

Ranking Technique to Improve Diversity in Recommender Systems

Vaishnavi.S
Final Year ME Software
Engineering,
SNS College of Technology,
Coimbatore

Jayanthi.A
Assistant Professor,
Department of CSE
SNS College of Technology
Coimbatore

Karthik.S
Professor & Dean
Department of CSE
SNS College of Technology
Coimbatore

ABSTRACT

The role E-marketing became important in everyday life. To help them to give correct products to human users, Recommender Systems have been evolved. It helps to find products related to user's interest. Mostly these systems employ collaborative filtering technique through which like minded users can be found. The key challenge is to attain quality in recommendations namely accuracy, diversity, novelty etc. Major algorithms have been focused on improving accuracy but diversity is important to have qualitative recommendations. On achieving diversity, distinct categories of items are taken for giving recommendations. Coverage is the main metric here. This paper gives one way to increase diversity by using LCM (Linear time Closed item set Miner) version 2 and I-Tree (Item-set tree). Data mining techniques are widely used in Recommender systems.

Keywords

E-marketing, Recommender System, Collaborative Filtering, diversity, LCM, I-tree, Data mining

1. INTRODUCTION

In recent years, there is drastic increase in amount of information in web. It leads to information overload. This explosive growth is mainly due to online world and E-Commerce. There are many e-commerce sites such as amazon.com, Netflix.com, MovieLens.com etc. These sites provides millions of products such as movies, books etc. E-commerce comprises core business process of buying, selling goods, services and information over the internet. Its usage is growing day-by-day. They help manufacturers in advertisement of their product also. Traditionally they advertise their product through TVs, radios, newspapers etc. These media create more cost and response of people to their product cannot be known [1].

These sites help in sales of products. Due to this, customers enjoy shopping from their home. Customers approach those websites to get products to which they are more interested. All categories of products can be found. Finding products of customer's interest is known as Personalization. They are a form of product differentiation. It differentiates one product from another competing one. It offers a unique, mostly used product to customer. Major problem is finding relevant content from huge information is a difficult task.

2. RECOMMENDER SYSTEM

For extracting relevant data, information filtering systems are mostly used. These systems removes redundancy in data sets i.e. it filter out the information from an information stream before presenting to user. In these systems such as search

engines (e.g. Google, Yahoo), user have to specify in advance what they are searching for. User's gives a query and they get relevant content to that query but it retrieves millions of relevant items. [2]

Main problem is giving products according to user's taste. To solve this, Recommender Systems (RS) were introduced. When the delivered information comes in form of suggestions, an information filtering systems is called Recommender Systems.

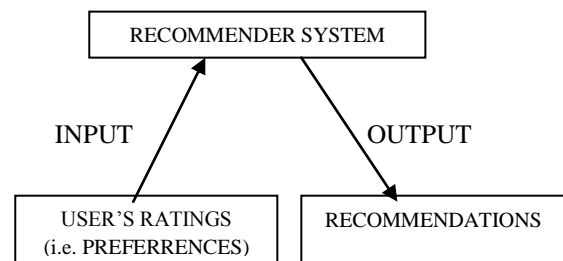


Fig 1: Recommendation Process

RS is an active information filtering system that attempt to present to user, information about item such as films, music, books, webpage, etc., to which user is interested in. It helps users in search of interesting items from large set of products. They give personalized recommendations to users. It allows users to weed out items which they want from huge set of choices. These personalized recommendations gather high importance since it allows user to get items from variety of products without loss in their taste.

Firms adopt these systems to provide increase in benefits and their popularity can be explained in the online world. If a customer adopts an RS and purchases a product and finds he does not like the products then he is unlikely to use the system again.

Recommender system understands user's needs and returns good recommendation in which items in top of list gain high relevant to user's need. User's preferences are analyzed by RS and it constructs a relationship model between items and users. User's interests are input to RS whereas output will be list of recommendations as shown in fig 1.

User's interests play a vital role in finding recommendations. User's interests are known by ratings they give to items. Rating indicates how a particular user liked a particular item e.g., Smith gave rating of 5 to movie 'Titanic'.

Utility function 'u' gives the usefulness of item 'i' user 'u' [3]. Two things are important; they are user space and item space which has user and item profile respectively. User profile has various user characteristics such as age, gender,

income etc. Similarly, item profile has characteristics of items such as price, name etc.

