

# A Study on Similarity Metrics for Recommendations

Sipra Sahoo

Department of Computer Science and Engineering,  
SOA University  
Bhubaneswar, India

Bikram Kesari Ratha

Department of Computer Science and Engineering,  
Utkal University  
Bhubaneswar, India

**Abstract**—Recommendations are used in every aspects of life. Due to generation of huge data for reviews it is now difficult to judge the exact content. Recommendation systems play an important role in providing suggestions to the customers in terms of possible recommendations. Collaborative filtering is the process of identifying similar users or items and providing recommendations based on the commonality of the items or users. In this work only item based collaborative filtering is considered. The item based collaborative filtering is done based on the similarity or nearness of items that are considered. The nearness of the items can be measured by using many similarity metrics like cosine based, jaccard, extended jaccard, adjusted cosine, correlation, etc. The objective of this work is to evaluate and analyse the total execution time of the generation of top recommendation list by using the item based collaborative filtering (CF) approach. The online jester jokes dataset was considered in the experiment here. These datasets contains many users and hundreds of jokes. This approach will predict the jokes that the user is most likely he/she may like. For a joke prediction for the user the recommender system will look into the jokes that the users have previously rated or liked. By this recommendation it will be easier for user to choose the jokes which they may prefer to read. Different similarity measure are used so as to observe how the algorithm behaves. The main aim to find which measure takes least time to generate the top-N recommendation list using a particular similarity measure.

**Keywords**- collaborative filtering, item based collaborative filtering, recommendations, similarity measures

## I. INTRODUCTION

Recommender system is an system which filters the information for recommending some items to its users, it filters the data and recommends the items. It is commonly used in movie lens, book-crossing, jester, wiki-lens that uses a collaborative filtering to present information on items and products that are likely to be of interest to the reader/consumers. User's interest in the past is seen and analysed for the recommendation of any items. While presenting the recommendations, the recommender system uses details of the account of the registered user's profile, behaviour, preferences and habit of their whole group of users and comparison of the information to present the recommendations.

### A. Types of Recommender Systems

There are various types of recommender systems.

**Content-based:** This system processes to recommend items that are almost alike to the ones that the user wished or desired in the past. It gives preferences to the features of the products, movie, jokes, book etc. Similarity is checked for the items which the user had likes previously. The similarity of items is calculated based on the features associated with the compared items. It has to recommend some item to the users, and there is the job to produce a list of items to be recommended, the most similar users are needed to be found out or items are found after evaluating the commonalities, and then consider the neighbours to get the top most common items as the list of items recommended.

**Collaborative filtering:** Collaborative filtering is a method of identifying the similar clients and recommending what the common clients prefer. This system recommends items to the active user or the target users with that of the other users with similar preferences in the past. The similarity in preferences of the two users is calculated/evaluated based on the similarity in the rating history of the different users. This is the reason why it's also known as "people-to-people correlation." This filtering is considered to be the most popular and widely implemented technique in Recommender Systems. There are several methods that implements collaborative filtering. Neighbourhood methods gives stress on relationships between items or, alternatively, between users. The item-item approach models the preferences of a user to an item based on ratings of similar items by the same user. Nearest-neighbours methods are more popular than the item-item approach models. Nearest-neighbours methods are mostly used due to the considerable popularity due to their simplicity, efficiency, and their ability to produce accurate result and personalized recommendations for relatively smaller datasets. Several collaborative filtering algorithms have been composed essentially for information sets where there are a lot of people a larger number of customers than items (e.g., the jester online recommender data set has 73,421 users and 100 jokes).

**Hybrid recommender systems:** It is the combination of content and collaborative filtering system. This recommender system is based on the combination of the content based system and the collaborative filtering system techniques. A hybrid system is a combining techniques where given X and Y, it tries to use the advantages of X to fix the disadvantages of Y. For example, Collaborative filtering system suffer from the new-item problems, i.e., they cannot recommend items that has not been rated by the users. Where as in case of the content based approach this problem doesn't limit its prediction for new items

as content based is dependent on the features and description that are typically easily available.

