A Survey of Recommender Systems

Vijeta
M.Tech. Student, Department of Computer Science & Applications, Kurukshetra University, Haryana, India
Email: vijeta100.sangwan@gmail.com
RajenderNath
Professor, Department of Computer Science & Applications, Kurukshetra University, Haryana, India
Email: rnath2k3@gmail.com

ABSTRACT
Recommender System (RS) is one of the appealing topics in data mining and machine learning fields. Recommender System is data analysis software that helps in decision making processes such as what item to buy, what music to listen, what online news to read and so forth. Given a set of users and their reviews of items, RS generates ranked list of items to recommend to individual users. It predicts the rating of an item that a user has not rated yet. Recommendation generation phase recommends novel and relevant data from tremendous amount of data overloaded over internet. It acts as a subset of information filtering system. Its main aim is to provide products or items to users which are of high interest to them. This paper reviews the various recommender system techniques in a systematic way and aid new researchers to gain new insights or vision in this field.

Keywords: **Collaborative-based Recommender system, Content-based recommender system, Hybrid recommender system, Recommender system**

1. Introduction

With the development of Web 2.0, the amount of information and Internet users have increased at an unprecedented rate. Abstraction of novel and relevant information becomes a tedious task. Recommender system acts as a subclass of information filtering system. RS is one of the key areas in data mining and machine learning fields. Data mining is the extraction of patterns and knowledge from large amount of data. Everybody, today rely on recommendations of peers or acquaintances either work of mouth, ratings provided by them, letters. News, reports and so forth. In the present generation of web where number of choices are overwhelming due to enormous amount of information overloaded over the internet. RS helps users to find and evaluate item of interest. Abstraction of information from large amount of data becomes an easy task with the help of recommender systems. RS handles all type of spotty and unstructured data present in divergent formats.

RS works on web in order to make recommendations on the basis of users choices and past ratings. When deluged with choices and options such that what to wear? What movie to watch? What music to listen to? What stock to buy? Then RS cope with these massive decision domains. The developers of the first recommender system, Tapestry (1) coined the phrase “collaborative filtering” and several others have adopted it. RS seek to predict the rating or preference that user would give to an item. RS reduces search and navigation time and enhances user experience by assisting user in finding information.

RS has been divided into three main categories: collaborative-based recommender system, content-based recommender system and hybrid recommender system. Further collaborative filtering divided into two parts: memory-based collaborative filtering and model-based collaborative filtering. Other famous recommender system techniques are: Agent-based RS, Trust-based RS, Context-aware RS, Utility-based RS, Demographic-based RS, and knowledge-based RS. This paper describes all types of recommender systems and different types of techniques and algorithms used by them.

Collaborative-based recommender system uses the known preferences of group of users in order to predict the preferences of other users of similar behaviour. Content based filtering is the complementary of collaborative filtering. Item-item matrix and profile of individual user preferences are used to make recommendations rather than user-item matrix. Various features of items are stored into a table and using these features predictions are generated for new item or product of similar features that might be of interest of new user. Hybrid recommender systems unify both prior described approaches into one model and overcome all the limitations (sparsity, cold start etc) of individual recommender systems. This paper describes various recommender system techniques and algorithms. Main focus of the paper is to study and understand the various novel techniques used to make recommendations by the RS and too gain insights into this field. This paper accentuates the importance of RS.

The rest of the paper is organised as follows: Section 2 presents the research methodology, Section 3 gives the survey of recommender system, Section 4 presents the comparative analysis of various RS techniques and Section 5 concludes the research paper.

2. Research Methodology

Research on RS is in continuation from the last 15 years and so many advancements have taken place in this field. In order to present these technical increments in RS firstly, research paper related to RS will be collected from various sources such as ACM, Springer, IEEE, Elsevier etc. published between 1995 and 2015. Then these research paper will be classified into different categories based on type of RS techniques used. After that research papers under each category will be analyzed and new vision on recommender system will be drawn.

3. Survey of Recommender Systems

Total 52 research papers were collected relating to RS and these were classified on the basis of types of information utilised by them in order to make predictions. The following three categories of RS were identified - Collaborative-based filtering, Content-based filtering and Hybrid filtering. Other classification of RS utilised any one of the upper stated techniques and comes under the hybrid RS. Therefore, RS were divided into two broad classes – Primary RS class and Secondary RS class.