2.1 Approaches

Three main categories of approaches have been developed for recommendations. First is Content-based system which recommends items similar to those users preferred in the past. Second is Collaborative Filtering method which recommends items to users based on similarity between items. Lastly, Hybrid Approach combines above two methods.

All approaches aim to give qualitative recommendations. Quality of recommendations includes accuracy, diversity, novelty and serendipity.

Accuracy gives items more relevant to user's query. Diversity includes gathering variety of items, Novelty refers to how each item is different and serendipity give interesting item which user might not have discovered.

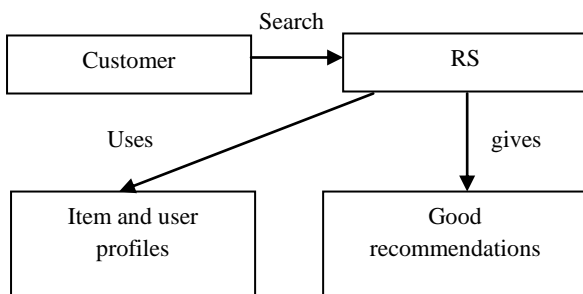


Fig.2: Work of RS

2.2 Recommendation Problem

Recommendations are based on ratings i.e., user preferences. Ratings can be acquired in two ways namely explicit and implicit. Explicit ratings are acquired by asking customer explicitly to rate the product and Implicit ratings are one which is acquired by observing customer purchasing behavior. Recommendation Problem can be defined as follows:

Let

$U \rightarrow$ set of users

$I \rightarrow$ set of items

and $u \in U$ and $i \in I$.

If item 'i' is useful to user 'u', user will give high rating to item. This is user preference. Known rating is denoted by $R(i,u)$. This rating value can be binary (i.e., yes or no) or numerical (i.e., 1, 2, 3,.....) or ordinal.

This known rating is used to estimate the unknown rating. Unknown rating is denoted by $R^*(i, u)$. It is estimated to recommend the products that are not yet discovered.

Recommending a product to user takes 2 phases namely

Phase I => estimating unknown ratings

Phase II=> calculating recommendation

for top N items.

It is given by the below figure:

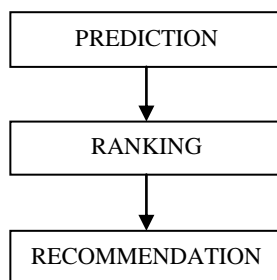


Fig 3: Task of Giving Recommendations

3. DATA MINING USAGE

Data Mining plays a significant role in extracting useful patterns from large sets. In recommender system application, data mining is a technique used to find recommendation rules. It consists of several tasks such as classification, clustering, Association rule mining, Anomaly Detection, Summarization etc.

Data Mining is dealt with producing different methods and techniques to process a large volume of data to get expected results within a minimum time gap [4].

Classification splits the data set into categories of items. Clustering aims at finding variables which have similar characteristics or behave in similar manner. Each similar item group is called cluster. Association rule mining allows the user to find frequent items from huge set of items i.e., it finds the relationship between items. E.g. If 'A' bought item 'X' and 'B' bought items 'X' and 'Y' then 'A' is likely to buy item 'X'. It determines the occurrence of item in the database. Various algorithms for these tasks have been developed.

Market basket analysis is an area where data mining is mostly used. It is used for recognition of dependencies in data. It tells 'if a collection of item is bought by a certain group of people then they are likely to buy by another group of people. It finds the relationship between purchases made by customer. It gives purchasing behavior of customer. So, it is useful for marketing.

There is always a trade-off between accuracy and diversity. Increasing accuracy decreases diversity. E.g. User A thought to buy cartoon movies. If 'A' gives cartoon movies to RS, if it gives movies of same directors then it is accurate but not diverse.

4. LITERATURE SURVEY

Recommender System has it's starting from information retrieval [5] and to consumer choice finding in marketing [6]. RS emerged in 1990's in order to overcome overload in information. This system wholly relies on rating concept. For acquiring true ratings, product rating acquisition problem [7] is one of the problem developed. The key input to RS is rating which can be implicit or explicit.

There is an inverse relationship between accuracy and diversity. Diversity should be increased by having minimal loss in accuracy [8]. Recommendation can not only be given to single product, it also allows recommending collection of products [9] [10].

Overcoming trade-off between accuracy and diversity, [11] describes an approach called variance based approach. This problem plays a significant role in RS since having diverse recommendations will give coverage of more products.

"Long-Tail" products are one that has high impact on sales. They are products which are less popular but earn high profit. Hongzhin, Bin says in [12] that by discovering these products, diversity can be increased.

