Hybrid Based Recommendation System Using Switching Technique

Ms. Preeti Bhole[1], Mr. Suman Kumar Swarnkar[2], Mr. Virendra Swarnkar[3]
[2, 3] Assistant professor
[1, 2, 3] B.C.E.T. DURG

Abstract- *Now a day, most of e-commerce and social media sites use recommendation systems to help users find more relevant products easily. The key feature of recommendation is personalization which means different products are being offered for different users according to each user's interests. In literature, there are a lot of algorithms and tools which implement recommendation systems. The most common techniques for recommendation systems include Collaborative Filtering (CF) and Content-Based Filtering (CBF). To increase efficiency and accuracy, these methods can be combined in a hybrid recommendation system. Hybrid Approach is one of the Technic which focuses primarily on algorithms in the areas of CF, clustering and classification. In this study, we used Hybrid Approach for blending item-based and user-based methods of CF Using the Euclidean Distance Algorithm with switching approach. The Pearson Correlation Similarity and Nearest N-User Algorithm is used in user-based CF, while Tanimoto Coefficient Similarity and Generic Boolean Preference is used in item-based CF. Moreover, we added genre-based average ratings as content-based filtering so that the final recommendation list becomes more relevant to user. The proposed hybrid algorithm is tested on MovieLens dataset and validated with k-fold cross validation. This new hybrid recommendation system that is used to find patterns in data and develop a model for the purpose of making accurate and efficient recommender systems is proposed.*

Keywords- Recommendation Systems, Hybrid Model, Collaborative Filtering, Content Based Filtering

I. INTRODUCTION

Recommendation systems aim to improve accuracy of suggestions which users might interest. Recommender systems are based on cognitive science, information retrieval, approximation theory. It emerged as an independent area in mid1990s with the focus on structures of ratings. In recommendation systems, the main job is to find unseen and unrated items for a user in order to choose the correct items with the highest estimation values. The systems try to estimate ratings by using domain knowledge, similarity algorithms and machine learning approaches. So, a recommender system is responsible for predicting the rating or preference that a user would give to an item. When a user creates his or her profile, the system has to get the user preferences in order to provide her interesting recommendations.

	Movie 1	Movie 2	Movie 3	Movie 4
User 1	4	2	5	-
User 2	-	1	4	-
User 3	-	-	3	5

Figure 1 : User x Item Matrix

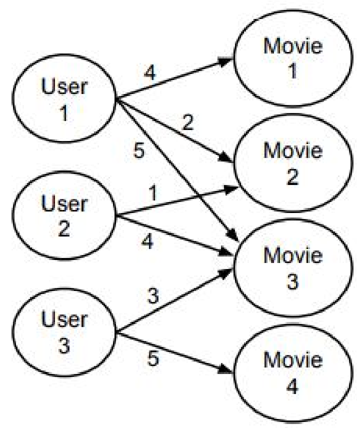


Figure 2 : User x Item Graph

In most of the recommendation systems, the dataset includes three main elements: user, item and rating. The data can be represented by a matrix. In the matrix, rows point out users, columns point out items and matrix entries are the ratings.

II. LITERATURE REVIEW

Data mining allows us to get recommendations or forecasts for future from very huge data. It is a process of getting meaningful information from meaningless raw data. The forecasts which data mining provides are crucial for recommender systems. The exact problem in this field is to get better and more accurate forecasts. Because of the better forecasts, not only the companies can sell more products, but also the customers can be satisfied because they are going to reach the desired products easily. The most common methods in this field are Collaborative Filtering, Content Based Filtering and hybrid approaches.

The first system that implemented the collaborative filtering method was the Tapestry project at Xerox PARC [15] in 1992. The project coined the collaborative filtering term. One of the other early systems is a music recommender named Ringo ([21], [25]) which is proposed in 1994. The other one is a system for rating USENET [30] in the same year. Group Lens is one of the first collaborative filtering recommendation system, which recommends movies [30]. Other examples are Amazon.com that recommends books, and the Jester system that recommends jokes.

