2. Literature Survey

 Literature survey was done on different types of recommendation systems and sentiment analysis. Each algorithm has different advantages and disadvantages. Several approaches and systems for the recommendation were developed, but many of them were focused on the integration of positive preferences in the recommendation process. Only some papers, mentioned at least negative preferences. Maximum papers used collaborative filtering which always gives better performance. Again some papers used clustering in different combinations. There are very less papers analysing negative ratings and reviews given by users. which has a wide area of scope. Literature survey is done four parts a)Clustering and fruit fly optimization b)Dynamic Clustering c) Collaborative filtering d)Textual review sentiments analysis

2.1 Clustering and fruit fly optimization

1.Zeng, Shan, et al. " collaborative A unified multikernel fuzzy clustering for multiview data." *IEEE Transactions on Fuzzy Systems* 26.3 (2017): 1671-1687.

Authors (Zeng, 2017)used multi-kernel fuzzy clustering for Multiview data efficiently. Authors used collaborative learning for individual views and multi kernels for the combination of all views in common kernel space. The authors tested the algorithm on synthesis datasets and proved that MKFCM obtained stable results than other existing clustering algorithms.

2.Xiao, Wenchao, et al. "Clustering algorithm based on FF optimization." *International Conference on Rough Sets and Knowledge Technology*. Springer, Cham, 2015.

The FF optimization algorithm is found in 2011. It is first used for clustering optimization by (Wenchao Xiao, Yan Yang(&), Huanlai Xing, and Xiaolong Meng, 2015) in 2015. The algorithm is easy for computation with less number of input parameters than all available swarm optimization algorithms. Authors have proposed shock factor for smell concentration value and tested initial range position on 10 datasets. The authors confirmed high precision, F-measure, and Dunn's index.

Zhou, Ruihong, et al. "Improved FF Optimization Algorithm-based density peak clustering and its applications." *Tehnicki vjesnik* 24.2 (2017): 473-480.

Author (Ruihong Zhou, Qiaoming Liu, Zhengliang Xu, Limin Wang, Xuming Han, 2017)proposed a density peak clustering algorithm based on FF optimization. FF algorithm is used to calculate cut-off distance and cluster centers. Authors shown that FF optimization converges fast and helps in correct clustering on seven UCI repository datasets.

4.Jiang, Zi-bin, and Qiong Yang. "A discrete FF optimization algorithm for the traveling salesman problem." *PloS one* 11.11 (2016): e0165804.

Authors (Zi-bin Jiang, Qiong Yang,2016) used a FF optimization algorithm for a famous large scale problem traveling salesman. The authors used FF to find neighboring cities from the source location. The authors solved the TSP problem efficiently by having an edge intersection elimination operator to find neighbors.

5.Zeng, Shan, et al. "A unified collaborative multikernel fuzzy clustering for multiview data." *IEEE Transactions on Fuzzy Systems* 26.3 (2017): 1671-1687.

Authors **(**Shan Zeng, 2017)used multi-kernel fuzzy clustering for Multiview data efficiently. Authors used collaborative learning for individual views and multi kernels for the combination of all views in common kernel space. The authors tested the algorithm on synthesis datasets and proved that MKFCM obtained stable results than other existing clustering algorithms.

6.Koohi, Hamidreza, and Kourosh Kiani. "User based CF using fuzzy C-means." *Measurement* 91 (2016): 134-139.

Authors (Hamidreza Koohi, Kourosh Kiani,2016) used fuzzy c means for CF. The authors stated that fuzzy c means is used the first time by them for CF. CF requires most similar users of the target user to give a recommendation. These most similar users can be collected together as neighbors. Fuzzy c means is applied on movilense dataset and results are compared with k-means and self-organizing map algorithms

