PRODUCT RECOMMENDATION FRAMEWORK BASED ON CUSTOMER REVIEW USING COLLABORATIVE FILTERING TECHNIQUESL

C. Bharathipriya[1], B. Swathi[2], X. Francis Jency[3]

[1,3]Assistant Professor, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, Tamilnadu, India

[2]Student, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, India

[1]bharathipriya.cs@gmail.com, [2]swathi.18mcs@kct.ac.in

Corresponding Author: C. Bharathipriya

Email: bharathipriya.cs@gmail.com

Abstract

Recently customers are exposed to large variety of products and information on Internet, there is a necessity to filter, prioritize and personalize appropriate information to increase e-commerce demand. Using Recommender system, Business to Consumer (B2C) relationship can be benefitted and optimal, product selection is generated by solving voluminous data dynamically .In this work, a collaborative filtering is proposed to achieve top N recommendation about products to the consumers for purchase. In this work, the proposed recommender system focuses on obtaining similar group of customers using novel method. Personalized customer product recommendation is obtained by using classification and clustering algorithms. Good product evaluation is done using metrics like root mean square error (RMSE), mean square error (MSE). Recommender system has proved to enhance quality of decision making procedure and it gives a great impact on people's decision making. This work gives a recommender system which increases the value of e-commerce websites and worthiness in encountering best products for customers.

Keywords: Recommender system, Collaborative filtering, decision making, Business-Consumer

I.   Introduction

Recommender system is a filtering approach that filters user's needs and gives suggestions based on their needs. Recommender system is evolved as a separate research area in mid 1970's in Duke University. In early days people buy products

dependent on suggestions provided by their friends/relatives, however options have raised a lot and buy things digitally so to fulfill people with products' efficiency and to offer confidence while purchasing products, recommender systems are build. The large amount of data in internet is handled and the required information is processed and finally top recommendation list is produced. Recently data availability is increasing tremendously. Every data is required at some point in-case of analyzing it and producing a result. Recommender system is utilized as a suggestion provider at times even when user is not clear about what they want it shows the best results through previous analysis of data. The system provides personalized recommendations by using the user's interest or the content of the product in case of cold start problems. Recommender systems are utilized in various areas like video and music services like Netflix, YouTube, product recommendations are done in Amazon, Flipkart, social media platforms such as Facebook, Instagram. One can match their resumes with job description and seek jobs, education, government, hospitals, tourism, library, business, etc. Collective behavior of the user is used to recommend a system.

The paper is structured as: Section II depicts background of prevailing works on filtering techniques, Section III gives overview of recommendation system.The section IV introduces the proposed system and experimental results are discussed in SectionV. Section VI gives the summary of the work and the future work of this work is outlined.

II. Literature Survey

This section deals with the existing works for the past decade, the background of various filtering techniques are discussed and work related to each category is dealt as part of the literature study.

Content Based Filtering

Michal Kompan et.al [I] worked on fast content-based news is recommended based on cosine similarity search and effective representation of news is displayed. It solves information overload and personalized recommendation is done. They computed article similarity followed by creating user model and with those result recommendations is done. Hui Li et.al [II] merges probabilistic model and traditional content-based filtering recommendation algorithm. They proposed a novel approach for recommendation system. They used HMM's recommended items for matching user model. With user data, items are recommended. Comparing VSM dependent algorithm to proposed approach is done. Proposed system is more effective on showing user's interest.

Kazuhiro Iwahama et.al [III] depicts content-based filtering aims at music data in Musical Instrument. Digital Interface (MIDI) format. Feature enumeration and extraction is done on MIDI data. User profile is designed along with content model and compared. Filtering system is implemented and evaluated decision tree is used classification algorithm.

Yan Li et.al [IV] proposes Content based recommendation is used. Deep convolution neural network is used to avoid cold start problem. Frame level features are extracted with prevailing approaches and anticipate utilization of synthetic anchor points to cope up with gap among training and testing data and solve data incompleteness. Audio feature and video-based Meta data evaluation is done and vision feature is superior to textual meta-data and audio information.

