ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 9, November 2012

Product Recommendation Techniques for Ecommerce - past, present and future

Shahab Saquib Sohail, Jamshed Siddiqui, Rashid Ali

***Abstract*— With the advent of emerging technologies and the rapid growth of Internet, the world is moving towards e-world where most of the things are digitized and available on a mouse click. Most of the commercial transactions are performed on Internet with the help of on-line shopping. The huge amount of data puts an extra overload to the user in performing on-line task. Product recommendation techniques are being used widely to reduce this extra overload and recommend the scrutinized product to the customers. Collaborative filtering, Association rules and web mining are on top amongst the techniques that is being used for recommendation technology. In this paper we try to give an overview of these recommendation techniques with suitable examples and illustrative diagrams, and change of trends in it with respect to time. Various diagrammatic representations are illustrated. Also a future direction of research in this area is indicated. And finally we conclude that there is a need of an extra effort to overcome limitations in existing techniques.**

***Index Terms*— Collaborative Filtering, E-commerce, Product Recommendation, Web Mining.**

I. INTRODUCTION

With the advent of emerging technologies and the rapid growth of Internet, the world is moving towards e-world where most of the things are digitized and available on a mouse click. Most of the commercial transactions are performed on Internet with the help of on-line shopping that makes e-commerce to become more popular. E-commerce is very much popular nowadays. Customers are buying more and more products on the Web and business organizations are selling more and more products on the Web. Whenever a user wants to buy a product on the Web, he visits an online store and looks for item of his interest. There are many popular e-commerce sites like ebay.com and amazon.com. Such online stores sell many items. For a single item, there are many brands and models available. The opportunity for the customer to select from a large number of products increases the burden of information processing before he decides which products meet his needs [1, 2]. If the customer is not sure about product of his choice, he may face the problem of information overload. He may come across a situation, where he may beunable to decide which product to buy. Whenever, a user visits a site and selects a product to buy, the sites recommend him some more products to buy. Product Recommender systems attempt to predict products in which a user might be interested, given some information about the product's and the user's profiles. Most existing recommender systems use collaborative filtering or content-based methods or hybrid methods that combine both techniques.

Collaborative filtering is considered to be one of the most successful product recommendation methods. Collaborative filtering identifies previous customers whose interests were similar to those of a given customer and recommends products to the given customer that was liked by previous customers. But, application of collaborative filtering to e-commerce has exposed some well-known limitations such as sparseness and scalability [3, 4].

As collaborative filtering requires explicit non-binary user ratings for similar products, the number of ratings already obtained is very small compared to the number of ratings that need to be predicted. Therefore, collaborative filtering based recommendations cannot accurately identify the products to recommend. Moreover, Algorithms to find the similar customers/products usually require very long computation time that grows linearly with both the number of customers and the number of products. With a large number of customers and products in a real world situation, collaborative filtering based product recommendations suffer serious scalability problems. These problems lead to poor recommendations. The quality of the recommendations has an important effect on the customer's future shopping behavior. If an online store recommends products that are not to be liked by the customer, customer may become angry and it is unlikely that he will visit the online store again [4]. If the online store target customers who are likely to buy recommended products and recommend products to only them, then this situation may be avoided.

Web mining is the application of data mining techniques to extract knowledge from the Web data [5]. Data mining refers to extracting unseen, hidden, novel and useful informative knowledge from a large amount of data. Web mining can be broadly divided into three distinct categories according to the kinds of data to be mined namely Web content mining, Web structure mining and Web usage

***ShahabSaquibSohail**, Department of Computer Science, Aligarh Muslim University., Aligarh, India, Mobile No, +91-9897771213, (e-mail: shahabsaquibsohail@gamil.com).*

***JamshedSiddiqui**, Department of Computer Science, Aligarh Muslim University., Aligarh, India, +91-9897267054, (e-mail: jamshed_faiza@rediffmail.com).*

***Rashid Ali**, Department of Computer Engineering, Aligarh Muslim University., Aligarh, India, +91-9411419429, (e-mail: rashidaliamu@rediffmail.com).*

mining. Web content mining is the process of extracting useful information from the contents of Web documents. Content data corresponds to the collection of information on a Web page, which is conveyed to users. It may consist of text, images, audio, video, or structured records such as lists and tables. Web structure mining is the process of discovering structure information from the Web. The structure of a typical Web graph consists of Web pages as nodes and hyperlinks as edges connecting related pages. Web usage mining is the application of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site.