If the relatively old studies on recommender systems are investigated, it is discernible that scientists mostly worked on text based domains to implement content-based filtering, collaborative and knowledge-based filtering. Sometimes, effective suggestions are produced on movie ([22]), music ([9], [18]) and web site domains. Besides, in recent years, efforts were made on e-government ([24]), e-learning and e-commerce domains (30] [26], [1], [10]). Recent studies show that filtering methods can give more effective results when they are used together. In this thesis, we aimed firstly the minimization of shortages of hybrid approach found in literature ([28], [12], [25], [23], [13]), and secondly the utilization of such an approach for another domain. The main purposes of this thesis are to investigate the new application areas of data mining and to calculate the effectiveness of our hybrid approach.

Data mining is the main part of recommender systems. In old recommender systems, data mining and data processing are utilized together and various methods are developed. In this thesis, data mining methods are studied together to reach at the hybrid approach. Similar studies have been performed since 1997 and have much more interest in recent years. Recommender systems are important not just for individuals but also for companies and governments. Future prediction is benefited in a lot of areas such as finance, shopping, internet, music, cinema, and e-government. In this thesis, one for such domain is selected applying our effective hybrid filtering approach.

Recommender systems are especially developed for e-commerce, e-government and e-learning ([29], [40], [19]) in recent studies. Moreover, a lot of recommender systems for music, cinema, book and entertainment domain (TV pro-gram recommendation, Flickr group recommendation) are developed. In development of such systems, data mining techniques and algorithms are effectively used. Some of the most effective filtering examples are content-based, collaborative, user-based and data-based. Hybrid recommender system is the combined usage of these filtering methods. In 2013, AR (association rule) and SVD (singular value decomposition) are utilized together as a hybrid solution for the recruitment of the partner filtering recommender system.

In recent years, hybrid recommender systems have gained significant importance. For example, Netflix Company organized a contest in 2006. The contest began on October 2, 2006 and continued through at least October 2, 2011. They wanted from contestants to create better movie recommendation by using their dataset. Netflix is already using a world-class movie recommendation system: CinematchSM. Its job is to predict whether someone will enjoy a movie based on how much they liked or disliked other movies. However, the contestants find better recommendations than CinematchSM. The contest by Netflix resulted in a big jump in the research of the recommender systems, more than 40,000 teams were trying to create a good algorithm. In the end, BellKor's Pragmatic Chaos [2] [29] [20] won the contest because their approach regards time unlike other studies. One of the other striking findings belong to Krishnan [21] who compared the recommendation results of machine against humans. According to the study of Krishnan, the machine won in most of the comparisons because machines can handle more data than humans can.

Hybrid approaches are developed for recommender systems used for e-commerce systems and TV programs. There are also some studies for social networks. One of them is about group recommendation system for Flickr. In another study, friend recommendation is performed using LinkedIn profiles. Music recommendation is investigated via auto-tagging and hybrid-matching. Some algorithms are developed for the recommendation used in search engines. Some studies made use of similarity trees such as fuzzy-tree. For example, recommender systems are developed for telecommunication products using this algorithm. Genetic algorithm and Bayes categorization are also utilized in some studies. These studies are beneficial to develop algorithms with scalable, accurate results.

In this thesis, some of the studies performed earlier are re-visited. Some of the proven algorithms are preferred to be utilized together. Firstly, Robin Burke's research studies [6] and [7] helped a lot in our work. We decided to use switching recommendation system.

The mentioned studies above could be found in resources section of the thesis. In this thesis, one of the promising sector "Movie" is selected as a domain to apply algorithms. The other data which we used is the registration records of junior and senior students in Middle East Technical University. We used the given scores of students to implement an elective-course recommendation system.

