ISSN: 2321-8169
1385 – 1395

A Survey on Personalized Recommendation Techniques

Saranya.K.G
Assistant Professor, Dept.of CSE
PSG College of Technology,
Coimbatore, India
saranyaa87@gmail.com

Dr.G.Sudha Sadasivam
Professor, Dept. of CSE
PSG College of Technology,
Coimbatore, India
sudhasadhasivam@yahoo.com

***Abstract*—** The large volume of information available on the web, the broad coverage of the web content, the phenomenal number of web users and their continued rapid growth has presented a major challenge to the web community. Identifying and accessing relevant information suited to individual needs is called personalized information retrieval and access. Recommendation systems are software agents that elicit the interests and preferences of individual users and make recommendations accordingly. Personalized recommendation systems are in need to provide proper recommendations based on user's requirements and preferences. This paper presents an overview of personalized recommendation techniques and identifies issues and describes different approaches for personalization. The goal of this survey is to present a study on the main concepts, approaches and practices in the area of personalized recommendation systems. As a result this paper is concluded by presenting a number of possible research directions.

Keywords- *Personalized recommendation process, Filtering techniques, Mining technique, Ontology.*

*****

I. INTRODUCTION

The main objective of Information Retrieval system is to mine and extract relevant information from a large data set. Current information retrieval system have to deal with heterogeneous, high volume, continuously changing nature of information's. Personalization can be defined as any set of actions that can tailor the Web experience to a particular user or set of users. The actions can range from simply making the presentation more pleasing to anticipating the needs of a user and providing customized and relevant information. To achieve effective personalization, organizations must rely on all available data, including the usage and click-stream data, the site content, the site structure, domain knowledge, as well as user demographics and profiles. Efficient and intelligent techniques are needed to mine data for actionable knowledge, and to effectively use the discovered knowledge to enhance the users' Web experience. In web personalization challenges include, scalability, integration of heterogeneous data, information retrieval and filtering, knowledge representation, information security and privacy, user modelling. Recommender systems represent one special and prominent class of such personalized Web applications, which particularly focus on the user-dependent filtering and selection of relevant information. Personalization in web search engines can be achieved with the help of Query adaptation, result adaptation or combination of both query and result adaptation. This paper focuses on techniques related to result adaptation in a recommender systems.

This paper is organized in the following way. Section 2 describes the process involved in personalization task. Section 3 describes user profile construction. Section 4 describes about classification of personalization systems. Section 5 describes different personalization techniques used in information search and retrieval process. Different evaluation approaches of personalized systems are described in section 6.

II. PROCESS OF PERSONALIZED RECOMMENDER SYSTEMS

Recommendation systems have the potential to support and improve the quality of the decisions users make while searching in the internet. Personalized recommender systems can deliver tailored service in a way that will be most appropriate and valuable to the user. Process of personalized recommendation can be divided into four distinct phases as listed below [1,2].

A. Extraction phase

From web repositories, information based on user interest is extracted. This personalized information can be as follows.

- Content data like text, images, etc, from heterogeneous sources.
- Structure data representing linkage between web pages.
- Usage data maintained in web server logs.

User Profiles maintained statically or dynamically.

B. Preprocessing phase

It is done to make it compatible with the analysis technique.

C. Analysis phase

This step applies machine learning techniques to discover interesting usage patterns and statistical correlations between web pages and user groups. This step frequently results in automatic user profiling. Collaborative filtering technique is one such approach used to extract information's personalized for a group.

D. Recommendation

This phase in personalization makes use of the results of the previous analysis step to deliver recommendations to the user. For personalization, user profile has to be maintained. Overall architecture of the personalized recommender system is depicted in the following figure (figure 1).