Some of the studies have suggested web usage mining as an alternative to collaborative filtering since it will reduce the need for obtaining subjective user ratings or registration-based personal preferences [6, 7]. One of the e-commerce data is click stream that means visitor's path through a web site. Click stream in an online store provides information essential to understand shopping patterns or purchasing behaviors of customers such as what products they see, what products they add to the shopping cart, and what products they buy. Through analyzing such information (i.e., web usage mining), it is possible to make a more accurate analysis of customer's interest or preference across all products than analyzing the purchase records only.

Another approach of product recommendation is to benefit from the experience of the others. It is natural that whenever a person intends to buy an item, it takes views of his friends or relatives to select the brand and the model. The business companies also advertise their products highlighting their features. But, a normal person never blindly trusts these advertisements. In this era of e-commerce, customers are turning towards online opinions for the purpose. There are many online opinion resources such as online news, forums, blogs and reviews. Opinions are subjective statements that reflect user's sentiments or perceptions towards an item. It is possible that by reading other's posted opinions, a customer can take decision on buying a product. On the web, there are hundreds of opinion sources available. A user may like to search for opinions on a particular item by utilizing a search engine such as Google. But, the search engine may provide the user not only the desired information, but also a large amount of irrelevant information. Hence, the user again has to face the problem of information overload. Opinion mining is a subclass of Web content mining, where reviews of various products are mined to extract people's opinion. Opinion extraction allows Web users to retrieve and summarize people's opinions scattered over the Internet. Automated opinion mining can provide quick search [8] and analysis [9] results to both consumers and manufacturers [10].

II. BACKGROUND

A good number of studies have been performed in the area of product recommendation. In earlier works, collaborative filtering has been used successfully in a number of different applications such as recommending web pages, movies, articles and products [11-13]. Since, collaborative filtering has some major limitations, researchers investigated to use Web mining techniques for product recommendation.

In literature, we find that majority of works on product recommendation using Web mining techniques are based on Web usage mining. Web usage mining is the process of applying data mining techniques to the discovery of behavior or patterns from web data. The pattern discovery tasks include the discovery of association rules, sequential patterns, usage clusters, page clusters, user classifications or any other pattern discovery method [6, 7].

In [14], Cho et al. proposed a personalized recommendation system based on Web usage mining. They recommended products based on web usage data as well as product purchase data and customer related data. In [15], Kim et al. discussed personalized recommendation based on Web usage mining. Their method focused on the problem of helping customers to get recommendation only about the products they would like to buy. For this, they suggested a list of top-N recommended products for a customer at a particular time. They performed experiments with the Web usage data of a leading Internet shopping mall in Korea for the evaluation of their methodology. Experimentally, they deduced that choosing the right level of product taxonomy and the right customers increases the quality of recommendations.

In [16], Liu and Shih developed a product recommendation methodology that combined group decision-making and data mining techniques. They applied the analytic hierarchy process (AHP) to determine the relative weights of recency, frequency, monetary (RFM) variables in evaluating customer lifetime value or loyalty. They then applied clustering techniques to group customers on the basis of the weighted RFM value. Finally, product recommendations to each customer group were provided using association rule mining. They concluded that recommending more number of items helps to improve the quality of recommendation for more loyal customers, but not do so for less loyal customers.

Zeng also discussed the development of a personalized product recommendation system in [17]. The recommender system utilized the web mining techniques to trace the customer's shopping behavior from his/her click streams and learned his/her up-to-date preferences adaptively. Here, the customer preference and product association were automatically mined from click streams of customers. Then, the matching algorithm which combined the customer preference and product association was used to score each product. The system then produced the recommended product lists for a specific customer. Experimentally, they showed that their system provides sensible recommendations, and enabled customers to save enormous time for Internet shopping.