### B. Overview

A particular recommender system introduces a collection of recommender job which categorizes the client aim. The proper and accurate datasets are chosen or selected for the purpose of evaluation. The calculation on the datasets can also be successfully done off-line by the use of the old available datasets and may be sometimes it also requires the on-line trial. The properties of the datasets is always considered and reviewed while selection of the datasets for the computation purpose. A survey is been done on the similarity metrics which will be used for the computation of the recommender system. Using those metrics we can also analyse the recommender system and its properties like its negative point and its positive points.

A report is been made on the obtained result and a comparison is been made by considering the different similarity metrics on the given datasets. By assessing a wide set of measurements on a dataset, we demonstrate that for a few datasets, while numerous distinctive measurements are emphatically connected, there are classes of measurements that are uncorrelated. We audit an extensive variety of non-exactness measurements, including measures of the degree to which proposals blanket the set of things, the variety and serendipity of suggestions, and client fulfillment and conduct in the recommender systems. To legitimately assess a recommender framework, it is vital to comprehend the objectives and errands for which it is, no doubt utilized. In this article, we concentrate on end-client objectives and undertakings (instead of objectives of advertisers and other framework stakeholders). We determine these undertakings from the exploration writing and from sent system. For each one undertaking, we talk about its suggestions for assessment. While the errands we've recognized are vital ones, in light of our experience in recommender frameworks research and from our survey of distributed examination, we perceive that the schedule is fundamentally deficient.

### C. Filtering Algorithms for Collaborative approach

A User based Collaborative filtering algorithmic approach is known for its straightforwardness and its productivity. Client-based Collaborative sifting algorithm produces suggestion of items for target client as per the perspective purpose of different clients. The suppositions which is made here is that if the appraisals of a few things are evaluated by some other clients are comparative, then the rating of different things appraised by these same clients will likewise be comparable or indistinguishable. Collective Filtering proposal framework utilizes the factual strategies to pursuit the closest neighbours of the target client and afterward contemplating on the thing rating evaluated by the closest neighbours to do the expectation of the thing rating appraised by the target client, and after that generate relating top N suggestion of the items. Collaborative Filtering framework that uses an area based algorithm is as takes after. In neighbourhood based algorithms, first step is that a

subset of clients are picked focused around their closeness to the dynamic client, and a weighted combo of their appraisals is utilized to generate expectations of items for the dynamic client.

Item based Collaborative filtering algorithmic approach in this item based approach, we mainly look into that how a particular item is liked or rated by the users. According to the ratings given by the users to the items, the items are recommended to the users. Rating of the items are given preferences so as to recommend the products to its users. In contrary to the user-based collaborative filtering algorithm, the item-based algorithm looks into and analyses the collection of the items the desired user has already liked in the past and computes how much and to what extend they are similar to the target item  $p$  using some similarity measures and

then select  $k$  most similar items  $\{p_1, p_2, p_3, \dots, p_k\}$ . Side by side their respective similarities, commonality  $\{s_{i1}, s_{i2}, s_{i3}, \dots, s_{ik}\}$  are also computed. Both calculation is done simultaneously. After this once the most alike items are found, next step is the prediction of those most alike items, which is then computed by getting a different formula's of norm of the target user's ratings or target items ratings on those similar items. There are two main processes here for the calculation of similarity between the items and then the prediction of the items to the desired users. The additional phase in the item-based CF algorithm is to evaluate the alikeness among items and then to selection of the most common items from them. The heart of the idea here lies in the similarity evaluation between two item  $p$  and  $q$ , we separate the users who have co-rated both of these items and after doing so we then to put forward a similarity measure computation method to determine the commonality between different items.

### D. Problem Statement

The main problem is to calculate the similarity between the different items in the given dataset. For calculating the similarity between different items extended jaccard coefficient, cosine similarity, correlation based similarity, adjusted cosine similarity is used which predicts the result that is top-N recommendation list. So we have to discover but which of the similarity measure gives output (that is the recommendation of the items) in least time and efficiently. And what are the advantages and disadvantages of different similarity measure.