Dima S. Mahmoud et al., [V] An enhanced content-based filtering (CBF) is performed where ABC model is applied in Content Based Filtering approach. Optimization is cleared using (ABC). A case study is done by implementing the algorithm and its result shows that applying improved CBF is extremely superior to prevailing CBF algorithm.

Hybrid Filtering

GMeryemet. al [VI] have discussed about existing software solution where that is not based on real information of the user and non- personal interests. Also they talk about the actors of academic orientation and factors that build's student's profile. In this work, collaborative and content-based filtering is cast off simultaneously by collecting information from student's about themselves and collecting similar people's interest and found the best solution to build a recommender system if both results are uncommon some more variables are added and reconstructed. This system helps the students to match their career path using real data.

Mathew et.al [VII] proposed a Book recommendation is done by combining both content and collaborative filtering and association rule mining is used to get effective and efficient recommendations.

Ashish Pal et.al [VIII] discussed Advantage of both Content and collaborative filtering is taken and similarity between the contents of two items is found. In content-based filtering correlation between two features are found using set interaction and similarities between two items are found and prediction is done using collaborative filtering.

Rudolf Turnip et.al [IX] conducted two experiments have been conducted CBF-CF-GL is better than CBF-GL (content-based filtering- good learning). Procedure followed is refining learning material, predictive rate and result of both methods. CBF-CF-GL depicts least MAE and hence anticipated method is superior than CF.

Vivek Arvind et.al[X] build a web recommendation precision in improving and enhancing item dependent collaborative filtering for effectual prediction rate and association mining made personalized recommender system for target users.

M. Ansari et.al [XI] depicts Content specific recommender as interactive programming learning scheme for Persian Speaking users. Personalized recommendations of educational materials are provided by evaluating user's activities and behavior inside system.

Collaborative Filtering

Rahul Kataryaet.al [XII] discussed recommender system that cast-off k-means clustering with cuckoo search optimization on Movie lens dataset. This model has shown improvements from the existing model. Their existing model is a hybrid one and obtained (0.78MAE) in this proposed model they have shown an improvement and obtained (0.75MAE).

Ting Bai et.al [XIII] Have taken an initiative on study about early reviewers and their impact on ecommerce platforms. 71% of e-shoppers follow only the early reviews and buy a product. Amazon and yelp data are compared. The review data is partitioned into three type's prior, majority and laggards. Prior reviewers are persons who are giving review in the earlier stage and also herding effect is understood on subsequent purchase of product. They have theoretically proven from sociology and economics that early reviews have major role in e-commerce platforms.

Due to space constrain they reported only Six products data set from Amazon and yelp and came to a conclusion that superior average rating score indicating superior popularity and finest helpfulness of reviewers is to enhance or diminish product's popularity. Evaluation metrics used here are Overlapping Ratio at rank, Hit ratio at rank, Ratio of Correct Comparison Pairs. Review spammers are removed as data cleaning procedure with Review text spamming (RT), Early deviation spamming (ED) and time-based spamming (TS). Linear regression is used to merge result of all the three techniques and find reviewer's spamming behavior score.

FengZhang et.al [XIV] explained about quickest procedure to compute collaborative filtering MAE and RMSE that are significant and representative methods. Compute MAE/RMSE based predictive ratings in contrary to true ratings. For analyzing item ratings, similarities among active users and neighbors has to be done.

Young-DukSeoet.al [XV] used one-month twitter dataset and gave a personalized recommender system dependent on strength in social networking. Conventional collaborative filtering technique is utilized to compute similarity among user Pearson correlation coefficient (PCC) and JMSD is used along with Cosine similarity to determine similarity among users and predict user's interest towards them.

Bat algorithm technique is used to compute features and neighbors for active user. They have compared bat algorithm with artificial bee colony based system and proved that bat algorithm performs better than artificial bee colony based system by mean absolute error and f1 score. Sambhav Yadava et.al[XVI].