3.1 Primary RS class

This class comprises of two main/basic techniques of Recommender System therefore, named as Primary RS class. This class contained two types of RS:

3.1.1 Collaborative-based RS

The developers of the first RS, tapestry (2) coined the phrase “collaborative filtering”. In Collaborative Filtering pre-maintained user-item matrix was used and similarity between new user and matrix users was computed. After that the known preferences of new user community were used in order to predict the preferences of new user for a particular item.

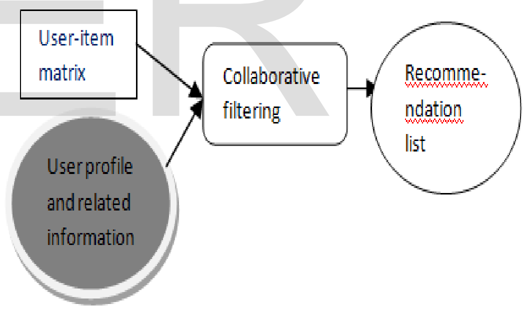


Fig.1 Recommendation based on Collaborative Filtering

Further Collaborative Filtering was divided into two types - Memory-based Collaborative Filtering: This filtering used the users rating in order to make recommendations. Abstract neighborhors for an active user then prediction of preferences on new item can be produced. We have an active user which becomes part of group of people with similar interests. Two good examples of memory-based recommendations were Neighbourhood-based recommender system and Top-N recommendations and Model-based Collaborative Filtering: Model-based Collaborative Filtering was a two step process:

first, a model could be estimated. Second, recommendations were generated for new user on based of learned model.

Table1: Determines Memory and Model-based Collaborative Filtering algorithms

Memory-based CF	Model-based CF
Similarity computation: Correlation based similarity (Pearson co-relation), Vector cosine based similarity. 1.Neighborhood-based CF(calculated neighbours of new user) 2.Top-n recommendations(k-most similar items or users)	Learned models: Bayesian belief net CF algorithm , simple Bayesian CF algorithm, clustering CF algorithm, regression based CF algorithm, mdp based, latent semantic CF

Following limitations of collaborative filtering were identified: (a) Cold start: when new item and user entered into the system, it becomes difficult to made prediction for them because of less number of ratings for new comers. (b) Scalability: due to increment in number of users and items, it made recommendation generating task time consuming and time complexity of algorithms becomes too large, degraded the performance of recommender systems. (c) Coverage: - It states percentage of items that an algorithm provides recommendations, reduced coverage occurs when number of items are greater than number of rating provided by users. (d) Data sparsity:- When user-item matrix becomes sparse and it affects the performance of recommendations and preferences of the collaborative filtering system.

3.1.2 Content-based RS

Content based filtering was the complementary of collaborative filtering. Item-item matrix and profile of an individual user preferences were used to made recommendations rather than user-item matrix. Various features of items were stored into a table and using these features predictions were made for new item or product of similar features that would be of interest of new user. This technique mainly recommend the items that were similar to those that are liked by a user in past.

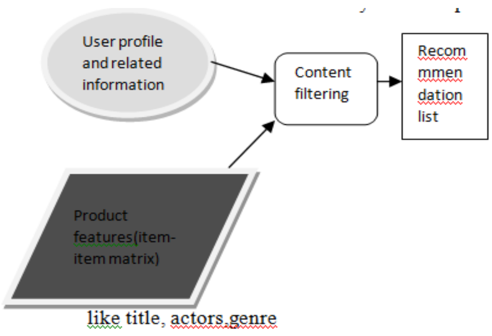


Fig.2 Recommendation based on Content Filtering

Limitations of Content-based Filtering – (a) Limited content analysis: Automatic features extraction was harder for multimedia data. (b) Item similarity: It becomes difficult to distinguish between two items if they had same features. (c) Cold start: This problem also existed in content filtering in some other form.

3.2 Secondary RS class

This class inherits features from Primary class and make use of techniques of Primary RSs. Therefore, named as Secondary RS class.

3.2.1 Hybrid Recommender system

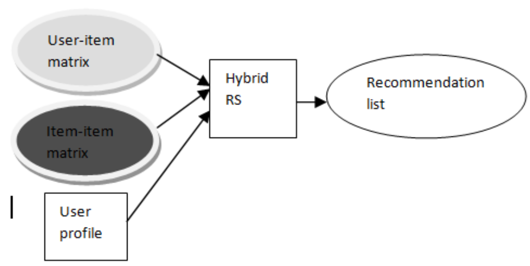


Fig.3 Recommendation based on Hybrid approach

Unify both prior described approaches into one model and overcome all the limitations (sparsity, cold start etc) of individual recommender systems.