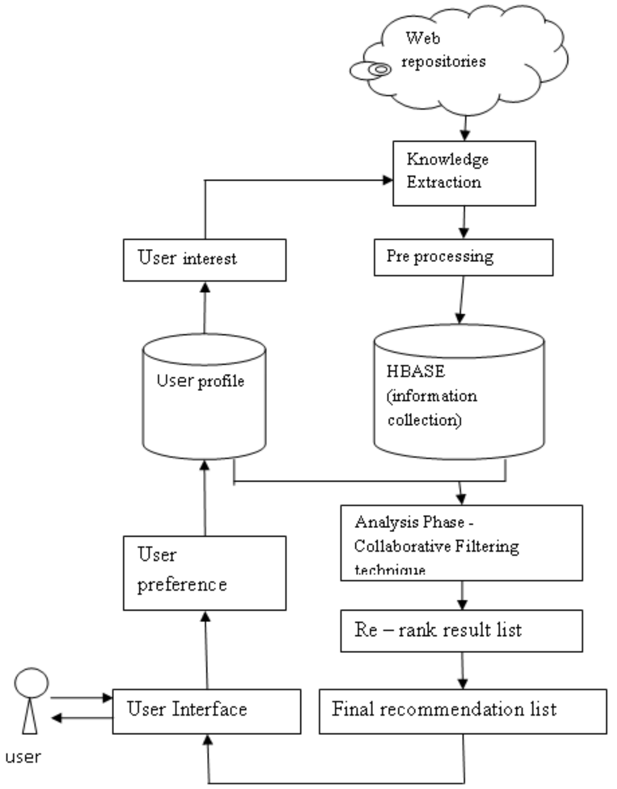

Figure.1 Architecture of personalized recommender system

The most important aspect of personalization is construction and maintenance of user profile. A user profile represents a collection of personal data associated with a specific user. There are four different design patterns for user models as listed below [35,36,37,38,39]:

1. *Static user profiles:* As the name indicates, once the main profile is built, it is not changed. Hence shifts in users' preferences are not registered. No learning algorithms are used to alter the model. [35,36]

2. *Dynamic user profiles:* Dynamic user models maintain up to date representation of users interests and interactions with the system. The user models are updated and hence consider the current needs and goals of the users [37].

3. *Stereotype based user profiles:* Stereotype based user models are based on demographic statistics. Based on the gathered information users are classified into common stereotypes. The system then adapts to this stereotype. Stereotype based user models rely on statistics. The main problem in this approach is that personal attributes might not always match the stereotype. However, they allow predictions about a user with very little information [38].

4. *Highly adaptive user profiles:* In contrast to stereotype based user models they do not rely on demographic statistics but aim to find a specific solution for each user. Although users can take great benefit from this high adaptivity, this kind of model needs to gather a lot of information first [39]. For construction of user profile, user information has to be collected [34,39]:

The process of collection of raw information about the user can be explicit or implicit. In explicit information collection, the user personally customizes the information source before the personalization starts. This approach results in static user profile. Implicit information collection is done by the system. It includes web usage mining based on user click throughs, browsing history, queries, user location, cookies and session id's. This approach results in dynamic user profile. Once the information is collected, user profile is constructed by analyzing and processing stored data. The compiled user profile is used in the actual web service (Figure. 2).

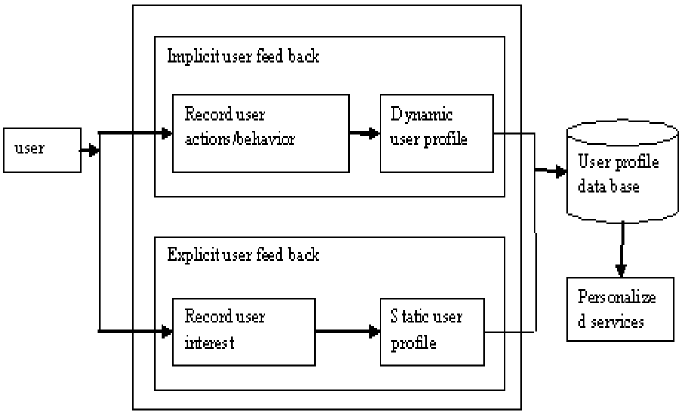

Figure 2. User Profile Construction.

As the vast majority of user's are reluctant to provide any explicit feedback on search results and their interests, many works on personalized web search focus on automatically learning user preferences without involving any direct user efforts[38]. Collected user information is processed and organized as a user profile in a certain structure depending on the need of personalization algorithm.