III. HYBRID APPROACH

Hybrid methods basically combine collaborative filtering (CF) and content based filtering (CBF) approaches to give recommendations. For Hybrid Filtering, we need to use same dataset for both CF and CBF algorithms, so the dataset should be suitable for both of them. Two algorithms have problems when they are used alone. Hybrid approaches not only tries to solve the problems but also tries to increase the accuracy of prediction. The common problems of recommender systems are explained below.

- **Early rating:** CF method has problem of early rating. This means that first user in the system is going to rate items without receiving any recommendation. So, we should decide which items to recommend without looking at the past ratings of this user. CF approach cannot provide recommendations for new users since there are no user ratings. Same problem is valid for new items since there are no user ratings on the item to forecast.
- **Data sparsity:** When there is not much information in dataset, the systems cannot calculate correct similarity. Sometimes the system cannot associate the user with other users, so it cannot find any recommended item. Besides this, if two users rated same items, both are going to see same recommendation. Therefore, it is a hard problem to compute similarity.
- **Cold start for user:** If a new user participates the system, he or she is going to see no recommendation, because there is no given rating for items. If user does not have sufficient number of ratings, he or she can suffer from unrelated recommendation. This may occur when similar users to this user cannot be found. This problem is called Cold Start.
- **Cold start for item:** If an item is introduced to the system, it cannot be recommended until it is rated by anybody. Also, if an item does not have enough ratings, it can suffer from not being recommended. This problem is called Cold Start too.
- **Attacks:** If there exist attacks to the recommendation system, the system should recognize it and try to avoid it. For instance, some user can copy the other user's profile and can get same recommendations. The system should separate the users who are attacker and the users who are very similar.

In order to do hybrid filtering, CF and CBF algorithms can be implemented individually and displayed separately. In a second option, both scores of CF and CBF algorithms can be multiplied the ranking scores in order to merge them into a single recommendation set. In literature, there exist seven hybridizations

Techniques that are explained briefly below [7]:

- **Weighted:** The score of both recommendation components are combined numerically.
- **Switching:** The system selects one of the recommendation components and applies the selected one.
- **Mixed:** Recommendations from different recommenders are presented in a combined list.
- **Feature Combination:** Features of different knowledge sources are combined together and given to a single recommendation algorithm.
- **Feature Augmentation:** A set of features can be calculated by one technique, and then they can be the input to the next technique.
- **Cascade:** Recommenders are given strict priority, with the lower priority ones breaking ties in the scoring of the higher ones.
- **Meta-level:** A model can be produced by one technique, which is then the input used by the next technique.