3.2.2 Demographic-based RS

ISSN 2229-5518

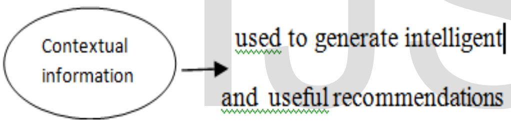
In demographic RS first, demographics (age, gender, country etc.) of users stored into the database. Next, figured out the group of users whose demographics were similar to new user. After that, used known item ratings of neighbours in order to predict the preferences for new user. This approach combined collaborative filtering with demographics and enhanced the CF technique.

3.2.3 Trust-based RS

This RS proposed by (3) was mainly used in social network with trust relations. It used user-item rating relations and friendship between users to made predictions. Relationships between users were calculated by propagating trust. Trust-based RS considered user own taste and trusted users' taste at the same time and made prediction without reducing coverage.

3.2.4 Context-aware RS

This RS proposed by (4) in order to generated more relevant recommendation by adapting the specific contextual situation of the user.



User interacted with system within a particular context. Context-aware RS labelled each user action with an appropriate context and generated the output to the user in that given context.

User X Items X Context → Ratings

3.2.5 Agent-based RS

In this RS agent maintained a user interest profiles and updating them based on the feedback whether the user like the item or not. Different agents were used for Collaborative Filtering.

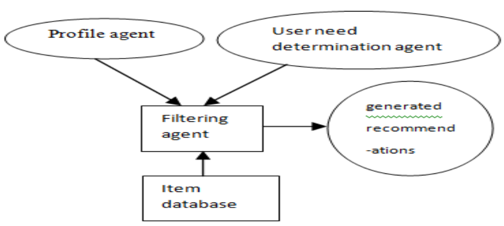


Fig.4 Agent based RS

3.2.6 Knowledge-based RS

A description about users' need or interest was used with features of items and recommendation predicted on the basis of knowledge of how these items meets a user needs. This RS inferred a match between items and users' need. This RS was of two types case-based and constrains- based RS.

4. Comparative Analysis of RSs

A comparative analysis of different recommender systems is carried out to find future research directions. The different RSs discussed above are compared on the basis of background, input data, and algorithm used by them.

4.1 Primary RS class

4.1.1 Collaborative-based RS

General meaning of RS was provided by (5). According to that people generally provided recommendations as input, that was used to generate output for users. Maintenance of RS was costly and many drawbacks of collaborative filtering also present in this approach such as: Sparsity, Cold start, Scalability problems. Further to improve collaborative filtering a new latent class model was proposed by (6). Two Latent class models were used: Aspect model and two sided cluster model. The advantage of this model was high degree of flexibility in modelling preference behaviour.

Next to increase scalability and quality of collaborative filtering prediction a clustering based approach was used (7) with collaborative filtering. Clustering algorithms were used to partition the set of items based on user rating data. Predictions were then computed independently within each partition. Clustering algorithm is better than nearest neighbourhood approach that discussed in previous section. (8) Found the limitations in filterbot proposed by (9) and overcome the limitations by combining collaborative filtering with personal information filtering agents for better recommendations. It also showed that using CF to create a personal

ISSN 2229-5518

combination of set of agents to produce better results as compare to individual agents.

Collaborative filtering is harnessed to make recommendations about movies, books, toys etc. And to overcome this flaw, Personality Diagnosis (PD) was proposed (10) was the combination of memory and model-based collaborative filtering. Given a user's preferences for some items, they computed the probability that he or she is of the same "personality type" as other users, and, in turn, calculates the probability that he or she will like new items. PD retains some of the advantage, that new data can be added easily and incrementally. This paper planed to extend the PD by incorporating user and title information beyond ratings.

User Rating Profile (URP) was a new model proposed by (11) modelled user rating profiles for collaborative filtering. The rating for an item was generated by selecting a user attitude for an item. Latent semantic model given by (12) was the powerful method for collaborative filtering and mining of user data based on statistical latent class model. Decomposition of user preferences using overlapping user communities was a novel idea and main advantage over memory-based method were higher accuracy, constant time prediction.