Gediminas [13] gave a graph approach by using Max Flow and Max Bipartite problem. It gives top-N recommendations with maximum diversity. Based on taxonomy [14] i.e., by classifying products based on category, diversity can be attained. Fuguo Zhang tells in [15], diversity can be achieved by using trust between neighbors and gave a trust based algorithm.

Swapneel says in [16], by using social network information, diversity can be improved. These networks can be 'real world' or 'virtual' groups. Nagaraj describes in [17], a cross check approach. This approach will not give recommendation to an item which is already purchased or

rated by user. It also includes the category among users and items

5. RECOMMENDATION ALGORITHMS

The following methods are used mostly in RS

1) Neighborhood Based Technique

It computes similarity between neighbors and then finds neighbors

2) Matrix Factorization Technique

It uses matrix vectors to compute ratings. These techniques are Collaborative Filtering (CF) techniques which are mostly used now-a-days since it provides preferences based on other users who have same taste. Above techniques predicts unknown ratings. Then the ranking technique is used to get top-N recommendations. Standard ranking approach [18] ranks item from high predicted rating to lowest ones.

6. PROPOSAL

All the recommenders in the online world (e.g. Amazon.com) should have updations in their database to give up-to-date recommendations. Some companies keep a 'count' to the number of users rated the item, database is updated by the admin to record new ratings. It has disadvantage that till a new updation takes place, wrong recommendations will be given to user.

One way is to make automatic updation of database when the preference given by user i.e., if one user who preferred the item increases then database is updated.

Here Neighborhood technique is used to predict unknown ratings. In this technique, similarity between two users say user 'u' and user'u1' is calculated. Based on this calculation value, nearest neighbor to user is found and ratings is acquired.

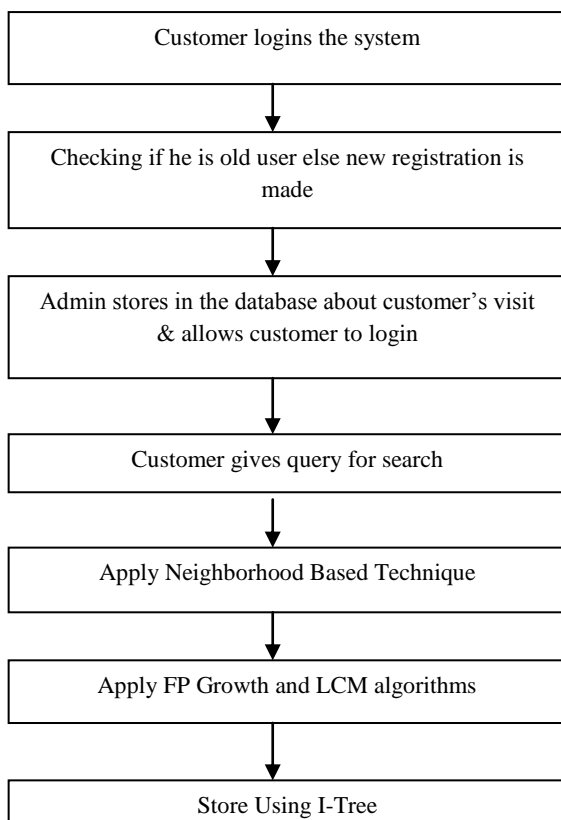


Fig 4: Framework for Proposed System

Then the frequent item set is extracted from the database by using FP-Growth algorithm. It is one of the algorithms in association rule mining. Tree structure of item sets in different transactions is used and pointer is used to attach the table to tree.

Next, LCM (Linear time Closed item set Miner) version 2 is used. It finds efficient frequent item set [20]. First, it reads the transactions to count item support. Then, for each transaction, it loads the subset including frequent items. Data are represented in memory by means of an array-based data structure, on which the extraction takes place. Above two algorithms are applied to get frequent item set [19].

The concept of I-tree is used to store transactional records. It provides a compact structure [18]. It uses prefix trees where each tree represents group of transactions. It has three layers namely top, middle and bottom layer.

7. EVALUATION

Here, we used datasets containing information about mobile products. System being developed is helpful to recommend mobile products to customers. Admin is created to maintain information about all mobiles and users.

If customers who have not yet registered approached this system he is first registered by getting his details. If the customer is already registered he can enter into the system directly by giving username and password. (Shown in Fig 5).