### Collaborative Filtering and Similarity Measures

Table 1:Comparision of literature survey papers

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sr**  **No** | **Paper reference number** | **Filtering**  **method** | **Algorithm and similar measure** | | **Evaluation Measures used** | **Findings** |
| 1 | 37  2003 | Collaborativeitem based | Cosine similarity | | -- | -In reality many product pairs have no common  customers.  -Therefore item based method is inefficient in terms  processing time and memory usage |
| 2 | 45  2009 | CF and  Content based | -Deep learning  -CNN | | Precision ,Recall,  F-measure | -Recallis less for top 5 or 10 items-  -for requiredperformance recall calculated for top 20 items. |
| 3 | 30  2010 | Demographic | LinearPredictors:Gradient  descent,Exponentiatedgradient  descent | | MAE, MSE,PCA, PCU | -EG is better than GD  -Loss of accuracy as no of input features reduced for better performane |
| 4 | 22  2012 | CF and  content based | - centering-bunching based cluste (CBBC) algorithm | | --- | helps to get further effectiveness and quality of  recommendations .  helps in the exploration of other clusters which have  similarity closer to the active user |
| 5 | 42  2012 | CF and  Content based | -item based cf  -semantic similarity | | TF-IDF | -MLR and cosine similarity gives improved accurac human tags.  -TF-IDF on human tags decrease performance that  shows IDF not accurately model for human tags |
| 6 | 31  2013 | Demographic | Kendal's rank coefficient,  -genetic algo | | Pearson chi-square test,  z-score | -Chi square test has result half of ideal chi-square te audio --   -Kendal's T is better for low amount of data. |
| 7 | 25  2013 | CF | -user based  -item based | | Avg Precision | -best accuracy for top 3recommendations  -user based get for top 10 and item based for top 20 |
| 8 | 35  2013 | Model based  Content based filtering | -- | | Precision,Std deviation | Audiovisuael features give resonable drop in  performance than user ratingl |
| 9 | 40  2014 | Collaborative  model based | -Singuler value decomposition  -k-means using squared euclideadistance | | nPrecision and recaall, Novelty | -increase in noof dimensions reduce novel  recommendations  -Precision shows SVDdoes not create novel  recommendations. |
| 10 | 8  2014 | Collaborative   model based | -Agglomorative hierarchical  clustering algorithm  -enhanced rating similarity | | MAE, Computing time | -require less computation time  -rating similarity is more relevat with each other in s cluster than in other cluster  -predictions on ratings in  same cluster is more accurate |
| 11 | 24  2015 | CF model   based filterin | single hidden layer feedforward  neural network (SLFN) and weig g  extreme learning machine (WEL | | receiver operating characteristic (ROC) graph | improves the system performance |
|  | 12 26 | Hybrid CF | -Dynamic time drift model  -item clustering(k-means) | | MAE,RMSE | -time impact factor matrix can be updated manually -this dynamic method has better performance than st method.SVD |
| 13 | 32  2015 | Content based | -Textual similarity using cosine  coefficient  -KNN model  -for visual features user eye track done | | TF-IDF,RMSE by varying numbof neighbour ,Top N prediction   10 fold cross validation | -Addition of visual attention improved result than  standard methods |
| 14 | 43  2015 | CF and  Demographic  filtering | -KNN  -cosine similaritys | | MAE,RMSE | -Better performance by combination.--item features  selection should be appropriate |
| 15 | 36  2016 | Collaborative user based | Time sequence based CF algorit | | h Precision and recall | For top 5 recommendations precision is desirable bu recall is out of approval |
| 16 | 23  2016 | Collborative   filtering | Time-oriented Discounting meth Mean , avg precision | |  | User-oriented and Item-oriented Discounting metho can improve the performance |
| 17 | 44  2016 | CF a nd  Content based | -Cosine similarity  - graph Laplacian and label  propagation | | -Mean of avg distance  -no of clusters | -performanceependson dynamicfeature selectioand  extraction |
| 18 | 18 34  2017 | Content based | Ontology schema by semantic search | | User interest seen by mesuaring clicks | Content based Ontology schema by semantic se User interest seen by mesuaring Accuracy for semantic similarity depend on ontology |
|  | |  | clicks | instace mapping which is manual in this paper  -Use positive ratings to rank recommendations |
| 19 | 33  2017 | Content based model based | | VSM, DeepRecvsi  calculate semantic   similarity | Precision,Recall,MAE,  Diversity | Allows only positive similarity for better performan -better performance than low level features |
| 20 | 41  2017 | Collaborative   memory base | | RACF | MAE | -UPCC requiresless time than RACF for low density regions  -RACF has better performance than NRCF for high  density region.  -NRCFrequires row,column normalisation ofutility  matrix. |
| 21 | 39  2017 | Collaborative  -model based | | -probabilistic matrix factorizatio  -EM | nRMSE | -has better RMSE than SVD |