IV. PROPOSED SYSYTEM

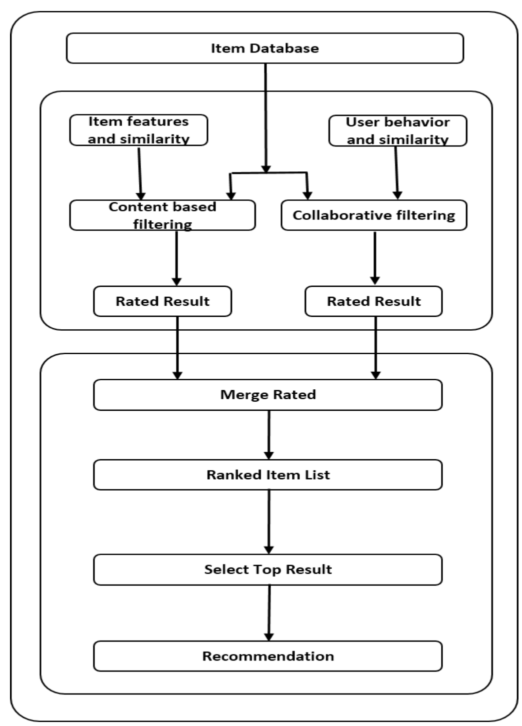


Figure 3: Block Diagram of Proposed System

V. METHODOLOGY

```
procedure Hybrid–Recommender
movie Data   movies.csv
for Fold k=1 to 5 do
model   trainingdatak:csv
test Model   testdatak:csv
for each user u in test model do
resultUB    UserBasedCF(model, u:ID)
CalcValidationMetrics(resultUB, testM odel)
resultIB    ItemBasedCF(model, u:ID)
Normalize(resultIB)
CalcValidationMetrics(resultIB, testM odel)
resultUB  IB    merge(resultUB, resultIB)
for each movie m in resultUB  IB do
if precisionIB > precisionUB then
resultCfHY BRID resultIB
else
resultCfHY BRID resultUB
end if
end for
Sort(resultCfHY BRID)
resultHY BRID    ContentBasedFiltering(resultCfHY BRID)
Normalize(resultHY BRID)
Sort(resultHY BRID)
CalcValidationMetrics(resultHY BRID, testM odel)
precisionk      add precisionu to end of file.
recallk         add recallu to end of file.
f measurek      add f measureu to end of file.
maek            add maeu to end of file.
end for
end for
end procedure
```

VI. EXPERIMENTS

For the execution, we utilized MovieLens datasets. It contains 1,000,209 unknown appraisals of 3,952 motion pictures made by 6,040 MovieLens clients who joined MovieLens in 2000. The rightness of information isn't ensured yet this information is extremely tasteful for motion picture suggestion frameworks. We are giving points of interest of the cross breed approach on MovieLens dataset to influence our calculation to clear. We executed a similar technique on our second dataset with a specific end goal to make understudy course suggestion framework.

In this investigation, we attempted weighted hybridization method, notwithstanding it didn't deliver wanted outcomes. In this way we chose exchanging approach. Our mixture arrangement works like the accompanying: In the first place, the aftereffect of client based CF (anticipated rating esteem is s1) and the consequence of thing based CF (anticipated rating esteem is s2) are figured. There were 100 motion pictures in each outcome list so that the vast majority of the films have both s1 and s2 anticipated rating esteems. In the event that a motion picture isn't in rundown of 100 consequences of client based CF, at that point it's s1 esteem equivalents to zero. Similarly, if a motion picture does not exist in 100 consequences of thing based CF, at that point s2 esteem equivalents to zero. Thusly, in weighted strategy, the s1 and s2 scores will be duplicated with weights and are totaled in a solitary esteem. In exchanging technique if a motion picture has both s1 and s2 comes about, the framework will choose the one which is higher exactness.

In weighted technique, we figured the aggregate weighted score of all motion pictures in result sets. The aggregate score is figured as (w1*s1)+(w2*s2). In the equation, w1 and w2 are weights in decimal number. The initial 100 films of the aggregate outcome are respected. Subsequent to attempting the distinctive numbers for w1 and w2, we can take a gander at accuracy and review. In basic terms, high exactness implies that a calculation returned considerably more pertinent things than unessential things, while high review implies that a calculation returned the greater part of the important outcomes.

In exchanging technique, we picked prescribed things by looking the accuracy and review of the arbitrarily picked 1000 clients in MovieLens 1M information. We run the calculations independently with weighted and exchanging hybridization procedures. We attempted the unadulterated cooperative sifting strategies (client based CF and thing based CF) alone and get the accuracy esteems. After that we picked the hybridization method. Exchanging technique gave preferable outcomes over the weighted strategy.

In this manner, we concentrate on it in later advances. In exchanging hybridization, CF technique with the most astounding accuracy esteem is chosen. At that point, at the last advance, we connected substance based sifting.