III. CLASSIFICATION OF PERSONALIZED RECOMMENDER SYSTEMS

Personalized recommender systems can be classified based on group size, type of user interaction, place where user profiles are maintained as individual vs. Collaborative, Reactive vs. Proactive, Client side vs. Server side.

A. *Group size*

Extreme or group personalization can be achieved based on group size. In Individual personalization, the system may choose to build an individual model of user likes and dislikes and use this profile to predict future interactions with the user. NewsWeeder is an example of such a system[118].Collaborative approach not only use the profile for the active user but also uses neighborhood users with similar preferences for recommendation. Group Lens is an example of such a system[118].

B. *Type of user interaction*

User interactions can be reactive or proactive in nature. Reactive approaches use personalization as the conventional process that requires explicit interactions with the user in the

 ISSN: 2321-8169
1385 – 1395

form of user queries and feedback. Examples of such personalized systems include Entrée, DIETORECS, and Expert Clerk [117]. Proactive approaches, on the other hand, learn user preferences and provide recommendations based on the learned information. User need not provide explicit feed back to the system to drive the recommendation process. Example of such personalized system include Genesys [116].

C. Location of user profile:

User profile can be maintained on client-side or server side. For client-side personalization, user information is collected and stored on the client side by installing a client software as depicted in figure.3. The user model is rich as the user's search behavior, contextual activities and personal information can be incorporated into the user profile. [8]. Server logs are nothing but previously visited URL's, and cookies.

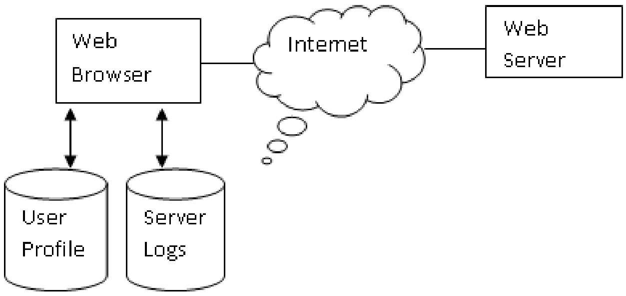

Figure.3. Client side personalization system.

For server-side personalization, user profiles are built, updated, and stored on the search engine. User information is directly incorporated into the ranking process, to help process initial search results (Figure 4) [6].

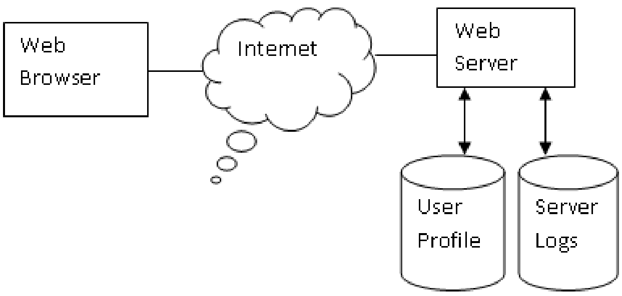

Figure 4. Server side Personalization System.

The main aim of personalization is to improve relevancy of recommendation by filtering relevant information from large data collections that are apt for the user. For this purpose personalization technique discussed in the following section are used.

IV. PERSONALIZATION TECHNIQUES

Personalization can be done using data mining approaches, filtering techniques, soft computing models or ontology models (Figure. 5) as described in this section.

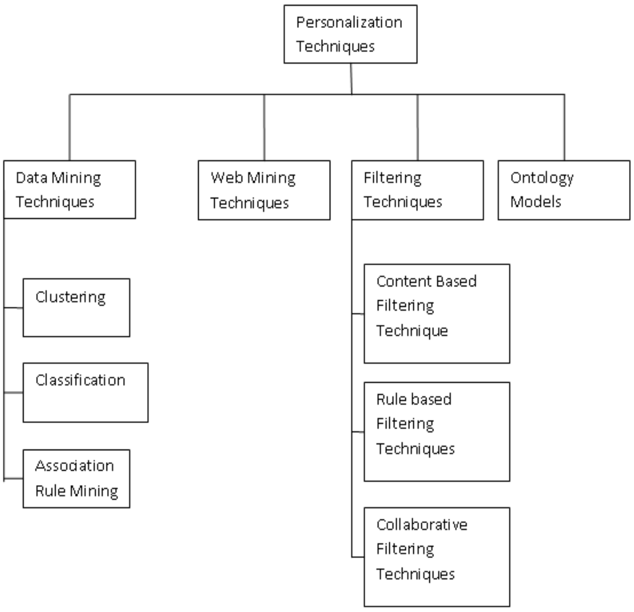

Figure 5. Taxonomy of Personalization techniques

1. Data Mining Techniques:

Data mining is a technique used to draw useful information from a large database to aid discovery of knowledge. Some of the approaches that can be used for personalization include clustering, association rule mining and classification [10,41, 42]..