Nilashiet.al [XVII]build a new hybrid approach to overcome two main disadvantages of recommender system such as sparsity and scalability. Theydealt with DR and ontology approaches. Ontology is used to enhance accuracy. Singular value decomposition (SVD) and DR approaches are used to determine similar items and users cluster to enhance scalability.

Kavitha Devi et.al [XVIII] has put forward a probabilistic neural network on movie lens dataset to analyze trust among users with rating matrix. With trust computation,

sparse rating matrix is determined, and trust is examined for active users. Calculated trust is cast off for recommending products.

Pujahariet.al [XIX] used movielens dataset and anticipated collaborative filtering approach which combines user-user and item-item collaborative filtering to produce effectual recommendation for homogeneous group. Soliman et.al [XX] recommends popular places using GPS traces of multiple users. Diverse reviews from users are collected and genetic algorithm is cast off to analyze unknown rates for unvisited regions dependent on explicit and implicit rating. Best places are recommended to users by highest rating.

Chengchao Yu et.al [XXI] have proposed restaurant recommender system by using historical orders and real time to enhance recommendations accuracy in contrary to FP growth and shown that anticipated approach has performed well in practical applications lesser in accuracy than FP-growth .F.O.Isinkaye et.al, [XXII] discussed the principle of recommender system, methods to implement recommender system using filtering techniques and effective evaluation metrics.Ivens Portugal et.al [XXIII] shows the recent trends of machine learning algorithms in recommender system, big data technologies used to exploit application domains such as movie is discussed. Precision and fmeasure are used to measure performance.

Yuanfeng Cai et.al [XXIV] used value-based neighbor selection to eliminate malicious users under shilling attack by detecting natural noise and abnormal noise. Random, average, bandwagon, segment is attacks are discussed. Beta protection-based beta distribution is used to eliminate malicious users however accuracy is merged with huge attack size. Deviation in rating pattern on targeted item are found and compared with group rating deviation from mean agreement to eliminate attackers. Performance is measured using fivefold cross validation, evaluation metrics is MAE, and accuracy is computed through f-measure, precision.

III. Recommmendation Stage

Recommendation system for a product or Item of Interest (IOI) contains three stages such as collecting information, learning, predicting. The following diagram shows the hierarchical representation of recommendation stages.
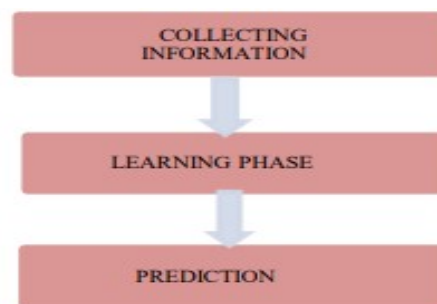

Fig. 1: Recommendation stages

Collecting Information

Collecting information in recommendation is essential as the recommendation is significantly dependent on information collected from consumer. It can be classified as explicit and implicit feedback.

Explicit Feedback

Asking user to rate the product and by using that data user's profile is created. The data may be sparse as many users don't care about ratings and it leads to low quality feedback. There may be cultural difference to rate it.

Implicit Feedback

Clicking a website, time spent on a website may be recorded and used it to recommend a system. Things you purchase are much more important data to use as recommender system and find the person's interest. Example: Amazon uses purchase data as they have large amount, YouTube uses the amount ofconsumption as the person spends time on viewing the video based on their interest. When we start a recommender system, we have to check the availability of data to recommend a system and source of user's interest is important.

Learning Phase

Learning algorithms are used to obtain filtering process after feedback collection. In this phase, training and testing of the data is carried out.