(13) Multiple Multiplicative Factor model was a discrete latent variable model. Data vector presented as a vector factors that had discrete, negative expression levels. The paper proposed by (14) further enhanced the recommendations effectiveness by introducing a personal recommendation procedure that was applied to internet shopping malls. The suggested procedure was based on Web usage mining, product taxonomy, association rule mining, and decision tree induction. They proved or suggested that choosing the right level of product taxonomy and the right customers increases the quality of recommendations. The recommender system named as WebCF-DT (Web usage mining driven Collaborative Filtering-based recommendation procedure using decision tree).

To cope with large data set a model was proposed in (15) (16) and used ClustKNN, a simple and intuitive algorithm. The method first compressed data tremendously by building a straightforward efficient clustering model. Recommendations were then generated quickly by using a simple Nearest Neighbour-based approach.

(17) Used Wikipedia as a ontology to solve the problem of text analysis in text based recommendation system that used traditional anthologies. The full system model combine semantic based analysis with collaborative via. content recommendation is presented. Semantic analysis was used at the place of lexical analysis because most of the data on the net is present in the form of text (unstructured data).

(18)Pattern-based workflow recommender systems employed historical usage patterns to generate recommendations. Semantics, on the other hand, could enable recommender systems to intelligently infer new relationships between workflow components. They combined both approaches to overcome drawbacks of individuals. "Workflow are a way to describe a series of computations on raw e-Science data. These data may be MRI brain scans, data from a high energy physics detector or metric data from an earth observation project. In order to derive meaningful knowledge from the data, it must be processed and analysed."Results showed that there was a clear improvement in the accuracy of the suggestions when semantics were combined with frequent usage patterns as opposed to only using patterns.

Table2: Year wise increment in Collaborative based RS

Year	Algorithm	Advantages
1999	Clustering/Partitioning	Improve quality of prediction, scalability, reduce computation time
1999	Latent class model	Flexibility and richness for prediction

ISSN 2229-5518

2000	Simple Bayesian Classifier	Better than co-relation based approach
2000	Eigen taste :A constant time collaborative filtering	O(1) processing time
2000	Personality Diagnosis	Made better predictions by combining memory + model approaches
2001	Scalable Neighbourhood Formation Using Clustering	quickly produce high quality recommendations, even for very large-scale problems
2003	Generative user rating profile model(URP)	It overcomes all the limitations of aspect model
2004	Regression Based Approach	Address the problem of data sparsity and prediction latency
2004	Model based Latent Semantic model	Higher accuracy, constant time prediction, explicit and compact model representation
2004	Multiple Multiplicative Factor Model	Discrete latent variable model
2005	Cluster based smoothing model	Overcome the limitations of memory and model based approaches with higher accuracy and increased efficiency of recommendations

2005	Co-Clustering model	Higher accuracy at lower computation cost
2006	ClustKNN	Improved nearest neighbourhood approach
2009	A hybrid collaborative filtering based web service recommender system	WSRec provides QoS information collection
2011	real time personalized recommender system based on slope one algorithm	Provided real time predictions

4.1.2 Content-based RS

Now to introduce title information a item-based top-n recommendation algorithm was proposed in (19) and in item-based technique a pre-made item matrix was used that represent relationship between items and then used these relationships to generate list of recommendation.

To eliminate the challenges of user-based collaborative filtering, item-based collaborative algorithm proposed in (20) (21)was major contribution in enhancing quality of recommendation. Looked into different techniques for computing item-item similarities (e.g., item-item correlation vs. cosine similarities between item vectors) and different techniques for obtaining recommendations from them (e.g., weighted sum vs. regression model). Finally, experimentally evaluated results and compare them to the basic k-nearest neighbour approach. At the end this paper concluded that item-based algorithms provided dramatically better performance than user-based algorithms.

Item based top-n recommendation algorithm proposed by (22) to address the scalability problem of model-based approach. Here first, similarity between items was computed after that, these similarities were used to compute similarity between basket of items and a candidate recommended item. In order to enhance the content-based filtering a Fuzzy set theoretic method (FTM) was proposed in (23)to address the problem of representation of item features, user feedback and relationship among them. FTM consist of a representation method for items' features and user feedback using fuzzy sets, and a content-based algorithm based on various fuzzy set theoretic similarity measures (the fuzzy set extensions of the Jaccard index, cosine, proximity or correlation similarity measures), and aggregation methods for computing recommendation confidence scores (the maximum–minimum or Weighted-sum fuzzy set theoretic aggregation methods).