One of the earlier works on opinion extraction was reported by Hu and Liu in [18]. They considered three

things in opinion extraction namely (i) extraction of Subject (the product), (ii) aspect (the attributes or features), and (iii) semantic-orientation. Semantic-orientation was binary-valued either positive or negative. Popescu and Etzioni in [19] additionally annotated Hu and Liu's corpus with tags. In [20], Kobayashi et al. discussed how customer reviews in web documents can be best structured. They proposed to structure opinion unit as a quadruple, that is, the opinion holder, the subject being evaluated, the part or the attribute in which it is evaluated, and the evaluation that expresses positive or negative assessment. They used a machine learning-based method for opinion extraction.

Aciar et al. in [21] used prioritized consumer product reviews to make product recommendations. Using Web content mining (also, called sometimes text mining) techniques, they mapped each piece of each review comment automatically into an ontology. Scaffidi et al. implemented a prototype system called Red Opal [22] to score each product on each feature for the users to locate products rapidly based on features. Sun et al. in [23] proposed an automated system to compare and recommend products for customers from both subjective and objective perspectives. For subjective comparison of products, they used results of opinion mining. They also included product technical details to improve the comparison results from the objective perspective.

III. OVERVIEW

In this section we give brief description for the approaches available for product recommendation techniques and the details are elaborated in the next section.

A. *Association Rules*

Association rule technique is one of the traditional data mining techniques and proved to be very effective in recommendation technology [4]. In this technique we try to find the association between set of purchased items. It searches for interesting relationship among items in a given data set.

Rule support and confidence are two measures of rule of interestingness.

B. *Collaborative Filtering*

It is believed that the collaborative filtering is the most successful technology for being used as a recommendation system till early decade of the new millennium [21, 17]. A good number of successful recommendation techniques on the web use collaborative filtering. The basis of collaborative filtering (C.F) is the user's opinion. There are three main parts of a recommender system, as classified in [4].

1. Representation of data
2. Neighborhood formation
3. Recommendation generation.

In spite of the success of C.F there exist some major issues with it such as 'sparseness' and 'scalability'.

C. *Web Mining*

Web mining is defined on the basis of different approaches. There are two approaches; first one is "process-centric view" and the second one is "data-centric view". Web mining is defined as "sequence of task" on the basis of "process-centric view". The data-centric view defines web mining in terms of the types of web data being used.

IV. Collaborative Filtering

.

The collaborative filtering is the most commonly used recommender system. The products are recommended based on the opinions of other customers. This opinion includes the trends of a particular customer on several products and several customers on a particular product. These systems try to find the neighbor of an item. Neighbors are the customers that either rated different product in a same way as the target customer or they seem to show their affinity for a particular product same as the target.

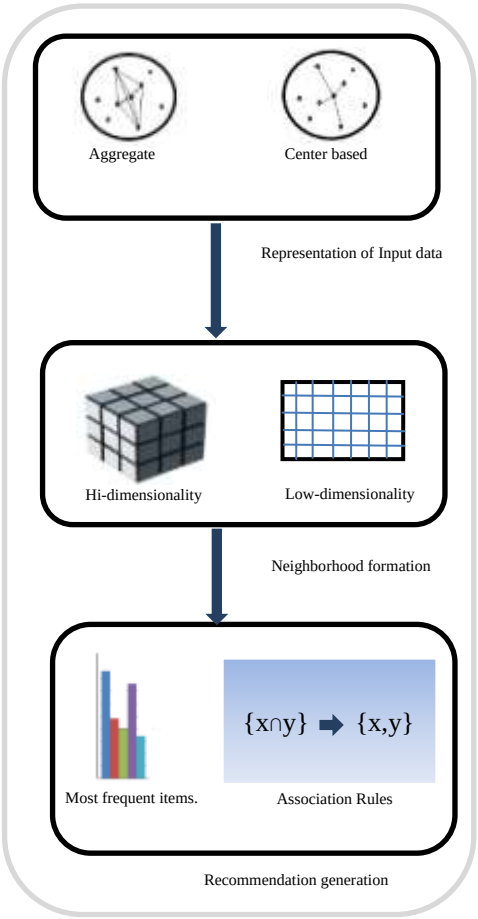

Fig.1 Main Components of recommendation system

Though C.F is a successful recommender system and widely being used, but still there are few major issues with this. One of the problems associated is Sparseness. [9, 27]