### E. Organization of the work

The data that we are given from different online sites like jester online, amazon.in contains large amount of data user-item matrix. So any user who is using the facility of recommender system must be recommended with the top -N list of items as quick as possible so that the user will be recommended with items and he/she can browse it and can view it. With the increase in the online shopping sites or online joke sites, the number of users are also increasing and within less time good result is demanded by the users which they will

also prefer. Hence, it is very important that recommendation list is generated in less time.

Recommender systems have several trends and its application of the information mining techniques to the problem of how to deliver the personalized recommendations to the users for information, products or items during a live interaction. High requirement of information and data mining is required here. All we need is to analyse over the data or the information in the database of the systems. The k-nearest neighbour is quite popular on the web. The rapid and everyday growth in the volume of available information and data, the count of clients to web sites in recent years poses some vital and tricky hurdles for recommender systems. These are: How to produce recommendations that are of good quality for the users which satisfies them, • How to recommend items to the users in the minimal time as possible, • predicting several recommendations per second for millions of users, • how to achieve high coverage when there is also data sparsity, • Fast analysis over the users-item rating matrix whose dimensions are large. In traditional as well as we can say today's collaborative filtering systems the amount of work to be done increases with the increase in the number of users in the system. Work done by the recommender system is directly proportional to the number of users increasing day in and day out. Hence it has become very much necessary that the recommender system should be such that in case of high demand it can very accurately and quickly give predictions of the high quality recommendation lists even if the dataset is really large enough. And to cope up with the above problem we approach for the item-based top n recommendation algorithm. Item-based techniques dissects over the client-item matrix to recognize connections between distinctive items, and after that further utilize these connections to in a roundabout way to compute and recommends or forecasts of items for the clients. And obviously the recommended items to the users should be such that the users may like it.

#### Organisation of article

The following chapters gives the outline and organization of the thesis with an emphasis on the contribution made. Section 1 gives the introduction and fundamental concepts about recommender system and collaborative filtering (CF) specifically item based collaborative filtering. It also gives idea why we are using different similarity measure and also gives some information about the similarity metrics. Evolution of recommender system from the generation of past till the present scenario. Challenges to be overcome by the recommender system as the number of users are increasing rapidly. In Section II we are going to discuss various journal, papers from where we collected the required information for the project. Analyse all the papers and cumulated all the data. What are the works done in the field of recommender system and what still is needed to be done is clearly known after reading these papers.

Section 3: Similarity Metrics In this chapter, we are going to discuss about different properties required for a similarity metrics. Correlation based similarity in detail and how user based collaborative filtering failed. There are number of

similarity measure but some of them are used which are efficient enough to use with the data sets. For example for the datasets other than the binary attributes we cannot use the jaccard co-efficient because it is only for the evaluation of the jaccard co-efficient. Therefore appropriate similarity metrics are used for the study. Section 4: Proposed Work The item based collaborative filtering algorithmic approach is taken for the top N recommendation generation. This item based approach does the prediction of items to any particular active user by analysing over his past ratings. Therefore the ratings of each items are used for measuring the similarity between the items. Different similarity measure has its own advantages and disadvantages, keeping this in mind we have used the similarity measures for comparison for prediction of items for the users. Section 5: Result and Analysis In this chapter the result are analysed based the output of the program. Large datasets is taken, computed the ratings of the items of different users and the result is taken out and is analysed. We get some result over which we do analysis. Chapter 6: The conclusion, Scope for Future Work are given in the last section.

## II. LITERATURE REVIEW

Resnick et al. [1] brought into the original approach that is the user based approach. He determined the recommendations for the active user, identify similar users and compute a weighted average of their ratings of items not yet seen by the active user. Similarity is computed based on the users' historical rating behaviours. Breese et al. [2] came up with an idea of prediction problem for collaborative filtering. He did empirical analysis of prediction algorithm. This prediction algorithm tries to guess the rating that a user is going to provide for an item. This user will be referred as active user and the item as an active item. These algorithms take advantage of the logged history of ratings and of content associated with users and items in order to provide predictions.