Clustering aims to divide a data set into groups or clusters where inter-cluster similarities are minimized while the similarities within each cluster are maximized. It can be used to identify group of users with common interest based on user's preferences. Personalized recommendations can be offered to members of the same peer group [43].

A popular data mining technique used in the implementation of recommender systems is the generation of association rules which define item-to-item correlations by associating objects in a database that share a particular relationship [44]. Association rule-based systems have difficulty producing recommendations when the database is sparse as larger item sets are unable to meet minimum support constraints. To alleviate this issue, dimensionality reduction techniques can be applied to the dataset [45]. Other techniques to help resolve issues with sparsity are to rank the discovered association rules according to the degree of intersection between the left-hand-side of the rule, and the user's current session, or by using a hybrid system that combines association rules with collaborative filtering, where recommendations are only based on the target user's neighborhood [46].

Clustering can analyze user patterns and form peer groups, while association rules can discover patterns between items, Classification is able to incorporate all of this information and more in making a recommendation [47]. Classification relies on a set of "training data". Each category based on browsing history, demographics of consumer is given a label and this is used to train the model. Once each label is optimized through the training steps, real user data is input and the appropriate recommendations are given [48].

ISSN: 2321-8169
1385 – 1395

2. *Web Mining Techniques*

Web mining is the process of extracting interesting patterns from web information repositories [43]. The various techniques used for web mining is shown in figure 6. Web content mining is classified into two categories namely web page mining and web search results mining. Web page mining is used to discover patterns directly from the contents of web pages [49,50]. Web structure mining reveals the structure of web sites and how they are connected. Links in web structure mining are classified as internal link and external link [56].

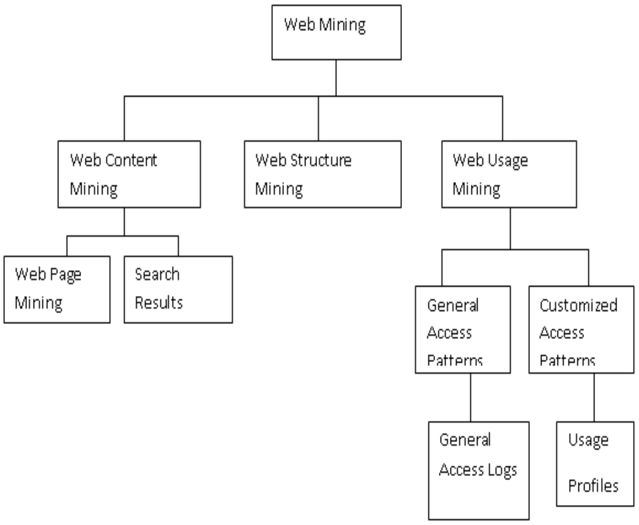

Figure 6. Taxonomy of Web Mining Techniques.

Web usage mining techniques read server log files to extract web site patterns [54]. Hence this technique is used to discover interesting user navigation patterns and can be applied to many real world problems like, improving web sites/pages, making additional topic or product recommendations and user behavior studies [56]. Web usage mining in the context of personalization involves data preparation and transformation, pattern discovery, and recommendation generation [55]. Speed Tracer [17] identifies user browsing pattern by investigating the web server log files with data mining procedures. Frequent Pattern (FP) trees [21] can discover users' frequent browsing patterns of users.