Training, Testing of the Data

Using training data we use machine learning and predict a recommender system and using test data we measure its accuracy as the learning doesn't know about the data and have not been trained, by using the test data we measure the accuracy.
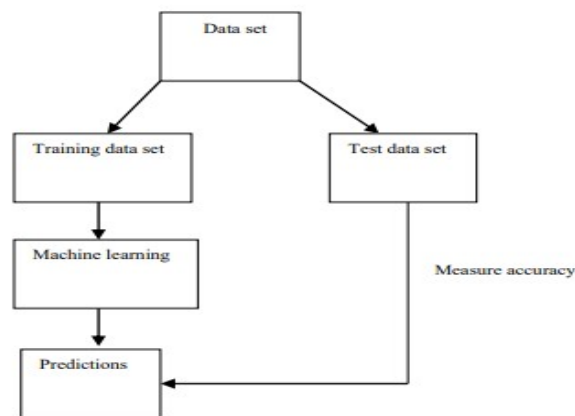

Fig. 2: Training and testing the data

K-Fold Cross Validation (CV)

K-fold CV is same as training and test set. Instead of one training set we divide it into many training set. Each training set is called as a fold. Each fold is used to train recommender system independently, and then we have to measure the accuracy against test result and take average. It takes lot of computational power.
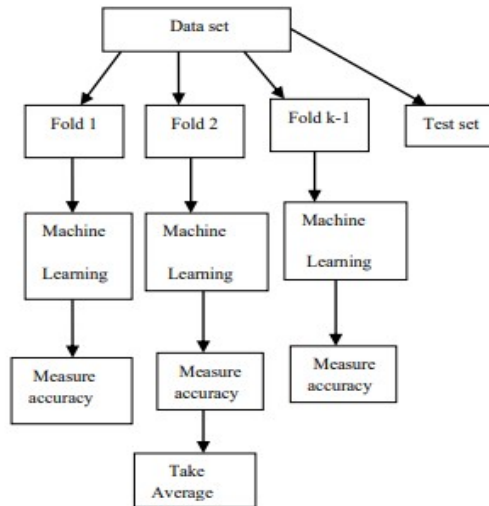

Fig. 3: K-Fold Cross Validation

Recommending Phase

Efficient techniques are to be used to provide good product for the consumers in order to develop good relationship with customers in e-commerce streams. Hence, it's found that there are three major filtering techniques exists and they are discussed below.

Content Based Filtering (CBF)

CBFis sourced on content or description given under category of product and also based on user's preference. Keywords are cast off to identify item and user's profile is created to identify user's likes and dislikes. It analyzes item attributes and produce predictions. CBF is recommending items dependent on its attributes instead of user behavior data. If we are yet to recommend contents like news, books, presentations, webpages. Then content-based filtering gives the best result in filtering the user's need personally. Details of user profile are not necessary in content-based filtering as it does not affect the process. Various technique used in identifying predictions are DT, NB classifier, NN, vector space model.

Hybrid Filtering

Combining both content and collaborative approaches and other approaches are called as hybrid recommender system. Hybrid approach can be obtained by combining both the results of content based and collaborative achieved separately. Several researches are going on to prove that hybrid filtering is more efficient than purely depending on content or collaborative techniques. Hybrid approach is helpful in cold start problems and handle data sparsity problem.

Collaborative Filtering

Collaborative filtering has a major role in recommender system. The system generates recommendations based on the rating profiles and review for users and items' rating history alike of present item or user. CF technique generates recommendations based on neighbor selection. Preference of item is done by user and it then matches the user with similar interest and so they are called as neighbors. Many algorithms are used to combine the choice of neighbor and suggest a recommendation set. It is capable of accurate recommendations in case of finding accurate neighbors. Many algorithms like k-nearest neighbor, Pearson correlation are used to find the similarities on user and item similarities.
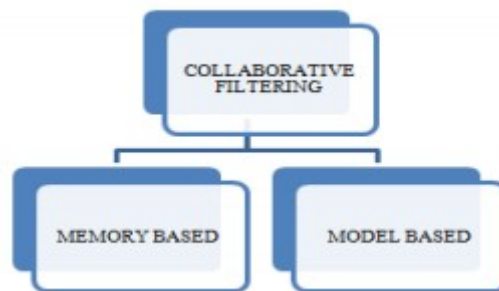