Herlocker et al [3] audits distinctive key choices in assessing collaborative filtering recommender systems: the client assignments being assessed, the sorts of dissection and datasets being utilized, the courses in which expectation quality is measured, the assessment of forecast properties other than quality, and the client based assessment of the system all in

all. Not with standing surveying the assessment techniques utilized by former specialists, experimental outcomes from the dissection of different precision measurements on one substance space where all the tried measurements broken down harshly into three proportionality classes were additionally inspected. Measurements inside every equivalency class were firmly related, while measurements from distinctive equivalency classes were uncorrelated.

Deshpande et al [4] gave some contribution over the model-based recommendation approach that initially evaluates the commonalities between the several items. After that it uses the commonality to list the collection of items to be recommended to the active users. He explained two important steps that are

(i) First how to use and which method to use for calculation of the similarity between the items.

(ii) The next step is the way applied to combine those similarities so as to compute the commonality between the items, the user already has purchased and is in his bag and a candidate recommender item.

Sarwar [6] analysed different item-based recommendation generation algorithms. He also did survey on the several ways so for computing item-item similarities and also discussed different techniques for obtaining the high quality predictions from them. And finally evaluation of the result and analyse them with the basic k-nearest neighbour approach. It gives us a conclusion that if we take the item based algorithm in consideration it will provide much better execution time, quality than the user based CF algorithm. But the better quality is provided by the user based one. Maddali Surendra Prasad Babu [7] expressed that collaborative filtering algorithms (CFAs) are the most prominent recommender system for teaming up each other to filter the records they read from the previous decade. CFAs have a few offers that make them not the same as different algorithms. The arrangement correctness is one among them. A client based collaborative oriented algorithms is one of the separating algorithms, known for their effortlessness and productivity. A study is led for its execution and its productivity as far as forecast unpredictability.

### III. PRELIMINARIES

Similarity are important because these are used by the number of data mining techniques for determining the similarity between the items or objects for different purposes as per requirement such as Clustering, Anomaly detection, Automatic categorization, Correlation Analysis, Nearest neighbour classification, search, and prediction, Discrimination and characterization

Similarity: It is the numerical measure of the degree of which two items are alike. Items which are more alike have higher similarity between them. Similarity are often non-negative numbers and are fall generally in the range of [0,1], 0 for no similarity and 1 implies complete similarity.

Example of similarity measure are: Jaccard coefficient, Cosine similarity, Adjusted cosine similarity, Dice coefficient etc., Correlation based similarity, Extended Jaccard coefficient, Mean squared difference.

Dissimilarity: It is also the numerical measure of the degree to which the objects are different. For more similar objects the dissimilarity are lower value. Dissimilarity fall in the range of [0,1], with the upper range varying may be from zero to infinity. Example of Dissimilarity metrics: Euclidean distance, Minkowski distance, Manhattan distance, Hamming distance, Jaccard Co-efficient similarity

Similarities have some well-known properties:

1.  $s(x, y) = 1$  (or maximum similarity) only if  $x = y$ ,
2.  $s(x, y) = s(y, x)$  for all  $x$  and  $y$ , where  $s(x, y)$  is the similarity between data objects,  $x$  and  $y$

**Cosine based Similarity:** Moving forward to this similarity measure, any two things are taken as two items in the  $s$  dimensional client-space. The concept of angle is used here to calculate the similarity among the different items. The similarity between the two items is calculated by finding out the cosine of the angle between the taken any two items. Formally, in the  $n \times m$  ratings matrix (that is user-item matrix), similarity between any items let suppose that we are taking the arbitrarily items  $i$  and  $j$ , denoted by... Advantage of Cosine-based similarity

It is simple

It is very efficient to evaluate

Gives the value in between [0,1]

Disadvantage of cosine based similarity

The variation in the ratings given to the items between the different users are not taken for the computation.