As collaborative filtering requires explicit non-binary user

ratings for similar products, the number of ratings already obtained is very small compared to the number of ratings that need to be predicted. Therefore, collaborative filtering based recommendations cannot accurately identify the products. It is evident that quality of recommendation plays very important role in identifying the customer's purchasing future behavior. If we do not use a good and reliable recommendation technique, there can be two major types of characteristics errors. One of the errors is "false negative". This is the error in which those items are missed to recommend which are likened by the customers. The second error is "false positive". There is a situation in which those products are recommended which customers do not like, and this is the worst condition as it irritates the customers and discourages them for any further purchasing. Another problem associated with C.F is scalability. As discussed in the previous sections that collaborative filtering uses neighbor algorithm that requires computations. And the computation increases proportionally to the number of customers and products both. In [4] Sarwar et al. divided the recommendation process in three tasks as discussed in Overview section. The figure (fig. 1) depicts their recommendation generation tasks. The first task is "representation" in which data is represented. They showed two approaches for representation; aggregate and center based respectively. Then the respective algorithm is used for neighborhood formation and can be reduced to low dimension from high dimension. Finally recommendation generation is performed by observing most frequent item and applying appropriate association rule.

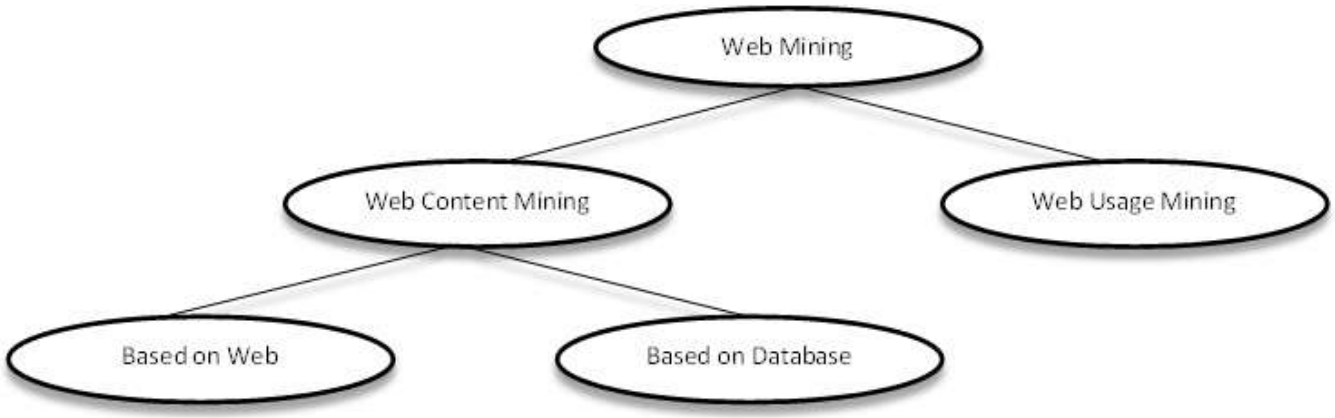

Fig 2. Web Mining Taxonomy

V. WEB MINING

A. *Taxonomy for Web mining*

Web mining technique is defined as the application of datamining technique to discover and extract information from web automatically [25]. Mobasher et al. in [24] categorize the web in two different category, web content mining and web usage mining. A similar taxonomy is represented in Fig 2.