Fig. 4: Types of Collaborative Filtering

Memory Based Filtering

Memory sourced collaborative filtering can be done with user and item based approaches. Grouping of items based on their similarity and the similar items are called as neighbors. Neighbors are selected by building a model which will group similar items by extracting items rated by active user with user-item matrix analyzing similarity among retrieved and target item then finding top most items are predicted based on these two popular methods called correlation-based and cosine based.
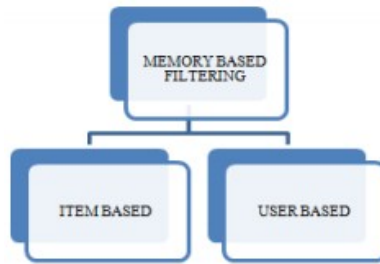
Fig. 5: Memory Based Filtering Types

Item Based

Candidatesmay appear more than once and they are combined together by boosting their scores and sorting the resultant recommendations of top N list of recommendation. Machine learning is used to find optimal ranking of candidates. Filtering process eliminates items that user already rated, remove items that are minimum rating and review threshold.
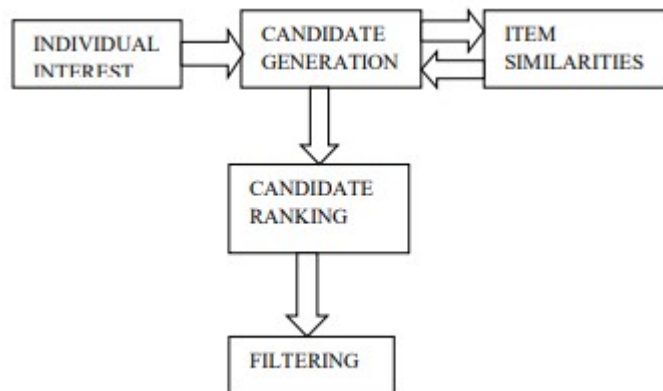

Fig. 6: Item Based Working

User Based Model

Grouping of similar persons and recommending things or contents what people like in common among their neighbors are recommended. Assumption is based on those alike users who go forsimilar items. Neighbors are found by comparing the ratings of similar items.
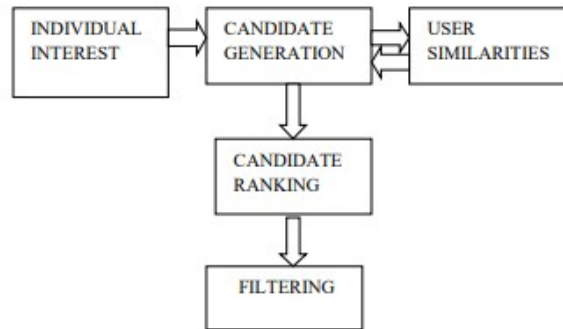
Fig. 7: Working of user basedmodel

Model Based

The significant disadvantage of memory based approach is necessity of loading huge amount of in-line memory. The serious crisis when determining rating matrix is extremely higher and there are numerous individuals utilized this system. Computational resources are consumed higher and its recital moves down; thus system will not react to users' request simultaneously. Model based technique resolves this crisis. There exists four general techniques for model based CF like classification, clustering, Markov decision process, latent and matrix factorization.

IV.   Proposed Work

Recommending best products for the customer is performed by analyzing the ratings and reviews. Here we use Amazon dataset containing ratings and reviews. Dataset is split to train and test followed by preprocessing. In preprocessing we perform stop words removal, lemmatization and tokenization using NLTK toolkit and polarity value is applied based on sentiment analysis. Using k-means clustering clusters are formed based on polarity values and later into classification algorithm like Support vector machine is used to classify the product groups and labels are formed. Then predicted value is Stored in database and recommending products will be listed in website.
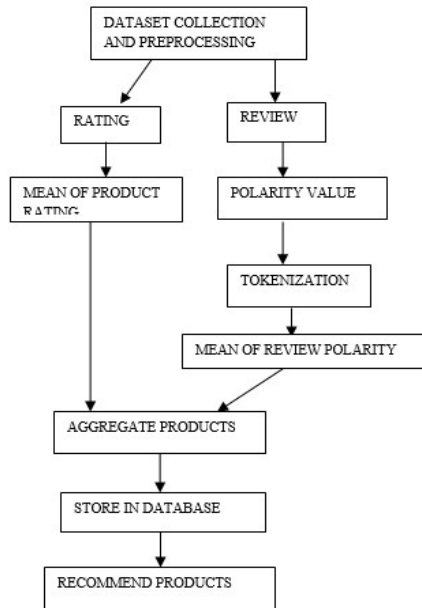
Fig. 8: Work Plan of Proposed System