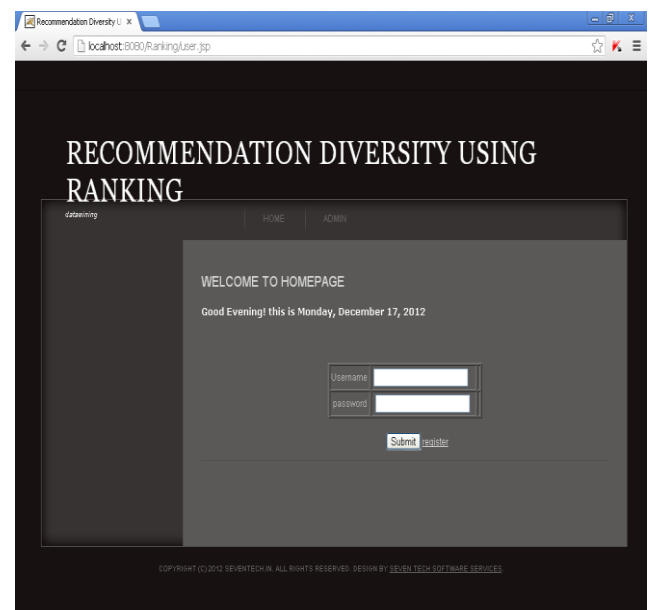


Fig 5: Homepage

Admin maintains the details in the datastore i.e., database and when a new customer registers or old customer logins, that information are stored, maintained and retrieved by the admin. When customer logins, information about them is cached by using index concept. (Shown in Fig 6).

System maintains the following details:

1. Number of times customer enters into the system per day/hour
2. What products he has searched for?
3. How much time he spent in the system?
4. What product he finally purchased?
5. What is the variety of products he searched?
6. What is the brand of product, time of purchase, day of purchase, price of product?
7. What rating he will give to that product?

8. Who and all already brought that product and when that is happened?

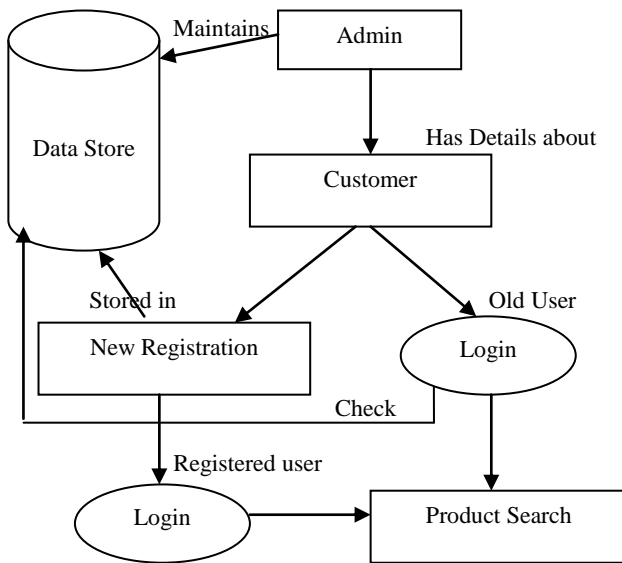


Fig 6: Admin Process

Then customer can get the details of the products (namely about mobiles here) and he can also view the purchase details of that products. Customer can also add information about mobiles.

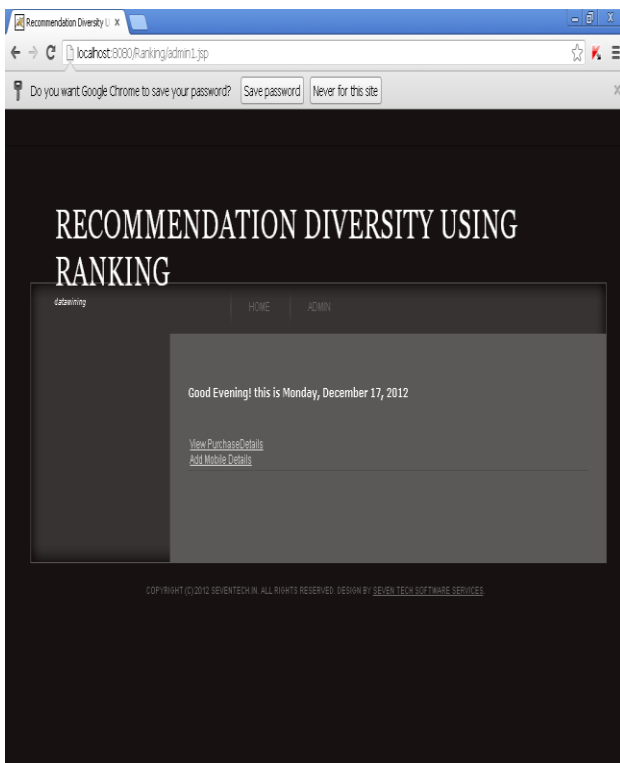


Fig 7: Information that customer uses

Then customer can view purchase details by giving mobile name, different types of brands, their price etc. If he is interested in that product by knowing its past history of purchases then may buy that product (Shown in Fig 8).