B. *Web Content Mining*

It is the process of extracting useful information by analyzing the contents of the web. In [24], the author classified web content mining in database approach and agent based approach. The agent based approach were again divided in to three category, Intelligent search Agents, Information filtering and Personalized Web Agents. They also divided database approach into Multilevel database and Web Query Systems.

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 9, November 2012

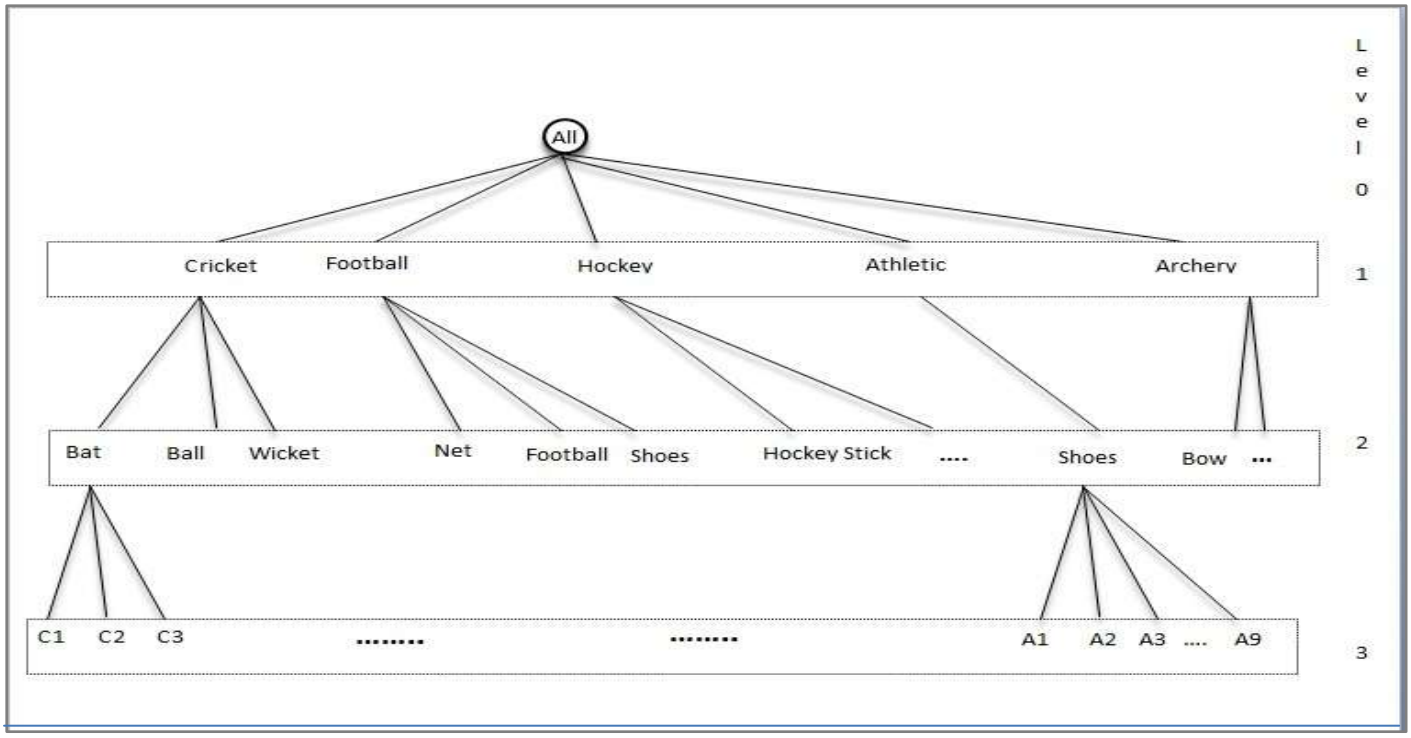

Fig 3. Product taxonomy for Web usage Mining

C. *Web Usage Mining*

Web Usage Mining is useful in predicting the user's behavior when they interact with the web. The data mined while interacting with the web are considered to be secondary data. [26.]