V. Experimental Results and Discussion

Top N- Hit Rate

Hits are the actual number of contents that user likes.

$$\text{Hit rate} = \frac{\text{HITS}}{\text{USERS}} \qquad (1)$$

Measuring Accuracy Of Top N-Hit Rate

To measure accuracy of average reciprocal hit rate, hit rate leave-one-out CV and cumulative hit rate is used.

Leave-One-Out CV

Top end recommendation for each user is computed and intentionally we remove one of them from user'straining data. In testing phase we can evaluate recommender system by checking whether the system recommends the left out item or not.

Average Reciprocal Hit Rate

We sum up the reciprocal rank of each hit instead of summing up hits. It works on slots.

$$\frac{\sum_{i}^{n}=1\frac{1}{rank\,i}}{users} \tag{2}$$

Evaluation Metrics

Mean Absolute Error (MAE)

MAE computes amount of average deviation or prediction rate error with actual rating.

$$\frac{\sum_{i}^{n}=1|y_i-x_i|}{n} \tag{3}$$

Root Mean Square Error (RMSE)

$$\frac{\sqrt{\sum_{i}^{n}=1(y_i-x_i)^2}}{n} \tag{4}$$

n=Ratings, y=Predicted value, x=Actual value.

MAE and RMSE should be as low as they are error.

Table 1: Average Reciprocal Hit Rate

SLOTS	RANK	RECIPROCAL RANK
Slot 3	3	1/3
Slot 1	1	1
Slot 2	2	½

Cumulative Hit Rate

Cumulative hit rate is cutting out the lower predicted rating from the list.

Table 2:Cumulative Hit Rate

HIT RANK	PREDICTED RANKING
4	5.0
2	3.0
1	5.0
10	2.0

VI. Conclusion

In this work stage of recommender system, approaches like content, collaborative and hybrid filtering is discussed to get into a solution for the needed

recommendation along with their evaluation metrics utilized in calculating quality of the production of recommender system, related works and their preferences in using the type of recommender system is discussed. Various learning algorithms are used in our review data to classify them and cluster them and to generate recommendation models is listed. Using these metrics implementation of a product recommender system is given.The future work of recommender system lies in integrating one's own actualization to do justice to serendipity while recommending which is also support rather than replace human decision making by understanding preferences.

References

I.   Ansari, M. H., Moradi, M., "CodERS: a hybrid recommender system for an E-learning system. 2nd Int Conf of Signal Processing and Intelligent Systems, IEEE, Dec 2016, pp. 1-5

II.   Arvind, B. V., Swaminathan, J., "An improvised filtering based intelligent recommendation technique for web personalization", In 2012 Annual IEEE India Conf IEEE, Dec 2012, pp. 1194-119

III. Bai, T., Zhao, W. X., He, "Characterizing and predicting early reviewers for effective product marketing on e-commerce websites", IEEE Trans on Knowledge and Data Engineering, 30(12), , 2018, 2271-2284.

IV.   Cai, Y., & Zhu, D. "Trustworthy and profit: A new value-based neighbor selection method in recommender systems under shilling attacks", Decision Support Sys, 124, 2019, 113112.

V.   Devi, M. K., Samy, "Probabilistic neural network approach to alleviate sparsity and cold start problems in collaborative recommender systems" In 2010 IEEE Int Conf on Computational Intelligence and Computing Research (pp. 1-4). IEEE, Dec 2012, pp. 1-4

VI.   Isinkaye, F. O., Folajimi, Y. O., "Recommendation systems: Principles, methods and evaluation", Egyptian Informatics Journal, 16(3), 2015, 261-273.