3. *Ontology Based Personalization Technique*

Ontology [62] is a formal description and specification of knowledge as a set of concepts and their relationships within the domain. It provides a common understanding of topics to be communicated between users and systems [63]. Personalized ontology's [64] are a conceptualization model that formally describes and specifies user background knowledge. Web users might have different expectations for the same search query [66]. Ontology-based user modeling system [65,67] integrates user ontology that specifies users and their relationships, domain ontology that captures application specific concepts and their relationships and log ontology that specifies the semantics of the user interaction with the system. Acharyya and Ghosh [68] propose a general personalization framework based on the conceptual modeling of the users' navigational behavior. The proposed methodology involves mapping each visited page to a topic, imposing a tree hierarchy on topics, and then estimating the parameters of a semi- markov process defined on this tree based on the observed user paths. Semantic characterization of the context is performed manually.

4. *Filtering Techniques*

Filtering techniques are a technique used to remove unwanted data from large volume of data using automated methods prior to recommend the data to the end user.

A. *Content Based Filtering Technique:*

Content-based filtering systems are solely based on individual users' preferences [51]. The system tracks each user's behavior and recommends them that are similar to the user likes in the past. Each object in the database is represented by the set of features and attributes that characterize that item. For each user, a profile is generated which includes descriptions of items based on the user interests. By using vector similarities such as cosine similarity, a prediction can be made on a user's interest would be on a particular item [52].
Some of the challenges in Content Based Filtering includes:

- Difficulty in Attribute identification for items is difficult.
- Obtaining the user feed back.
- Overspecialization problem, as recommendations cannot be given for unknown terms.

B. *Rule Based Filtering Technique*

A decision tree is constructed to represent various rules. In rule-based filtering the users are asked to answer a set of questions derived from a decision tree. This approach is unable to make recommendations for patterns that do not appear in association rules. Accuracy of recommendation can be improved by combining content-based, rule-based, and collaborative filtering techniques.

C. *Collaborative Filtering Techniques*

This technique predicts the opinion of the user and recommends items based on the users opinions and the opinions of the other like minded users. Collaborative filtering techniques are classified as follows.(Figure.7).

i) Memory Based Collaborative Filtering (Cf) Techniques

Memory-based CF algorithms use the user-item database to generate a prediction. Every user is part of a group of people with similar interests. By identifying the neighbors of a user, user preferences can be predicted. Memory based collaborative filtering techniques can be classified as follows.

a) The neighborhood-based CF where, a subset of nearest neighbors of the active user are chosen based on their similarity, and a weighted aggregate of their ratings is used to generate predictions for the active user [92].

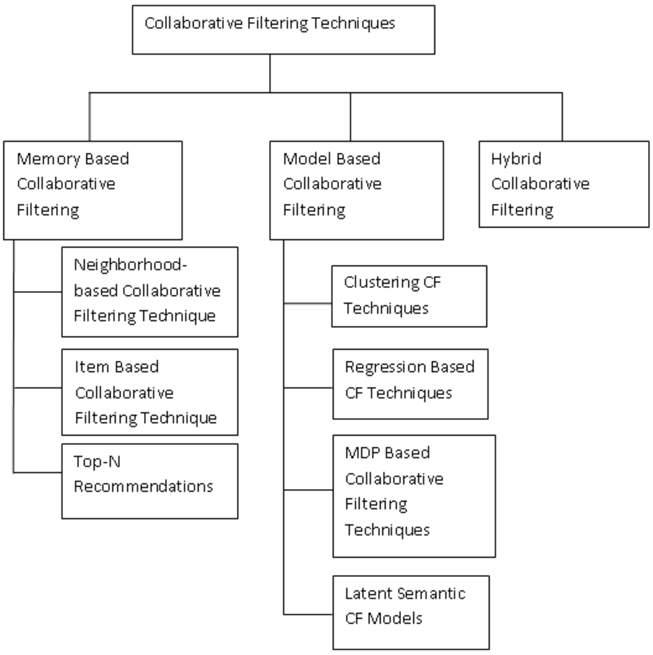

Figure 7 Taxonomy of collaborative filtering techniques

b) Item-based CF approach where, user rating of items is used to find their similarity to the target item. Prediction is computed by taking a weighted average on the target users rating on the most similar items. Similarity between the items is identified based on their Correlation, Vector Cosine, and Probability measures. [95].

c) Top-N Recommendations where, a set of N top-ranked items are recommended based on user interest. This technique analyzes the user-item matrix to discover relations between different users or items and provides suitable recommendations [96].

ii. Model Based Collaborative Filtering Techniques

Models can be trained to recognize complex patterns and then make intelligent predictions for the collaborative filtering tasks based on their learning[101]. The following approaches can be used.