7.  FULI ZHANG, A Personalized Time-Sequence-Based Book Recommendation     
        Algorithm for Digital Libraries,2016,IEEE access

FULI ZHANG,2016[36] uses traditional memory based collaborative filtering algorithm to recommend books.It studies student's learning trajectories and combine with time sequential  cf algorithm.Algorithm take two inputs as time sequence information of book and circulation time of book,Then it calculate distance.

8. Xiaokun Wu, Bo Cheng, and Junliang Chen, Collaborative Filtering Service    
          Recommendation Based on a Novel Similarity Computation Method,2017,IEEE.   
          Transactions

              Xiaokun Wu, Bo Cheng, and Junliang Chen,[41]propose ratio based method for service recommendation.This paper uses memory based collaborative filtering. It computes similarty by using RACF.Result shows comparision of RACF and all other similarities like PCC,NRCF,UPCC,IPCC etc.Using this similarity prediction is done about service recommendation.

9. Greg Linden, Brent Smith, and Jeremy York,amazone.com, Amazon.com    
        Recommendations*Item-to-Item Collaborative Filtering,2003,IEEE access*

         Greg Linden, Brent Smith, and Jeremy York,2003[,Amazon.com,37]proposes item based memory based collaborative filtering algorithm. It calculate similarity by using cosine similarity. Utility matrix can be calculated by iterating through all item pairs.But in reality many product pairs have no common customers. Therefore item based method is inefficient in terms of processing time and memory usage. This algorithm is extremely time intensive, with *O*(*N*2*M*) as worst case. In practice, however, it’s closer to *O*(*NM*), as most customers have very few purchases.

10. John Z. Sun, Dhruv Parthasarathy, and Kush R. Varshney, Collaborative Kalman      
         Filtering for Dynamic Matrix Factorization,2017,IEEE transactions

John Z. Sun, Dhruv Parthasarathy, and Kush R. Varshney*,*2017[39]uses model based collaborative filtering algorithm.It uses probabilistic matrix factorization. Paper proposes new CKF algorithm ,it improves temporal dynamics.CKF requires model parameters at start, but parameters are unknown .Therefore EM algorithm is used to learn the parameters.The E-step finds mean and covariance of initial states and M-step refines these parameters.

11.Kibeom Lee and Kyogu Lee, Using Dynamically Promoted Experts for Music R        
       ecommendation,2014,IEEE Transactions

                Kibeom Lee and Kyogu Lee,2014[40] gives music recommender system.It uses matrix factorizaion method Paper performs dimension reduction by singular value decomposition and then prepare for k-mean clustering. Relevant recommendations are checked by average rating in that cluster.

**12.**Rong Hu, Wanchun Dou and Jianxun Liu, "ClubCF: A Clustering-Based               
             Collaborative Filtering Approach for Big Data Application," IEEE Transactions on     
              EMERGING TOPICS IN COMPUTING, volume 2, no. 3, september 2014

               Rong Hu, Wanchun Dou and Jianxun Liu[8] have proposed clustering based collaborative filtering technique for big data application. In this they have calculated characteristic similarity as summation of functionality and descriptional similarity  and then clustered services using agglomorative hirarchical clustering algorithm. Then for collaborative filtering they have calculated rating and enhanced rating similarity in same cluster .If this enhanced rating similarity is greater than already set threshold then service is recommended to the active user. As the number of services in a cluster is much less than that of in the whole system, ClubCF costs less online computation time. Moreover, as the ratings of services in the same cluster are more relevant with each other than with the ones in other clusters, prediction based on the ratings of the services in the same cluster will be more accurate than based on the ratings of all similar or dissimilar services in all clusters.