Web usage mining is concerned with the behavior of the customer that how they visit and what are the trends of their shopping, what are the products they visit before purchasing an item, and what are the items they purchase.If a customer purchases items A and B in the first week of consecutive months then it is probable that in next month that particular customer will purchase both the items. So these combinations are made with the help of web usage mining. Product taxonomy for web usage mining is shown in Fig 3.

If we consider the kind of data to be mined, we can categorize web mining in one more category, web structure mining. The taxonomy for this is depicted in Fig 4.

D. *Web Structure Mining*

Mining the web structure implies finding out the structure information of the web. It is considered as the process of extracting structure information of the web. If we draw graph for a typical web structure, it consists of web pages and hyperlinks as node and edges respectively.

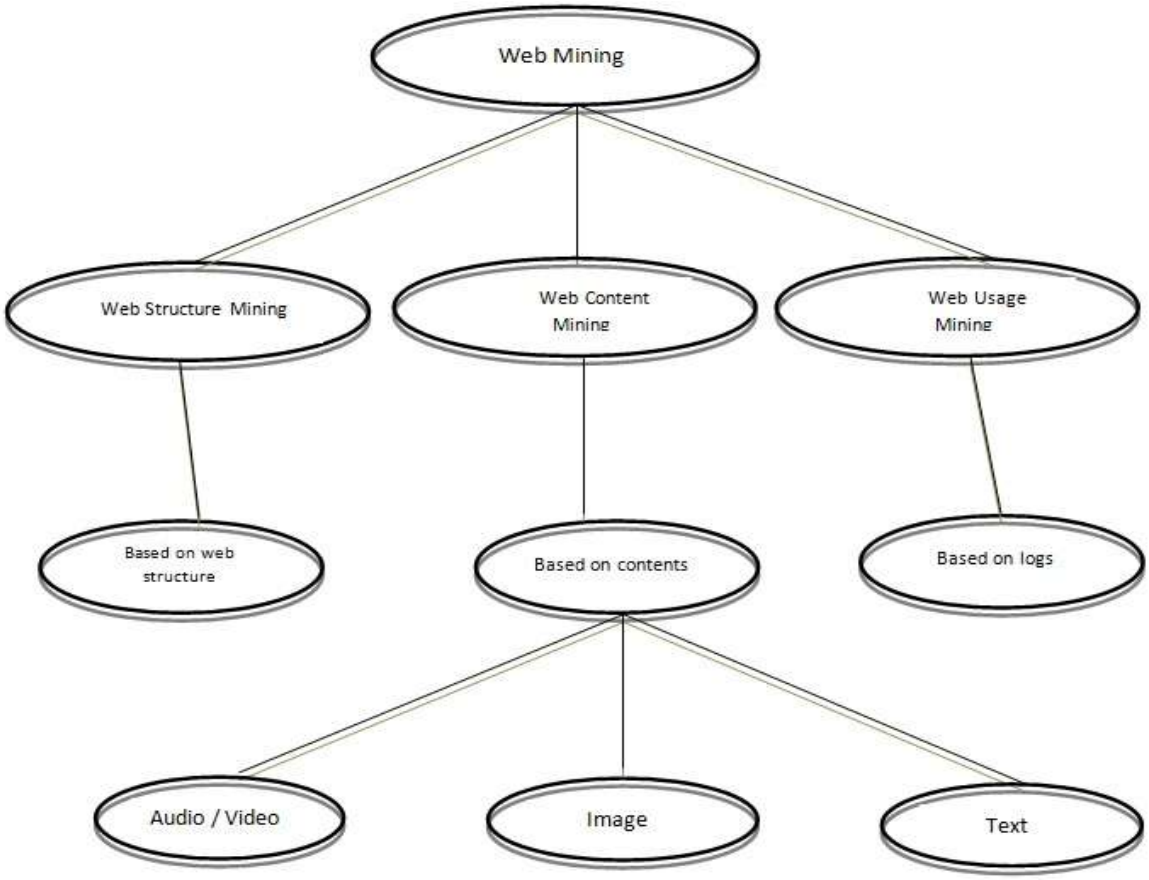

Fig 4. Web mining taxonomy based on data to be mined

VI. FUTURE DIRECTION

We have presented the overview and general approach of various recommendation techniques. There is a need of relative comparisons between these techniques over same data sets. Also one can tabulate the comparison of their respective characteristics. Also, one can categorize these techniques on the basis of a number of criteria.