A. Clustering based CF Algorithms.

A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [102]. The measurement of the similarity between objects is determined using metrics such as Minkowski distance and Pearson correlation. Clustering methods can be classified into three categories: partitioning methods, density-based methods [103], and hierarchical methods [104]. A commonly-used partitioning method, k-means[105], has two main advantages namely, efficiency and ease of implementation. Density-based clustering methods typically search for dense clusters of objects separated by sparse regions that represent noise [103]. DBSCAN [106] and OPTICS [80] are well-known density-based clustering methods. Hierarchical clustering methods, such as BIRCH [81], create a hierarchical decomposition of the set of data objects.

B. Regression-Based CF Algorithms

A regression method uses an approximation of the ratings to make predictions based on a regression model.

C. Markov Decision Process (Mdp) -Based Cf Algorithms.

An MDP is a model for sequential stochastic decision problems, which uses an agent to influence its surrounding environment through actions [104]. An MDP can be defined as a four-tuple: S,A, R, Pr, where S is a set of states, A is a set of actions, R is a real-valued reward function for each state/action pair, and Pr is the transition probability between every pair of states given each action.

D. Latent Semantic CF Models

A Latent semantic CF technique relies on a statistical modeling technique that introduces latent class variables in a mixture model setting to discover user communities and prototypical interest profiles. Conceptionally, it decomposes user preferences using overlapping user communities. The main advantages of this technique is its higher accuracy and scalability [109,110].

E. Other Model-Based CF Techniques

Association rule based CF algorithms [111] are used for top-N recommendation tasks when compared to prediction. Top-N items are chosen by identifying rules with sufficient support and confidence values and sorting them in a descending order [112]. Maximum entropy approach clusters the data, and then uses maximum entropy in a cluster to make predictions [113]. A dependency network is a graphical model for probabilistic relationships, whose graph is potentially cyclic. The probability component of a dependency network is a set of conditional distributions, one for each node given its parents. Although less accurate than Bayesian belief nets, dependency networks are faster in generating predictions and require less time and memory to learn[114]. Decision tree CF models treat collaborative filtering as a classification task and use decision tree as the classifier [109]. Horting is a graph-based technique in which nodes are users and edges between nodes are degrees of similarity between users [109]. Multiple multiplicative factor models (MMFs) are a class of causal, discrete latent variable models combining factor distributions multiplicatively and are able to readily accommodate missing data [110]. Probabilistic principal components analysis (pPCA) [112] determines the principal axes of a set of observed data vectors through maximum-likelihood estimation of parameters in a latent variable model closely related to factor analysis. Matrix factorization based CF algorithms have been proven to be effective to address the scalability and sparsity challenges of CF tasks [112,113].

iii. Hybrid Collaborative Filtering Technique

Hybrid CF systems combine CF with other recommendation techniques like content-based systems to make predictions or recommendations [113]. The content-boosted CF algorithm uses naive bayes as the content classifier, it then fills in the missing values of the rating matrix with the predictions of the content predictor to form a pseudo rating matrix, in which observed ratings are kept untouched and missing ratings are

 ISSN: 2321-8169
1385 – 1395

replaced by the predictions of a content predictor. It then makes predictions over the resulting pseudo ratings matrix using a weighted Pearson correlation coefficient based on number of active users[113] The content boosted CF recommender has better prediction performance when compared to pure content-based recommenders and pure memory-based CF approaches. It also overcomes the cold start problem and tackles the sparsity problem of CF tasks. Probabilistic memory-based collaborative filtering (PMCF) combines memory-based and model-based techniques [114]. A mixture model built based on user profiles and the posterior distribution of user ratings is used to make predictions. Personality diagnosis (PD) is a representative hybrid CF approach that combines the advantages of memory-based and model-based CF approaches [115].