13. Subhash K. Shinde , Uday Kulkarni "Hybrid personalized recommender system using   
         centering-bunching based clustering algorithm, "ELSEVIER Expert Systems with   
         Applications 39 (2012) 1381–1387

                 Subhash K. Shinde , Uday Kulkarni[22] proposed a novel centering-bunching based clustering (CBBC) algorithm which is used for hybrid personalized recommender system (CBBCHPRS). The proposed system works in two phases. In the first phase, opinions from the users are collected in the form of user-item rating matrix. They are clustered offline using CBBC into predetermined number clusters and stored in a database for future recommendation. In the second phase, the recommendations are generated online for active user using similarity measures by choosing the clusters with good quality rating. This helps to get further effectiveness and quality of recommendations for the active users. The hybrid approach proposed in this paper extracts user’s current browsing patterns using web usage mining, and forms a cluster of items with similar psychology to obtain implicit users rating for the recommended item. in the proposed CBBCHPRS, similarity is combined with density of the clusters. This helps in the exploration of other clusters which have similarity closer to the active user and provide him/her with good set of recommendations.

14. Kento Kawai, Hiroyuki Kitagawa "Collaborative Filtering with Implicit Feedbacks by   
         Discounting Positive Feedbacks",2016 IEEE computer society, 978-1-5090-2179-6/16.

     Kento Kawai, Hiroyuki Kitagawa[23] proposed three discounting methods for observed values in implicit feedbacks. The key idea is that there is hidden uncertainty for each observed feedback, and effects by observed feedbacks of much uncertainty are discounted. The three discounting methods do not need additional information besides ordinary user-item feedbacks pairs and timestamps. Paper addresses the recommendation problem with implicit feedbacks that implicit feedbacks has only positive feedbacks. In particular, the Time-oriented Discounting method significantly improves the baseline model, indicating that feedbacks from users will turn to noises over time. Additionally, the User-oriented and Item-oriented Discounting methods can improve the performance although these methods use only information about which users gave feedbacks to which items.

15.Marko Krstić, , and Milan Bjelica," Impact of Class Imbalance on Personalized Program   
        Guide Performance**,"** IEEE Transactions on Consumer Electronics, Vol. 61, No. 1,   
        February 2015

                Marko Krstić, and Milan Bjelica[24] proposed*,* a personalized TV program guide based on neural network is described The proposed system uses single hidden layer feedforward neural network (SLFN) and weighted extreme learning machine (WELM) as its learning algorithm. It is shown how class imbalance information can be exploited in learning the user preferences. This not only improves the system performance, but increases the user satisfaction as well*.*

16. Zhao Xu, Qiao Fuqiang," Collaborative Filtering Recommendation Model Based on User’s Credibility Clustering" IEEE,CPS, 978-1-4799-4169-8.

                                  Zhao Xu, Qiao Fuqiang[26] proposed a collaborative filtering recommendation model based on user’s credibility clustering by taking movie database.This model divides recommendation process into offline and online phases. Offline, it uses the result of user’s credibility for clustering and then writes the clustered information into a table in database.Online,finds the cluster that target user belongs to and then gives recommendation.

17. Latifa Baba Hamid, Sofiane Abbar and Amine Haquari "The Impact of Negative   
         Preferences on a Recommendation Process,"2012, IEEE, 978-1-4673-1520-3.

 Latifa Baba Hamid, Sofiane Abbar and Amine Haquari[27] proposed a system for recommending movies which combines positive and negative preferences to estimate the utility of a given movie for a given user.A new approach is defined over here that two profiles are created for each user positive above preset threshold and negative profile below preset threshold. User profiles are created using quantitative preferences on movie genres i.e multi-criteria ratings. To recommend a movie to user, system calculates similarity between descriptor of movie and positive  profile of user  .If this is greater than similarity between descriptor of  the movie  and negative profile of user ,then movie is recommended to user.