#### Correlation-based similarity

For this comparability measure, commonality between any two items let assume  $i$  and  $j$  is acquired by registering the given association based on likeness. Keeping in mind the end goal to make the association based calculation effective and doable we should from the beginning confine the co-evaluated conditions (i.e., situations where the clients have offered rating to both the thing  $i$  and  $j$ ). Give us a chance to take that the set of clients who both appraised  $i$  and  $j$  are indicated by  $U$  then the correspondence similarity is given by the following formula

$$Sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - R_i) - (R_{u,j} - R_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - R_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - R_j)^2}}$$

Advantage of Correlation based similarity

The computation is accurate.

It does not calculate for the users, we isolate the condition where customers have rated both the item  $i$  and  $j$ .

#### Adjusted cosine similarity

The adjusted cosine similarity overcomes the drawback of the cosine based similarity. The main variation between the likeness evaluation in client-based Collaborative filtering and item-based Collaborative filtering is that in a condition where the client-based CF the similarity is to be calculated taking considering the rows of the matrix where as in case of

the item-based CF the commonality is evaluated taking along the columns. While the computation of the similarity using the ground level cosine measure in item-based consideration it has one vital flaw that is the differences in rating given by the users between different clients are not taken into view. The adjusted cosine similarity overcomes the above weaknesses by eliminating the corresponding user

average or item rating norm from every co-rated items. By using the formula, the similarity between items i and j using the adjusted cosine similarity is given by

$$Sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u) - (R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}}$$

Advantage of Adjusted cosine similarity

- Overcomes the drawback of cosine based similarity
- It subtracts the user average from each co-rated pair

### Extended Jaccard Coefficient

This extended jaccard co-efficient can be used for the document data and this similarity measure gets reduced to Jaccard coefficient in case of the binary attributes. The extended

Jaccard coefficient is also known as the Tanimoto coefficient. This co-efficient, which is represented as EJ, is defined by the following equation:

$$EJ(x, y) = \frac{x \cdot y}{|x|^2 + |y|^2 - xy}$$

## IV. PROPOSED WORK

Difficulty of the User-based Collaborative Filtering Algorithms These filtering systems are in trend and are in popularity and also has been very flourishing in past, but their more use has brought into some troubling flaws such as:

**Sparsity:** Recommender system works over the large datasets comprising of the set of users and the items. So, in reality several recommender systems are extensively used to calculate over the given large data sets ( for example, Amazon.com recommends books, jester.com recommends jokes and recommender music albums). If we consider the case that the users have purchased the items in a very less percentage from a large percentage of items, then as a result the concept of the on nearest neighbour algorithms seems to be incapable to predict any of the items for any of the target users. Aftermath is that the accuracy of recommendations will be poor.

**Scalability:** Day in and day out, the data and its users are increasing leading to the increase in the scalability of the datasets. Therefore the nearest neighbour algorithms is in utmost need so as quick evaluation that promotes both with the increase in count of users and the count of items. With long range of number of users and items, a well-known recent recommender system running with the algorithms which is in existence, undergoes severe scalability problems. Nearest –nearest neighbour algorithm for large, differential databases is weak which leads to explore any alternative recommender system algorithms. So to handle the problem of

scalability the item based collaborative approach is used for improvement.

### Proposed Improvement

We take the item based approach for the improvement of challenges. The sole agenda here is to analyse the user-item given matrix so as to identify existing likeness and relations among different products and then to use those similarity to derive the prediction of items for a given user for any particular item. The thought for taking the item based approach is that a customer will be more inclined to purchase those items that are alike in the features to the items the user had liked in the past and there will be high probability that he might obviously try to neglect items that are alike to the items the user who have not liked those in the past. In this technique there is no need to identify the neighbourhood of alike users when a prediction is to be made for the users, therefore as a output they will produce much faster items recommendations..”

## V. RESULTS

Similarity measure Taken: The cosine based similarity, extended jaccard based similarity, adjusted cosine based similarity and the correlation based similarity is taken for the comparison between different similarity metrics for item based top n recommendations. The above similarity measures are taken and used in the item based top n recommendation algorithm. For the generation of recommendation list it has two phase.