Table3: Year wise increment in Content based RS

Year	Algorithm	Advantages
1998	Using Social and Content-Based Information in Recommendation	Combined both rating information and other information and provided better flexibility
2001	Item-based collaborative filtering recommendation algorithm	Enhanced the user based collaborative filtering
2004	Item-Based Top-N Recommendation Algorithms	up to two orders of magnitude faster than the traditional user-neighbourhood based recommender systems and provided recommendations with comparable or better quality.
2005	Feature based RS	Overcome the drawbacks of top-n recommendation algorithm, provided recommendation in the domains where no sufficient historical data is present about users or items
2009	Fuzzy set theoretic method (FTM)	treat the vagueness and imprecision in the context of the application

4.2 Secondary RS class

To overcome the drawbacks of content and collaborative based approach proposed by [resnick Paul, 1997] a new RS Fab was introduced by (24) that used a hybrid approach. Fab used two step processes: collection agent (collection of items in a database) and selection agent (selection of items from this database for particular user). The drawback of this approach was the construction of accurate collection profile. To overcome the prior defined drawback a Filterbot model a statistical approach was proposed by (9) addressed the sparsity problem of collaborative filtering. It combines both content and collaborative filtering approaches and increased the utility of agent software. Combined information filtering agents and opinion of community of users to produced better recommendations. Future directions suggested by this paper were selection of appropriate agents and personal agents.

Trust based recommendation system enhanced the prior described all types of recommender system (content, collaborative and hybrid based) by introducing trust based recommender system. Trust based recommendation systems were based on giving recommendation using only trust scores or combination of trust and similarity scores while giving suggestions. Trust based recommender enhance the level of approaches that were based only on user similarity. It improved the reliability of the seller/buyer in a transaction. Here sellers and buyers rated each other and that obtained data was used to make future recommendations. Trust/reputation system provided reliability scores of buyers/sellers that were calculated by scores given by the users of the transactions.

(25) Proposed regionKNN, hybrid collaborative filtering used for large scale web service recommendation. Region was formed using QoS and based on that model recommendation would be generated quickly. Highly accurate than traditional collaborative filtering algorithms.

(26) Given a survey of hybrid recommender system, new hybrid RS named as EntreeC was proposed that combined Knowledge based approach with collaborative techniques. Semantic ratings gathered by the knowledge RS enabled more accurate predictions and enhanced the effectiveness of collaborative filtering In order to achieve accuracy and speed a new recommendation technique based on hierarchical clustering was proposed by (27) a RS system based on Hierarchical Clustering is proposed. The user or item specific information is grouped into a set of clusters using Chameleon Hierarchical clustering algorithm. Further voting system was used to predict the rating of a particular item. In order to evaluate the performance of Chameleon based recommender system, it was compared with existing technique based on K-means clustering algorithm. The results demonstrated that Chameleon based Recommender system produced less error as compared to K-means based Recommender System.

Hierarchical clustering algorithm was mainly described in two phase's phase 1 and phase 2. Phase 1 was responsible for forming the initial graph and then partitioned into arbitrary number of partitions. Partitions were done on the basis of minimum edge cut. In phase 2 the partitions were combined until natural number of clusters are obtained. Here clusters were combined on the basis of two parameters Relative Closeness (RC) and Relative Interconnectivity (RI), at the end this paper proved that Chameleon clustering algorithm produced lower error thus identified better quality clusters as compared to K-Means clustering algorithm.

5. Conclusion

This paper has reviewed the algorithms proposed in the literature of RSs systematically. Though a number of algorithms have been proposed for efficiently discovering preferences or ratings of novel items for a user from large set of information. Firstly, they are classified in two broad categories – Primary and Secondary RS class. Then under each category the representive papers are studied and the limitations are identified and the paper further has suggested future research direction based on limitations encountered during analysis. There are few demanding issues in RSs that can be comprehensive for future research such as developing a strategy that will recommend items actually needed by a user, covers all type of choices for items. That is, an algorithm needs to be generated that should be able to generate accurate recommendations according to the users taste.

This paper has analyzed the existing RSs algorithm and techniques systematically in order to draw future research direction. So the main research direction which has come up is that future research will concentrate on advancing in existing methods and algorithms to improve the quality of recommender systems predictions and recommendations.

References

1. *Using collaborative filtering to weave an information tapestry.* **Goldberg, Nicholas D, Terry, B.M.** 1992, acm, pp. 61-70.