VII. CONCLUSION

With an extra information overload over Internet, users need good and sound recommendation techniques. In this paper, we describe various recommendation techniques and briefly their advantages and limitations are elaborated. This gives a clear idea about the recommendation approaches and easy to understand the phenomena of recommendation, even for a native user.Finally we conclude that there is a need to make a lot of efforts to overcome the limitations of the existing techniques.

REFERENCES

[1] E. Kim, W. Kim, and Y. Lee (2000). "Purchase propensity prediction of EC customer by combining multiple classifier based on GA", In Proceedings of International Conference on Electronic Commerce 2000, pages 274–280.

[2] J. B. Schafer, J. A. Konstan and J, Riedl (2001). "E-commerce recommendation applications", Data Mining and Knowledge Discovery, volume 5, issue 1–2, pages 115–153.

[3] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes and M. Sartin (1999). "Combining content-based and collaborative filters in an online newspaper". In Proceedings of ACM SIGIR '99 Workshop on Recommender Systems, Berkeley, CA.

[4] B. Sarwar, G. Karypis, J. Konstan and J. Riedl (2000). "Analysis of recommendation algorithms for e-commerce", In Proceedings of ACM ECommerce 2000 Conference, pages 158–167.

[5] G. Xu (2008). "Web Mining Techniques for Recommendation and Personalization", PhD Thesis submitted to The School of Computer Science & Mathematics, Faculty of Health, Engineering & Science, Victoria University, Australia.

[6] B. Mobasher, R. Cooley, and J. Srivastava (2000). "Automatic personalization based on web usage mining", Communications of the ACM, volume 43, issue 8, pages 142–151.

[7] B. Mobasher, H. Dai, T. Luo, Y. Sun, and J. Zhu (2000). "Integrating web usage and content mining for more effective personalization", In Proceedings of the EC-Web 2000, pages 65–176.
[8] J. Liu, G. Wu and J. Yao (2006). "Opinion Searching in Multi-product Reviews", In proceedings of the sixth IEEE International Conference on Computer and Information Technology.
[9] B. Liu, M. Hu, and J. Cheng(2005). "Opinion observer: analyzing and comparing opinions on the Web", In Proceedings of the 14th international conference on WWW.
[10] P. Chaovalit and L. Zhou(2005). "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches", In Proceedings of 38th Hawaii International Conference on System Sciences, Big Island, HI, USA. IEEE Computer Society.
[11] W. Hill, L. Stead, M. Rosenstein and G. Furnas (1995). "Recommending and evaluating choices in a virtual community of use", In Proceedings of the 1995 ACM Conference on Human Factors in Computing Systems , pages 194–201.
[12] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl (1994). "Grouplens: An open architecture for collaborative filtering of netnews", In Proceedings of the ACM 1994 Conference on Computer Supported Cooperative, pages 175–186.
[13] U. Shardanand and P. Maes (1995). "Social information filtering: algorithms for automating word of mouth" In Proceedings of Conference on Human Factors in Computing Systems, pages 210–217.
[14] Y. H. Cho, J. K. Kim and S. H. Kim (2002)."A personalized recommender system based on web usage mining and decision tree induction", Expert Systems with Applications, volume 23, Elsevier Science, pages 329–342.
[15] J. K. Kim, Y. H. Cho, W. J. Kim, J. R. Kim and J. H. Suh (2002). "A personalized recommendation procedure for Internet shopping support", Electronic Commerce Research and Applications, volume 1, Elsevier Science, pages 301–313.
[16] D. R. Liu and Y.Y. Shih (2005). "Integrating AHP and data mining for product recommendation based on customer lifetime value", Information & Management, volume 42, Elsevier Science, pages 387–400.
[17] Z. Zeng (2009). "An Intelligent E-commerce Recommender System Based on Web Mining", International journal of business and management, volume 4, issue 7, 2009, pages 10-14.
[18] M. Hu and B. Liu (2004). "Mining and summarizing customer reviews", In Proceedings of the Tenth International Conference on Knowledge Discovery and Data Mining, pages 168-177.
[19] A. M. Popescu and O. Etzioni (2005). "Extracting product features and opinions from reviews", In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 339-346.
[20] N. Kobayashi, K. Inui and Y, Matsumoto (2007). "Opinion Mining from Web Documents: Extraction and Structurization", Information and Media Technologies, volume 2, issue 1, pages 326-337.
[21] S. Aciar, D. Zhang and S. Simoff and J. Debenham (2007)."Informed Recommender: Basing Recommendations on Consumer Product Reviews", IEEE Intelligent Systems, 2007, volume 22 Issue 3, pages 39 – 47.
[22] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng and C. Jin (2007). "Red Opal: Product-Feature Scoring from Reviews," In Proceedings of ACM EC, pages 182-191.
[23] J. Sun, C. Long, X. Zhu, and M. Huang (2009). "Mining Reviews for Product Comparison and Recommendation", Polibits, Research journal on Computer science and computer engineering with applications, volume 39, pages 33-40...
[24] R. Cooley, B. Mobasher, and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web", In support of NSF grant ASC 9554517.
[25] O. Etzioni, "The World Wide Web: quagmire or goldmine", Communication of ACM, 1996, 39 (11), pages 65-68C.
[26] R. Kosala, H. Blockeel, "Web mining research: A survey", ACM SIGKDD Exploration, vol. 2, issue-1, pp. 1-15, July 2000.