VII. EXPECTED RESULT

The Precision, Recall and F-Measure, the MAE (Mean Absolute Error) are calculated. The formula is –

$$precision = \frac{|relevantMovies| \cap |retrievedMovies|}{|retrievedMovies|}.$$

$$recall = \frac{|relevantMovies| \cap |retrievedMovies|}{|relevantMovies|}.$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}.$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|.$$

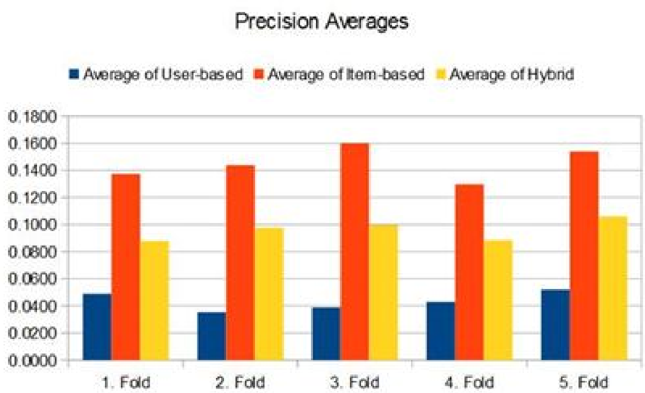


Figure 4: Precision Averages of Switching Hybridization

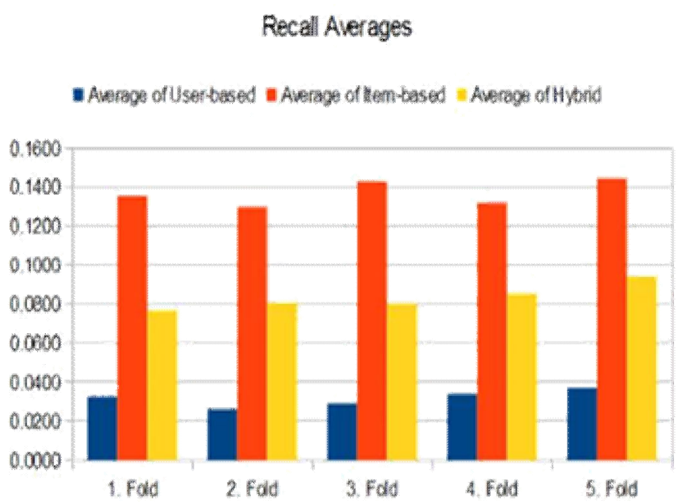


Figure 5: Recall Averages of Switching Hybridization

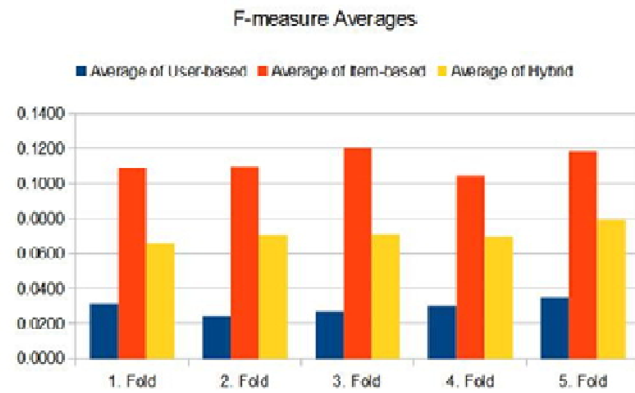


Figure 6:F-measure Averages of Switching Hybridization

Method	Avg. Precision	Avg. Recall	Avg. F-measure
User-based CF	0.093	0.024	0.036
Item-based CF	0.407	0.117	0.175
Hybrid Method	0.153	0.041	0.062

Table 1: Average Values of Precision, Recall and F-measure of Switching Hybrid Method without Euclidean Distance Algorithm CBF on MovieLens data set after 5-fold Cross Validation

Method	Avg. Pre.	Avg. Re.	Avg. F-mea.	Avg. MAE
User-based CF	0.093	0.024	0.036	2.876
Item-based CF	0.407	0.117	0.175	1.041
Hybrid Method	0.270	0.073	0.111	0.941

Table 2: Average Values of Precision, Recall and F-measure of Switching Hybrid Method with Euclidean Distance Algorithm CBF on MovieLens set after 5-Fold Cross Validation