  18.  Wei Zeng, Ming-Sheng Shang" Effects of negative ratings on personalized   
           recommendation", IEEE, Hefei, China. August 24–27, 2010.

Wei Zeng, Ming-Sheng Shang [28] firstly divided the whole ratings into two categories: positive ratings and negative according to preset threshold -3. constraint clustering .For example, if a recommender system’s ratings structure is 1 to 5, we denote the ratings which is no smaller than 3 as positive ratings and negative ratings otherwise. From these two types ratings, proposed system build two bipartite networks which are positive bipartite network and negative bipartite network. As a result, system is able to obtain a new bipartite network by mixing these two networks linearly. Our recommendation is based on the new bipartite network. From the  result, it is observed  that lower ratings are also important as higher ratings in the recommender system. As it is shown , in the Amazon network, the accuracy has been increased by 8% if we consider lower ratings in recommendation.

19. Maryam Khanian Najafabadi\*, Mohd Naz'ri Mahrin, Suriayati Chuprat, Haslina Md   
           Sarkan,"Improving the accuracy of collaborative filtering recommendations using   
          clustering association rules mining on implicit data"ELSEVIER, Computers in Human   
          Behaviour  (2017) 113e128.

            Maryam Khanian Najafabadi, Mohd Naz'ri Mahrin, Suriayati Chuprat, Haslina Md Sarkan[29] proposed a novel technique for recommendations based on user profiles created from implicit user feedback which is well suited to CF with data sparsity. The goal of this research was to improve the accuracy of recommendations by efficiently profiling user's item preferences from user's listening activities and tags associated with items and identifying similar preferences on the same category of songs. To overcome the problem of data sparsity, they have employed association rules mining technique to discover similar interest patterns among users from implicit information instead of the explicit information. In recent years it has been proved that a single algorithm is not generally able to pale the drawbacks in using basic CF and optimize recommendation accuracy. Hence, this proposed technique has improved the effectiveness of recommendations by employing clustering analysis technique to reduce the size of data and by using the association rules mining technique to find similar interest patterns between users.

20. BAOSHAN SUN AND LINGYU DONG, Dynamic Model Adaptive to User Interest       
        Drift  Based on Cluster and Nearest Neighbors,2017,IEEE Transactions

                   Baoshan Sun and Lingyu Dong,2017,[38] proposes hybrid CF method combining memory based and model based algorithm .It uses user-user cf  alongwith clustering. The algorithm I.e dynamic time drift model establishes the time impact factor matrix *T* based on the item clustering result and time decay function. For item clustering k-means algorithm is used.The paper  presents a time impact factor matrix that can be adjusted. The advantage of the matrix is that we can further reduce the prediction error by manual adjustment.

21.Wu, Xiaokun, Yongfeng Huang, and Shihui Wang. "A New Similarity Computation Method in CF Based RS." *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2017.

22. Liu, Haifeng, et al. "A new user similarity model to improve the accuracy of CF." *Knowledge-Based Systems* 56 (2014): 156-166.

23.Bobadilla, Jesús, Fernando Ortega, and Antonio Hernando. "A CF similarity measure based on singularities."*Information Processing & Management* 48.2 (2012): 204-217.

Xiaokun Wu (Wu, Xiaokun.,2017) has given a new similarity measure called JACRA which uses consistency in preferences of two users. The author has shown the proposed method which is working efficiently as compared to cosine, Jaccard, and Pearson. Haifeng Liu(Liu.,2014) also proposed a new user similarity model to improve accuracy. The author proposed a new similarity measure based on PIP. Similarity using PIP is not normalized, therefore the new heuristic similarity measure(NHSM) proposed by the author is normalized and has shown considerable improvement in accuracy. The measure differentiates all users with a minute difference in their similarities because of normalization.J.Bobaldia (Bobadilla, Jesús.,2012)suggested a new measure by calculating singularities for every product. He divides relevant and non-relevant votes for every product. He has considered positive, negative agreement, and disagreements between users. Here the author has shown improvements in the recommendation quality.