2. *Using collaborative filtering to weave an information tapestry*. **Goldberg, et al.** 1992, Commn. ACM, pp. 61-72.

3. *Recommender Systems Based on Social Networks.* **Sun Zhoubao, Han Lixin, Wenliang Hwng, Yan Hong.** 2015, Elsevier, pp. 109-119.

4. *Collaborative Filtering and Inference Rules for Context-Aware Learning Object Recommendation.* **Lemire Daniel, Boley Harold , McGrath Sean , Ball Marcel.** Canada : s.n., 2005.

5. *Recommender Systems.* **Resnick paul, Varian R. Hal.** 1997, Communication of ACM, pp. 56-58.

6. **Hofmann Thomas, Puzicha Jan.** *Latent Class Model for collaborative filtering.* 1999.

7. **O'Connor Mark, Herlocker Jon.** Clustering Items for Collaborative Filtering. 1999.

8. **Nathaviel Good, Schater Ben, Konstan Joseph A, Borchers Al.** *Combining Collaborative Filtering with Personal Agents for Better Recommendations.* 1999.

9. *Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System.* **Sarwar M Badrul, Konstan A Joseph.** 1998, ACM, pp. 1-10.

10. *Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach* . **Pennock M. David, Horvitz Eric.** 2000, pp. 473-480.

11. *Modeling User Rating Profiles For Collaborative Filtering.* **Benjamin, Marlin.** Canada : s.n., 2003.

12. *Latent Semantic Models for Collaborative Filtering.* **THOMAS, HOFMANN.** 2004.

13. *The Multiple Multiplicative Factor Model For Collaborative Filtering.* **Marlin Benjamin, Zemel Richard S.** Canada : s.n., 2004.

14. *A personalized recommendation procedure for Internet shopping support.* **Kim Jae Kyeong, Cho Yoon Ho , Kim Woo Ju , Kim Je Ran , Suh Ji Hae.** 2002, Elsevier, pp. 301-313.

15. **A Rashid Mamunur, Shyong K. Lam , George Karypis, and John Riedl.** *ClustKNN: A Highly Scalable Hybrid Model- & Memory-Based CF Algorithm.* 2006.

16. **Koren, Robert M. Bell and Yehuda.** *Improved Neighborhood-based Collaborative Filtering.* Calofornia : ACM, 2007.

17. *Wiki-Rec: A Semantic-Based Recommendation System Using Wikipedia as an Ontology* . **Ahmed Elgohary, Hussein Nomir, Ibrahim Sabek, Mohamed Samir, Moustafa Badawy, Noha A. Yousri.** 2010, ieee, pp. 1465-1470.

18. *Incorporating Semantics in Pattern-Based Scientific Workflow Recommender Systems.* **Kamran Soomro, Kamran Munir, Richard McClatchey.** London : s.n., 2015. pp. 28-30.

19. **George, Karypis.** *Evaluation of Item-Based Top-N Recommendation Algorithms.* Minneapolis : s.n., 2000.

20. *Item-Based Collaborative Filtering Recommendation Algorithms.* **Sarwar Badrul, George Karypis, Joseph Konstan, and John Riedl.** 2001, pp. 285-293.

21. **Greg Linden, Brent Smith,and Jeremy York.** *Amazon.com Recommendations Item-to-Item Collaborative Filtering.* s.l. : ieee computer security, 2003.

22. *Item-Based Top-N Recommendation Algorithms.* **DESHPANDE MUKUND, KARYPIS GEORGE.** 2004, ACM.

23. *Representation, similarity measures and aggregation methods using fuzzy sets for content-based recommender systems.* **Azene Zenebea, ∗, Anthony F. Norciob.** 2009, Elsevier, pp. 76-94.

24. *Content-Based,Collaborative recommendation.* **Balabanovic Marko, Shoham Yoav.** 1997, Communication of the ACM, pp. 66-72.

25. *RegionKNN: A Scalable Hybrid Collaborative Filtering Algorithm for Personalized Web Service Recommendation* . **Chen Xi, Liu Xudong, Zicheng Huang, Hailong Sun.** 2010, IEEE, pp. 9-16.

26. *Hybrid Recommender Systems: Survey and Experiments1*. **ROBIN, BURKE.** 2001, pp. 331-367.

27. *Recommender System Based on Hierarchical Clustering Algorithm Chameleon.* **Gupta Utkarsh, PatilDr Nagamma.** 2015, IEEE, pp. 1006-1010.