VII. CONCLUSION

There is a data over-burden in the web and individuals barely discover the things which fit their taste. Individuals require to channel the data, so proposal frameworks help them along these lines. Suggestion frameworks become well known because of their advantages for the two organizations and clients. The organizations need to over to the client particular data to build the buys while clients need to achieve the applicable things effortlessly. There are three won methodologies in writing for better suggestion: content based sifting (CBF), community oriented separating (CF) and crossover approaches. The majority of the suggestion applications utilize the calculations of these three methodologies. The execution of the half and half approach is contrasted and the unadulterated synergistic and substance based techniques in a few examinations. Subsequently, the examinations exhibit that unadulterated methodologies have issues and are insufficient for a superior suggestion framework and the half and half strategy is required. The CBF and CF strategies can be utilized as a part of a half breed way to deal with beat a portion of the regular issues in recommender frameworks, for example, chilly begin and the sparsity issue. We enhanced a half breed technique and performed tries different things with greater and more shifted informational collection. After we done on tests, we contrasted results and approval measurements, for example, exactness, review, f-measure and mean total blunder (MAE). We utilized k-overlay cross approval procedure for this examination. After the examinations are done, we additionally did a study on clients to whom the dataset has a place.

VIII. FUTURE SCOPE

In future work, this study can run on distributed computing platform. Apache Mahout works alongside Hadoop which is distributed computing platform, so we can scale it out easily. We are using MapReduce which supports Hadoop and MapReduce. We did not use Hadoop and Map reduce for this study, however they can be used in the future work to implement movie recommendation system with scalable format.

REFERENCES

[1] A. Albadvi and M. Shahbazi. A hybrid recommendation technique based on product category attributes. Expert Systems with Applications, 36(9):11480–11488, 2009.

[2] R. M. Bell, M. Jahrer, and A. Tauscher. The Big Chaos Solution to the Netflix Prize 2008. Netflix Prize, Report, pages 1–17, 2008.

[3] D. Billsus and M. J. Pazzani. Learning Collaborative Information Filters. In Icml ’98, pages 46–54, 1998

[4] J. Bobadilla, F. Ortega, a. Hernando, and a. Gutiérrez. Recommender systems survey. Knowledge-Based Systems, 46:109–132, 2013.

[5] J. Breese, D. Heckermanx, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Uai, 1998.

[6] R. Burke. Hybrid recommender systems: Survey and experiments. User Modeling and User Adapted Interaction, 12(4):331–370, 2002.

[7] R. Burke. Hybrid web recommender systems. The adaptive web, pages 377–408, 2007.

[8] A. Dhond, A. Gupta, and S. Vadhavkar. Data mining techniques for optimizing inventories for electronic commerce. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’00, pages 480–486, 2000.

[9] J. Donaldson. A hybrid social-acoustic recommendation system for popular music. In Proceedings of the 2007 ACM conference on Recommender systems - RecSys ’07, page 187, 2007.

[10] F. Dong, J. Luo, X. Zhu, Y. Wang, and J. Shen. A Personalized Hybrid Recommendation System Oriented to E-Commerce Mass Data in the Cloud. In 2013 IEEE International Conference on Systems, Man, and Cybernetics, pages 1020–1025, 2013.

[11] S. Dooms, T. De Pessemier, and L. Martens. Offline optimization for user-specific hybrid recommender systems. Multimedia Tools and Applications, pages 1–24, 2013.

[12] S. J. E. Lee. A Study on the Hybrid Recommendation System. International Journal of Research in Engineering and Technology, 2, 2013.

[13] M. H. Esfahani and F. K. Alhan. New hybrid recommendation system based On C-Means clustering method. In The 5th Conference on Information and Knowledge Technology, pages 145–149, 2013.

[14] D. Fang, K. R. Varshney, J. Wang, K. N. Ramamurthy, A. Mojsilovic, and J. H. Bauer. Quantifying and Recommending Expertise When New Skills Emerge. In 2013 IEEE 13th International Conference on Data Mining Workshops, pages 672–679, 2013.

[15] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12):61–70, 1992.