VII. Iwahama, K., Hijikata, Y., "Content-based filtering system for music data. In 2004 IntSym on Applications and Internet Workshops. IEEE, Jun 2004, pp. 480-487

VIII. Katarya, R., Verma, O. P. "An effective collaborative movie recommender system with cuckoo search. Egyptian Informatics Journal, 18(2), 2017, 105-112

IX.   Kompan, M., "Content-based news recommendation. In International conference on electronic commerce and web technologies" Springer, Sep 2010, pp. 61-72

X.   Li, H., "Content-based filtering recommendation algorithm using HMM". In 2012 Fourth Int. Conf on Comp and Info Sciences. IEEE., Aug 2012, pp. 275-277

XI.   Li, Y., Wang, H., "A study on content-based video recommendation. In 2017 IEEE International Conference on Image Processing IEEE, Sep 2017, pp. 4581-4585

XII.   Mahmoud, D. S., & John, "Enhanced content-based filtering algorithm using Artificial Bee Colony optimisation. In 2015 SAI Intelligent Systems Conf, IEEE, Nov 2015, pp.115-163

XIII.   Mathew, P., Kuriakose, B., "Book Recommendation System through content based and collaborative filtering method. In 2016 Int Conf on Data Mining and Advanced Computing (pp. 47-52). IEEE, Mar 2016, pp. 47-52

XIV.   Meryem, G., Douzi, K., "Toward an E-orientation platform: Using hybrid recommendation systems" In 2016 11th International Conference on Intelligent Systems: Theories and Applications, IEEE, Oct 2016, pp. 1-6

XV.   Nilashi, M., Ibrahim, O., "A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. Expert Systems with Applications, 2012,507-520

XVI.   Pal, A., Parhi, P., "An improved content based collaborative filtering algorithm for movie recommendations", In 2017 Tenth Int Conf on Contemporary Computing, IEEE, Aug 2017, pp. 1-3

XVII.   Portugal, I., Alencar, P., "The use of machine learning algorithms in recommender systems: A systematic review", Expert Systems with Applications, 97, 2018, 205-227.

XVIII. Pujahari, A., Padmanabhan, V. "Group Recommender Systems: Combining user-user and item-item Collaborative filtering techniques." Int Conf on Information Technology, IEEE, Dec 2015, pp. 148-152

XIX.   Satpathy S, Prakash, M, DebbarmaSwapana, SenguptaAditya S.C, BhattacaryyaBidyut K.D, "Design a FPGA, fuzzy based, insolent method for prediction of multi-diseases in rural area", Journal of Intelligent & Fuzzy Systems, ISSN 1064-1246, vol. Pre-press, pp. 1-8, 2019.

XX.   Soliman, T. H. A., Mohamed, S. A. E. M., "Developing a mobile location-based collaborative Recommender System for GIS applications", 10th Int Conf on Computer Engineering & Systems, IEEE, Dec 2015, pp. 267-273

XXI.   Turnip, R., Nurjanah, "Hybrid recommender system for learning material using content-based filtering and collaborative filtering with good learners' rating. In 2017 IEEE Conf on e-Learning, e-Management and e-Services, IEEE, Nov 2017, pp. 61-66

XXII.   Yadav, S., Nagpal, S. "An Improved Collaborative Filtering Based Recommender System using Bat Algorithm", Procedia computer science, 132, 2018, pp.1795-1803

XXIII. Yu, C., Tang, Q., "A Recommender System for Ordering Platform Based on an Improved Collaborative Filtering Algorithm" Int Conf on Audio, Language and Image Processing, IEEE, Jul 2018, pp. 298-302

XXIV.   Zhang, F., Gong, T., "Fast algorithms to evaluate collaborative filtering recommender systems. Knowledge-Based Systems, 96, 2016, 96-103