Steps for generation of recommendation list using different similarity measures are as follows:

Phase 1

**Input:** User-item matrix n\*m that is R and k that indicates the count of product to product similarities that will be stored for each product.

**Result:** m\*m matrix M

**Step 1:** The user-item matrix is taken and item to item similarity is calculated using the different similarity measure that we have considered in the above for all the items that are not similar.

**Step 2:** The value of M(i,j) is compared with the k most similar items. If it is same then it is left as the value is else it is made zero. This is what we get the output matrix M, which will be used in the next phase of the algorithm.

Phase 2

**Input:** The output matrix M from the previous phase, the matrix m\*1 U which store the products that has been purchased beforehand by the users, and the variable N that specifies the number of items that will be recommended to the users.

**Output:** m\*1 matrix x that stores number of items to be recommended. Its non-zero value indicates that the items that is in top n and is recommended to its users.

Steps Involved in phase 2

Step 1: The matrix M and U are multiplied to get the result in x.  
 Step 2: Purchased items are found out and if the item is purchased then just the value is put as zero in matrix x.  
 Step 3: If any value in x is not equal to the n largest value among the matrix x, then it is made as zero.  
 Step 4: The resultant value which we get at last in the matrix x that is the non-zero value is the top N items are recommended.

The two phases of the algorithm is implemented and run repeatedly using different similarity measures. While running the program the execution time of the whole program is being calculated to track which similarity measure is taking how much time to recommend items to the user's such that he may like it.

**Analysis**

The item based top n recommendation algorithm was executed repeatedly by deploying different similarity metrics. The results obtained for various similarity metrics are shown in the table. Using those execution time all the similarity are compared with each other. To compare a graph is plotted to compare the performance of different similarity. How they differ from each other. Graph showing comparison And it is observed that the cosine and the extended jaccard similarity takes less execution time as compared to the adjusted based similarity and correlation based similarity. Among these four the extended jaccard takes the least time for execution.

**VI. CONCLUSION**

Similarity metrics are used to calculate how much similar all the items are to each other in the matrix. We implemented the algorithm and get the result accordingly. Comparison is made between them by plotting the graph which depicts which similarity measure takes how much time. Hence, from the above table and graph it is concluded that the cosine based similarity and the extended jaccard similarity takes less time to recommend items to the active user in comparison to the adjusted cosine based similarity and the correlation based similarity. Final conclusion is that among the taken four similarity measures the extended jaccard takes the least time to recommend items. At initial stage it is observed that when the users are less cosine similarity behaves better but as the number of users goes on increasing the extended jaccard similarity behaves much better than the all other taken similarity measures.

In this field there are several scope of doing future work such as: By using different similarity measure we can see which gives the most accurate answer when compared with the other similarity measures. If we take into consideration the recommendation of the recommender system in contrast with the real life preferences we can compare their mean absolute

error. For collaborative filtering work can be done for serendipity and novelty.

TABLE I. EXECUTION TIME FOR DIFFERENT SIMILARITY METRICS

Users	<i>Cosine</i>	<i>A-Cosine</i>	<i>E-Jaccard</i>	<i>Correlation</i>
1000	55.99	289.80	57.00	299.88
2000	57.33	290.50	57.66	301.22
4000	59.20	291.80	58.34	302.10
8000	61.98	293.00	58.99	303.43
16000	64.31	293.98	59.80	304.22
20000	67.54	295.20	60.60	305.36
24000	68.99	296.60	61.11	306.51
28000	71,71	297.80	62.00	307.99
32000	73.00	299.30	62.78	309.16