Evolutionary algorithms (EA) belongs to a family of iterative stochastic search and optimization methods based on mimicking successful optimization strategies observed in nature [70,71,72,73]. Together with fuzzy sets, neural net-works and fractals, evolutionary algorithms are among the fundamental members of the class of soft computing methods [72]. EA operate with population of artificial individuals (referred often as items or chromosomes) encoding possible problem solutions. Objective function are used for evaluating encoded individuals, which assigns a value to each individual. The quality (ranking) of each individual is represented by fitness value, as solution of given problem. Competing individuals search the problem domain towards optimal solution [73]. Evolutionary approaches can be used for optimization in large search spaces for various personalization techniques.

TABLE.1. Comparison of different recommendation techniques.

S.No	Example Recommender System	Technique used	Advantages	Disadvantages
1.	DBSCAN [106],OPTICS [80],BIRCH [81].	Clustering Algorithms	• Faster Recommendation. • Better performance	• Not accurate
2.	MovieLens [119].	Association rules	• Provide accurate Prediction about the user interest. • Fast to implement. • Not much storage space required. • Fast to execute.	• Difficulty in producing the recommendations when the data base is sparse. • Not suitable if preferences change rapidly. • Rules can be used only when enough data validates them.
3.	Levis [92], Netflix [95].	Memory Based Recommender Systems.	• Easy implementation • New data can be added easily and incrementally • Need not consider the content of the items being recommended. • Scale well with correlated items	• Dependent on human ratings • Performance decrease when data are sparse. • Cannot recommend for new users and items • Limited scalability
4.	GroupLens [93].	Model based Collaborative Filtering.	• Better at addressing the sparsity and scalability • Improve prediction performance	• Expensive model building. • Lose useful information for dimensionality reduction.
5.	TAN-ELR, NB-ELR [121].	Classification Technique	• Accurate. • Deals Sparsity problem effectively	• Scalability. • Time Consuming in model training.
6.	Personality Diagnosis [115].	Hybrid Recommenders	• Improved prediction performance. • Overcomes sparsity problem.	• Increased complexity. • Expensive. • Need external information that usually not available.

ISSN: 2321-8169
1385 – 1395

7.	Newsweeder [118],LIBRA	Content-Based Filtering	• Efficiently address Cold start problem and Sparsity Problem	• Requires content that can be encoded as meaningful features. • User interest must be represented as a learnable function. • Overspecialization problem.
9.	ATHENA [106].	Ontology	• No cold start issue. • More accurate	• Knowledge engineering is required. • Expert opinion may not match with user preferences.
10.	PEN recsys [120].	Web usage mining	• More accurate	• Much storage space is required.

V. PERFORMANCE MEASURES

Performance of the personalization technique is evaluated based on the following measures.

A. System performance

System performance is concerned with measuring retrieval effectiveness. It can be measured with the help of following metrics.

1. *Precision*, is the ratio of number of retrieved relevant documents over the total number of retrieved documents.

a. Precision at K, which measures the fraction of retrieved relevant documents within the top K retrieved documents.

b. Mean Average Precision (MAP), It is the average precision at K values computed after each relevant document has been retrieved for a query, where the mean of all these averages is calculated across all the test queries.

c. R-precision, measures precision with respect to a given number of documents that are known to be relevant.

d. 11-point Precision, measures the precision of retrieved results at 11 fixed values of recall.

2. *Recall*, which is the number of relevant documents that are retrieved over the total number of known relevant documents in the document collection.

a. Recall at K, which measures the fraction of retrieved relevant documents within the top K documents over the total number of relevant documents in the document collection.

3. *F-Measure*, is the weighted harmonic mean of precision and recall.

4. *Normalised discounted Cumulative Gain (NDCG)*, is a precision metric that is designed for experiments where documents are judged using a non-binary relevance scale like highly relevant, relevant, or not relevant. It gives higher scores for more relevant documents being ranked higher in the ranked list of results.

5. *Break-even Point*, which is determining the point at which precision equals recall;

B. *Usability*

Usability is concerned with the user's perception of the system. This aspect of evaluation measures the degree of user satisfaction with respect to the adaptive service. This type of evaluation is hard to standardize across different systems as it is subject to user bias. It can be qualitatively evaluated using usability questionnaires [75]. It can also be quantitatively evaluated by measuring the user's performance in fulfilling certain tasks using the system, for example by keeping track of the time and number of actions needed to complete the task.