Shahab Saquib Sohail is a research scholar in the Department of Computer Science, A.M.U, Aligarh. He has completed his master degreein Computer Science and Applications (M.C.A) from Aligarh Muslim University (A.M.U). He is working with Dr. JamshedSiddiqui and Dr. Rashid Ali on Web Mining. His area of interest includes web mining, communication technology, security and hacking. He has research papers at International Conferences and journals of high repute such as IEExplore, etc.Mr. S.S.Sohail has attended several conferences and seminar and presented research papers there. He has also authored a book on security titled "Chaos-based Encryption".

Dr. Jamshed Siddiqui is an Associate Professor at Computer Science Department, Aligarh Muslim University, Aligarh, India. He holds Master's degree in Computer Science and obtained the degree of Ph. D. in Information Systems from Indian Institute of Technology, Roorkie. India. His research areas and special interests include Information Systems, MIS, Systems Analysis & Design, Knowledge Management Systems, E-Business, Data Mining and Parallel Computing. His areas of teaching interest includes Analysis and design of Information system, Software Engineering, Performance evaluation of computer systems, Computer oriented Numerical methods. He has published various papers in international journals and journals of international repute such as Journal of Information Technology, TQM Magazine, (Emerald Group Publishing Ltd.), Business Process Management Journal, (Emerald Group Publishing Ltd.), Journal of Information, Knowledge, and Management, Journal of Systems Management, International Journal of Services and Operations Management etc.

Dr. Rashid Ali obtained his B.Tech. andM.Tech. from A.M.U. Aligarh, India in 1999 and 2001 respectively. He obtained his PhD in Computer Engineering in February 2010 from A.M.U. Aligarh. His PhD work was on performance evaluation of Web Search Engines. He has authored about 75 papers in various International Journals and International conference proceedings. He has presented papers in many International conferences and has also chaired sessions in few International conferences. He has reviewed articles for some of the reputed International Journals and International conference proceedings. He has supervised 15 M.Techdissertations. Currently, he is supervising three PhD candidates. His research interests include Web-Searching, Web-Mining, soft computing (Rough-Set, Artificial Neural Networks, fuzzy logic etc.), and Image Retrieval Techniques.