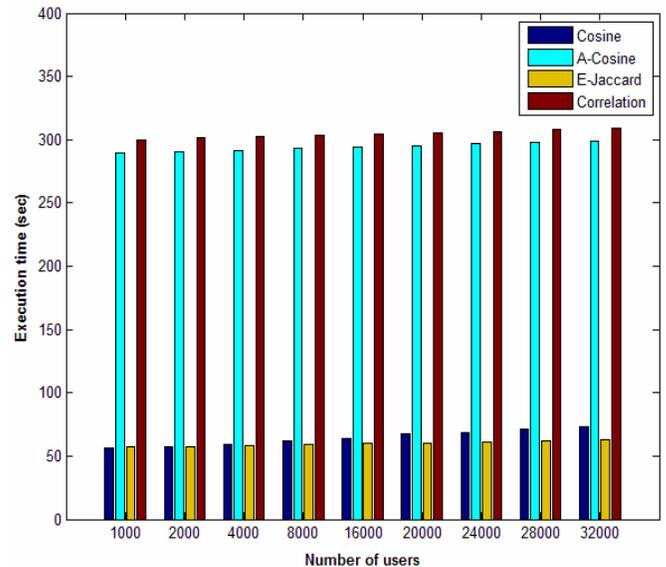


Figure. Execution time of different similarity metrics.

**REFERENCES**

- [1] P. Resnick., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, 1994. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of CSCW '94, Chapel Hill, NC.
- [2] J.S. Breese, D. Heckerman, C. Kadie, 1998. Empirical analysis of predictive algorithms for collaborative filtering. Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence.
- [3] J. Herlocker, J.A. Konstan, L. Terveen, J. Riedl., 2004. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS) 22.

- [4] M. Deshpande and G. Karypis. Item-based top-n recommendation algorithms. *ACM Trans. Inf. Syst.*, 22(1):143–177, 2004.
- [5] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, “Using collaborative filtering to weave an information tapestry,” *Communications of ACM*, vol. 35, no. 12, pp. 61–70, 1992.
- [6] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proc. of the WWW Conference*, 2001.
- [7] M. Surendra Prasad Babu, and B. Raja Sarath Kumar. 2011, An Implementation of the User-based Collaborative Filtering Algorithm / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (3) , 1283-1286
- [8] M. Papagelis and D. Plexousakis, 2005, “Qualitative analysis of user based and itembased prediction algorithms for recommendation agents” *Engineering Applications of Artificial Intelligence* 18 781–789
- [9] R. Bell and Y. Koren, 2007, “Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights”, *IEEE International Conference on Data Mining (ICDM'07)*, pp. 43–52
- [10] R. Aggarwal, H. Mannila, R. Srikant, H. Toivonen, and A. Verkamo, 1996. Fast discovery of association rules. In *Advances in Knowledge Discovery and Data Mining*, U.Fayyad, G. Piatetsky-Shapiro, P. Smith, and R. Uthurusamy, Eds. AAAI/MIT Press, 307–328.

#### AUTHORS PROFILE



**Sipra Sahoo** is working as an assistant professor in ITER, SOA university. She was born on 6<sup>th</sup> May 1984. She has completed her B.E (Information Technology) from BPUT in 2005 and M.E.(Computer Science and Engineering) in 2008 from Utkal University. and currently pursuing Ph.D in Utkal University. Her areas on interest include data mining, recommendation systems, soft computing, web personalization and sentiment analysis. She is a member of Institute of Engineers (India).

Sipra is currently working as an assistant professor in SOA university in the department of Computer Science and Engineering since 2009. Prior to this job she has worked in MIET and NIST during 2008 to 2009.



**Bikram Kesari Ratha** was born on 20<sup>th</sup> May 1965. He completed his MCA (Tech) from NIT Rourkela in 1988 and subsequently completed his ME (CSE) from the same institute. He completed his Ph.D from Utkal university. He has teaching experience over 25 years in several programmes like B.E., MCA, MSc, ME, M.Tech in different universities and institutes.

Dr. Bikram has worked as a professor in Nepal Engineering College (a constituent college of Pokhara university, Katmandu, Nepal) from 2000 to 2005.

Later he has worked as a professor in SRTM university. He has published more than 50 research papers and delivered many talks on different areas of computer science. His research interests include software engineering, data mining, ICT applications, computer networks and image processing.