24.Chen, Zihao, and Zhengmin Li. "A collaborative recommendation algorithm based on user cluster classification." *2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS)*. IEEE, 2016

Zihao Chen(Chen.,2016) stated that the performance of CF is improved by using clustering. It worked on building a neighborhood set of users. He concluded that Cf recommends ratings without needing textual information and solve the problems caused by overloaded information of textual description of products.

25.Wang, Shaohua, Zhengde Zhao, and Xin Hong. "The research on CF recommendation algorithm based on improved clustering processing." *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*. IEEE, 2015.

26.Gao, Xiang, et al. "An effective CF algorithm based on adjusted user-item rating matrix." *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(*. IEEE, 2017.

Shao Hua Wang, ZhengDe Zhao(Wang.,2015) stated recommendation quality is improved by using clustering. They have suggested improvement in distance function for calculating similar users and items. Authors have used hamming distance and slope one method for clustering. Xiang Gao, Zhu, Hao(Gao.,2017) shown improvement in cf by adjusting the user-item matrix. They stated each rating given by the user has some deviation as the user's mental states are changing.

### Textual Review Analysis

On the literature survey, it is found that RSs based on ratings can not test the quality of products. Textual reviews are needed to confirm the quality of products. Sentiment-aware recommendation considers reviews given by customers to assess the quality of products. A survey is done on sentiment aware recommendations.

Despotovic, Vladimir, and Dejan Tanikic. "Sentiment Analysis of Microblogs Using Multilayer Feed-Forward Artificial Neural Networks." *Computing and Informatics* 36.5 (2017): 1127-1142.

Authors (Vladimir Despotovic, Dejan Tanikic,2017) have used an artificial neural network with a sigmoid activation function to analyze microblogs.Authors compared results with naïve Bayes, Maximum entropy, and SVM techniques and obtained good performance.

David Vilares, Carlos Gomez-Rodrõguez, Miguel A. Alonso, “Universal, Unsupervised (Rule-Based), Uncovered Sentiment Analysis”, Knowledge-Based Systems, Vol. 118, pp. 45–55, February 2017.

Authors (David Vilares *et al.*,2017) have proposed a universal and unsupervised model for compositional sentiment analysis (SA) driven by syntax-based rules for semantic composition. Their model outperforms existing unsupervised approaches as well as state-of-the-art compositional supervised models on domain-transfer settings and shows that the operations can be shared across languages, as they are defined using universal guidelines. They introduced a new concept of compositional operations for defining arbitrarily complex semantic relations between different nodes of a dependency tree. For future enhancement they decided to design algorithms for the automatic extraction of compositional operations that capture the semantic relations between the tree nodes.

Kumar Ravi, Vadlamani Ravi, “A survey on opinion mining and sentiment analysis: tasks, approaches and applications”, Knowledge-Based Systems, Vol. 89, pp. 14-46, November 2015.

Authors(Kumar Ravi et al,2015), have proposed a method for sentiment analysis to presents a comprehensive, state-of-the-art review on the research work done in various aspects during 2002-2014. They examined their work in six different dimensions such as subjectivity classification, sentiment classification, review usefulness measurement, lexicon creation, opinion word, and product aspect extraction, and various applications of opinion mining. The understood some intelligence techniques that have not been used thoroughly are random forest, evolutionary computation, association rule mining, fuzzy rule-based systems, rule miner, conditional random field theory (CRF), formal concept analysis, radial basis function neural network (RBFNN), and online learning algorithms. They concluded their work by proving that sentiment analysis has found various promising applications like market prediction, political sentiment determination, equity value prediction box office prediction, etc.

Weishi Zhang and Guiguang Ding, Li Chen, uli and Chengbo Zhang[25] proposed review-aware recommender algorithms that particularly exploited the sentiment classification results to automatically derive virtual ratings, and then fused them into item-based and user-based CF algorithms by which the User- Item Rating Matrix can be inferred by decomposing item reviews that users gave to the items.