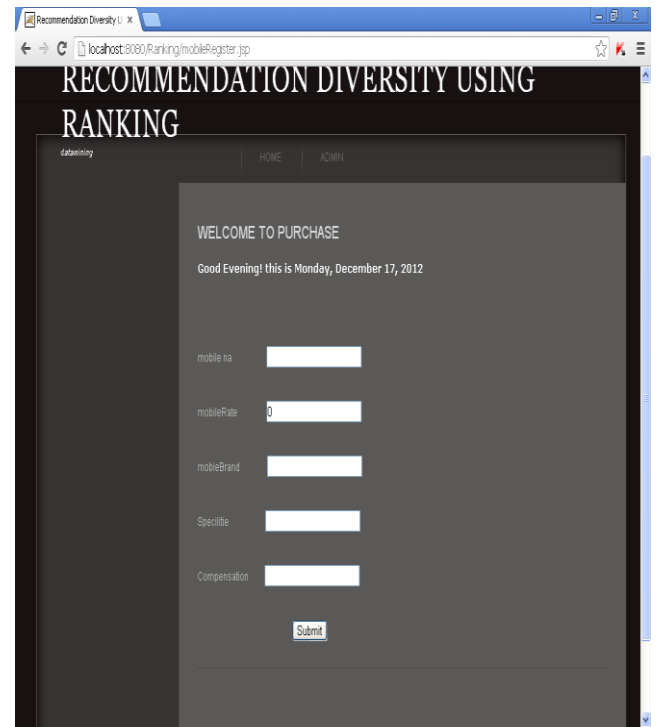


Fig 8: Purchase Details

The proposed algorithm maintains the details then-and-there i.e., it updates the database time-to-time. For every entry of customer into system database makes updates. It gives correct details since update is not done after specified amount of time.

8. RESULTS

Quality of recommendation increases when there is good and correct updates in database. The information it poses should be correct or else wrong recommendations will be given. Similarly, if specified time is fixed for making an update in database then new recommendations occurring during that time will be unknown to user approaching in that interval. User can know about those new recommendations only after the update happening after the specified time.

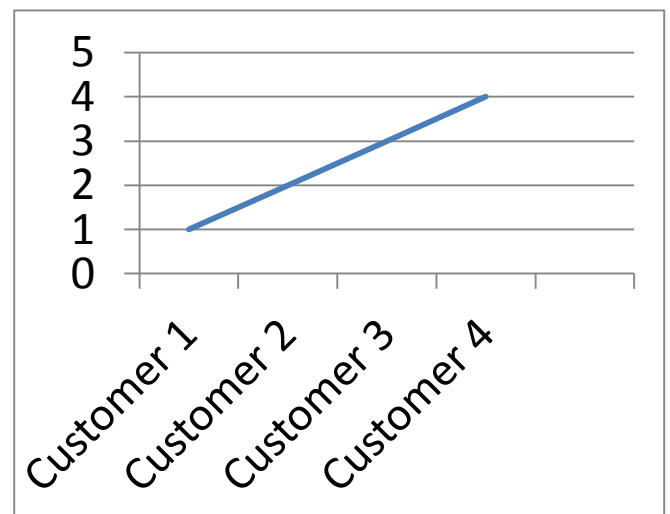


Fig 9: Update Time Vs Number of customers

New customer arrives update is done. This is shown in Fig 9.

9. CONCLUSION & FUTUREWORK

People search for products eagerly to buy good product. This is due to enormous production of large number of products on the world. Their decision to choose a product highly depends on word-of-mouth. Customers greatly observe the views of different people to make decisions. Recommender Systems help people a lot for making decisions. In the online world, Data is usually in the form of purchase information (i.e., what items customer has purchased), ratings given by user, purchase behavior of other customers etc. This makes recommender systems to help in E-Commerce. While preferring for E-commerce applications, these systems should be qualitative and be able to answer about huge variety of things. They have to work well by considering all parameters to give worthy recommendations. Diversity is the main issue. On applying the LCM and I-Tree, we can get time to time updates so admin can give correct recommendations. This helps people to get useful suggestions. Also, Accuracy should not be degraded. Diversity increased with less loss in accuracy. In future, I-Mine algorithm [18][19][21] can also be applied for database update. Greedy optimization techniques can be used for prediction of ratings and their ranking. Many re-ranking techniques specified in [22] can be applied additionally to rank a product. More-over, recommendations can be done based on reviews or suggestions of customer or manufacturer of products.