[16] Ò. C. Herrada. Music Recommendation and Discovery in the Long Tail. PhD thesis, Universitat Pompeu Fabra, 2009.

[17] G. Ingersoll. Apache Mahout: Scalable machine learning for everyone. IBM Developer Works, pages 1–19, 2011.

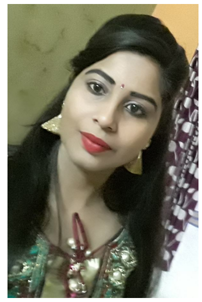
[18] M. Kaminskas, F. Ricci, and M. Schedl. Location-aware music recommendation using auto-tagging and hybrid matching. In Proceedings of the 7th ACM conference on Recommender systems - RecSys ’13, pages 17–24, 2013.

[19] A. Klašnja Milićević, B. Vesin, M. Ivanović, and Z. Budimac. E-Learning personalization based on hybrid recommendation strategy and learning style identification. Computers & Education, 56(3):885–899, 2011.

[20] Y. Koren. The BellKor Solution to the Netflix Grand Prize. Baseline, (August):1–10, 2009.

[21] V. Krishnan, P. K. Narayanashetty, M. Nathan, R. T. Davies, and J. a. Konstan. Who predicts better? Results from an online study comparing humans and an online recommender system. In Proceedings of the 2008 ACM conference on Recommender systems - RecSys ’08, volume October, page 211, 2008.

[22] G. Lekakos and P. Caravelas. A hybrid approach for movie recommendation. Multimedia Tools and Applications, 36(1-2):55–70, 2008.
[23] H. Liu, M. Amin, B. Yan, and A. Bhasin. Generating supplemental content information using virtual profiles. In Proceedings of the 7th ACM conference on Recommender systems - RecSys ’13, pages 295–302, New York, New York, USA, 2013. ACM Press.
[24] B. M. Marlin. Modeling User Rating Profiles For Collaborative Filtering. Advances in Neural Information Processing Systems, 16:627–634, 2003.
[25] F. Mourão, L. Rocha, J. A. Konstan, and W. Meira. Exploiting non-content preference attributes through hybrid recommendation method. In Proceed-ings of the 7th ACM conference on Recommender systems - RecSys ’13, pages 177–184, 2013.
[26] M. G. Ozsoy and F. Polat. Trust based recommendation systems. In Pro-ceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM ’13, pages 1267–1274, New York, New York, USA, 2013. ACM Press.
[27] M. Papagelis and D. Plexousakis. Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents. Engineering Applications of Artificial Intelligence, 18:781–789, 2005.
[28] M. Papagelis, D. Plexousakis, and T. Kutsuras. Alleviating the sparsity problem of collaborative filtering using trust inferences. In Proceedings of the Third international conference on Trust Management, pages 224–239, 2005.
[29] M. Piotte and M. Chabbert. The Pragmatic Theory solution to the Netflix Grand Prize. Working Paper, (August):1–92, 2009.
[30] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. G. Riedl, and An. Open Architecture for Collaborative Filtering of Netnews. In in Proceedings of CSCW’94 (Chapel Hill NC, October, pages 175–186, 1994.



Miss. Preeti Bhole, receive B.E. degree in computer science and engineering in 2013 from C.S.V.T.U Drug, Chhattisgarh, India and pursuing the M.TECH. degree in computer science from Bharti College of Engineering and Technology Drug, Chhattisgarh, India.

Mr. Suman Kumar Swarnkar, received the MTech degree in Computer Science and Engineering in 2015 from R.G.P.V BHOPAL, INDIA and Pursuing Ph.D. In computer Science and Engineering From Kalinga University, Raipur, Chhattisgarh.

Mr. Virendra Kumar Swarnkar, received the MTech degree in Computer Science and Engineering in 2013 from C.S.V.T.V Bhilai, INDIA and Pursuing Ph.D. In computer Science and Engineering from Kalinga University, Raipur, Chhattisgarh.