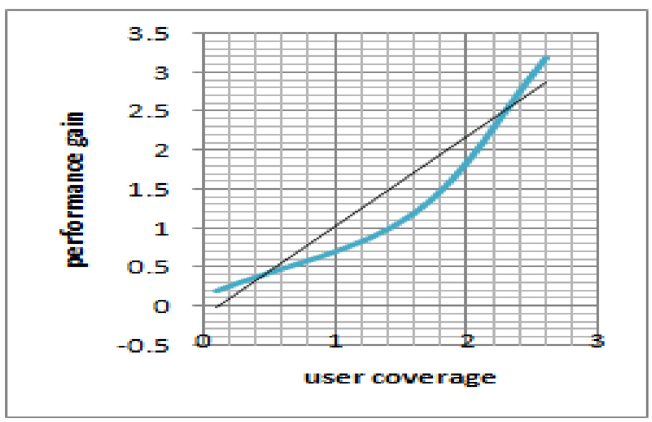

Fig.3 Hybridization coverage

ISSN:2319-7900

The above graph shows coverage increases as attempt to use of this method. There always less probability in accomplishing complete coverage in any technique but to maximum extent coverage can be attained using hybridization principles with collaborative.

By considering few social websites, such as youtube, Amazon and Netflixs [7] a graph is proposed. To give an idea, to show hybrid with collaborative approaches is suitable for social websites, for online video recommendations, an assumption is made based on the further studies using comparison of performance between them with these filtering aspects in a graph.

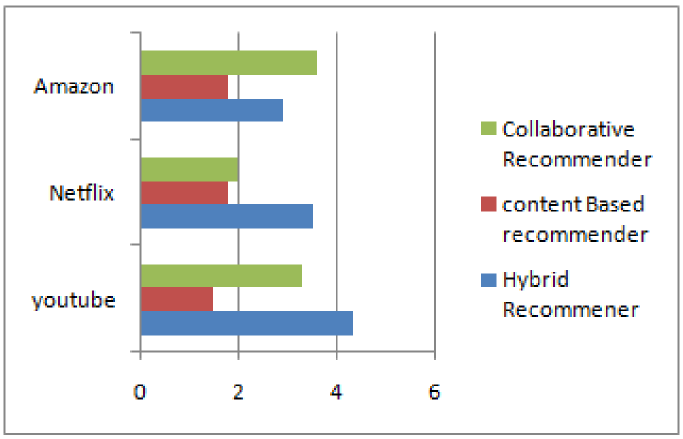

Fig. 4 Assumption graph for different filtering performance

The above Fig.4 shows a graph of comparison between different filtering techniques on different social network on basis of their performance. In Youtube, the performance factor rises as hybridization and collaborative filtering [8] approach is used, while for other sites such Movielens, Netflix also show hybridization collaborative is much effective [9]. In this survey, from studies done on various filtering approaches, one of the most effective filtering is hybrid with collaborative filtering. It is better to avoid content based in that sense it is much tedious, for collection of history of the users and contents increments simultaneously. This approach is focused only for multimedia sharing, in which it is possible to provide an easier search. By this filtering technique search can be made with reduce in delay, with good performance, high recall and better precious. In upcoming video sharing sites, as per research conducted more features can be added to improve the filtering process.

7. CONCLUSION:

A perfect recommendation system should be as dynamic as possible, so that it can keep up with the changing global trends. This indicates that the updates on profiles can be performed in real time, to certain proximate. The data flow in every second that may be videos, images or web documents. Computation complexity is major issue in almost all algorithms devised. The innovations of designs advance the computational speed, algorithms and techniques for obtaining low time computational complexity are expected in the recommendation system developments.

The hybrid collaborative filtering provides better efficiency and scalability compared to rest of the techniques in recommendations. It can be further extent in sense of adding extra features or tools, which help to make it more dynamic in nature. In future scenarios, recommendations can be applied along with, algorithms for removal of spammer videos or duplicated video contents, which would limit the content retrieved on search, to gain more accuracy.

REFERENCES:

[1] D. Li, X. Xie, L. Shang, H. Xia, T. Lu, and N. Gu, “Interestbased real-time content recommendation in online social communities,”Knowledge.-Based Syst., vol. 28, pp. 1–12, Apr. 2012.
[2] Yijun Mo, Jianwen Chen, *Senior Member, IEEE*, Xia Xie, Changqing Luo, and Laurence Tianruo Yang, *Member, IEEE* “Cloud-based mobile multimedia recommendation with user behavior information” vol. 8, no. 1,pp.186, March 2014
[3] D. Poirier, I. Tellier and F. Fessant “Reducing the cold-start problem in content recommendation through opinion classification,” in Proc.IEEE/WIC/ACM 2010, pp. 204–207.
[4] Z.-D. Zhao and M.-S. Shang, “User-based collaborative-filtering recommendation algorithms on Hadoop,” in WKDD, 2010, pp. 478–481.
[5] Kumar Abhishek, Saurabh Kulkarni, Vipin Kumar.N and Archana, P. Kumar, ‘’A Review on Personalized Information Recommendation System Using Collaborative Filtering” vol 2(3), IJCSIT,2011,1272-1278.
[6] Greg Linden, Brent Smith, and Jeremy York(2003), Amazon.com recommendations, item-to-item collaborative filtering. IEEE Computer Society, February 2003.
[7] Tuzhilin, A., Umyarov, A.; "Improving Collaborative Filtering Recommendations Using External Data," Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on , vol., no., pp.618-627, 15-19 Dec. 2008
[8] Bellogín, Alejandro, Iván Cantador, Fernando Díez, Pablo Castells, and Enrique Chavarriaga. "An empirical comparison of social, collaborative filtering, and hybrid recommenders", ACM Transactions on Intelligent Systems and Technology, 2013.
[9] X.Wu, Y. Zhang, J. Guo, and J. Li, “Web video recommendation and long tail discovering,” in Proc. IEEE ICME, 2008, pp. 369–372.
[10] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar,D. Ravichandran, and M. Aly, “Video suggestion and

ISSN:2319-7900

discovery for YouTube: Taking random walks through the view graph," in *Proc. WWW*,
2008, pp. 895–904.
[11] Referred to http://en.m.wikipedia.org/wiki/Recommender_system

Acknowledgement

Reshma.M.R ,PG Student of the Department Of Computer Science and Engineering, School of Engineering, Vels University, Chennai, India.
Dr.S.Neduncheliyan, Director of the Department Of Computer Science and Engineering, School of Engineering,Vels University, Chennai, India
Kalaivani K, HOD and Assistant Professor of the Department of Computer Science and Engineering, School of Engineering, Vels University, Chennai, India.