REFERENCES

- [1] "E-Commerce book", P.T. Joseph, Prentice Hall of India-New Delhi, 2009.
- [2] Ricci.F, Rokach.L "Recommender systems handbook", Springer, 1st edition, 2011.
- [3] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.
- [4] J. Ben Schafer, "The Application of Data-Mining to Recommender Systems" Encyclopedia of Data Warehousing and Mining, Second Edition, 2009.
- [5] G.Salton, Automatic Text Processing, Addison-Wesely, 1989.
- [6] B.P.S Murthi and S.Sarkar, "The role of the management Sciences in research on Personalization", Management Science, vol. 49, no. 10,pp. 1344-1362,2003.
- [7] Z. Huang, "Selectively Acquiring Ratings for Product Recommendation," Proc. Int'l Conf. Electronic Commerce, 2007.
- [8] Tao Zhou, zoltan Kuscsik, "Solving the apparent diversity-accuracy dilemma of recommender systems", proceedings of the national academy of Sciences of the USA 2010.
- [9] R. Garfinkel, R. Gopal, A. Tripathi, and F. Yin, "Design of a Shopbot and Recommender System for Bundle Purchases,"Decision Support Systems, vol. 42, no. 3, pp. 1974-1986, 2006.
- [10] Jennifer Golbeck," A framework for Recommending collections". Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011), held in conjunction with ACM RecSys 2011. October 23, 2011, Chicago, Illinois, USA.
- [11] Gediminas Adomavicius, YoungOk Kwon,"Overcoming Accuracy-Diversity Tradeoff in Recommender Systems: A Variance-Based Approach", Proceedings of 18th on Information Technology and Systems,WITS 2008,Paris, France, December 2008.
- [12] Hongzhi Yin Bin Cui,"Challenging the Long Tail Recommendation",Proc of VLDB Endowment,vol 5, No 9,Aug2012.
- [13] Gediminas Adomavicius, YoungOk Kwon," Maximizing Aggregate Recommendation Diversity: A Graph-Theoretic Approach" Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011), held in conjunction with ACM RecSys 2011. October 23, 2011, Chicago, Illinois, USA.
- [14] Cai-Nicolas Ziegler, Georg Lausen, "Making Product Recommendations More Diverse", In Proceedings of the 18th Workshop on Information Technology and Systems (WITS'08) (December 2008).
- [15] Fuguo Zhang, "Improving Recommendation Lists Through Neighbor diversification" IEEE International Conference on Intelligent Computing and Intelligent Systems, 2009. ICIS 2009.
- [16] Swapneel Sheth, jonathan Bell, "Towards Diversity in Recommendations using Social Networks", Columbia University Computer Science Technical Report, 2011.
- [17] Nagaraj. K, Saleem Malik. Kammadi, Somanath. J. Patil, Doddegowda. B. J and Sonali Shwetapadma Rath, "Diversity in Recommender System using Cross-Check Approach", International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2012) Penang, Malaysia.
- [18] T.Senthil Prakash, Dr.P.Thangaraj," IMine: Index Support for Item Set Mining in Item Set Extraction", International Conference on Advanced Computer Technology (ICACT), Proceedings published by International Journal of Computer Applications® (IJCA), 2011.
- [19] Thadi Ananda Ravi Kumar, N Tulasi Raju, D.Chitti Babu, Subhakar Rao Golla," Item Set Mining using IMINE Index Support ", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 1, Jan-Feb 2012, pp.189-194.
- [20] Takeaki Uno, Masashi Kiyomi," LCM ver. 2: Efficient Mining Algorithms for Frequent/Closed/Maximal Item sets".
- [21] T.Sunitha, G.Srujana, "IMine: Index Support for Item Set Mining", International Journal of computer Trends and Technology, July to Aug issue 2011.
- [22] Gediminas Adomavicius, "Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques", IEEE Transactions on Knowledge and Data Engineering', vol 24, No 5, May 2012.