C. *Accuracy*

Accuracy is concerned with the ability of the the system to implicitly infer and represent user information, such as the user's interests. It can be measured with the help of Receiver Operating Curve (ROC). The ROC- curve is a plot of the systems sensitivity by the complement of its specificity. Generally to compare the personalization accuracy in two systems, the size of the area under the ROC- curve is measured. The curve with a larger value indicates better performance [78].

VI. CHALLENGES

Some of the challenges of personalization include,

- Clustering users into user communities..
- Handling heterogeneous data sources.
- Identifying new metrics to evaluate the performance of recommender systems.
- Maintaining large amount of frequently changing information.
- Improving accuracy and relevancy of new recommendation.

hybrid collaborative filtering technique can be used to take care of user and group interest. This can improve relevancy of retrieved information based on user interest. To handle large volume of heterogeneous, continuously changing information, big data platforms like hadoop and NoSQL data bases like HBase can be used.

VII. CONCLUSION

Information retrieval systems are used in every field and personalization improves relevancy of extracted information. This paper has presented an overview of recommendation systems and personalization techniques. Several recommendation systems have been designed based on collaborative filtering, content based filtering and hybrid recommendation methods. Their complementary advantageous and disadvantageous has been discussed. Personalized recommender systems are an active area of research. The need for improvement in accuracy and handling large volume of heterogeneous information necessitates the use of specialized big data platforms with suitable mining techniques for personalization.

References

[1] Schafer J.B., Konstan J., and Reidel J. (1999). Recommender Systems in E-Commerce, In Proc. ACM Conf. E-commerce, 158-166.

[2] Mobasher B., Dai H., Luo T., and Nakagawa M. (2001). Effective personalizaton based on association rule discovery from Web usage data, ACM Workshop on Web information and data management, Atlanta, GA.

[3] Spiliopoulou M. and Faulstich L. C. (1999). WUM: A Web utilization Miner, in Proc. of EDBT workshop WebDB98, Valencia, Spain.

[4] Nasraoui O., Cardona C., Rojas C., and Gonzalez F. (2003). Mining Evolving User Profiles in Noisy Web Clickstream Data with a Scalable Immune System Clustering Algorithm, in Proc. of WebKDD 2003 – KDD Workshop on Web mining as a Premise to Effective and Intelligent Web Applications, Washington DC, 71-81.

[5] Srivastava, J., Cooley, R., Deshpande, M., And Tan, P-N. (2000). Web usage mining: Discovery and applications of usage patterns from web data, SIGKDD Explorations, 1(2), 12-23.

[6] Abhinandan Das, Mayur Datar, Ashutosh Garg, "Google News Personalization: Scalable Online Collaborative Filtering" ACM 978-1-59593-654-7/07/0005, 2007, Canada..

[7] Nasraoui O., Cardona C., Rojas C., and Gonzalez F. (2003). Mining Evolving User Profiles in Noisy Web Clickstream Data with a Scalable Immune System Clustering Algorithm, in Proc. of WebKDD 2003 – KDD Workshop on Web mining as a Premise to Effective and Intelligent Web Applications, Washington DC, 71-81.

[8] Mikhail Bilenko , Matthew Richardson , " Predictive Client side profiles for Personalized advertizing" , KDD"11, ACM 2011, August 24, San Diego, California, USA.

[9] OARD, D. W. & DIEKEMA, A. R. 1998. Cross-Language Information Retrieval. In: WILLIAMS, M. (ed.) Annual Review of Information Science (ARIST), pp. 472-483.

[10] OARD, D. W. 2010. Multilingual Information Access. Encyclopedia of Library and Information Sciences, 3rd Edition**,** 3682 - 3687.

[11] FURNAS, G. W., LANDAUER, T. K., GOMEZ, L. M. & DUMAIS, S. T. 1987. The Vocabulary Problem in Human-System Communication. Communications of the ACM, 30**,** 964-971.

[12] MANNING, C. D., RAGHAVAN, P. & SCHUTZE, H. 2008. Introduction to Information Retrieval, Cambridge University Press.

[13] ZHOU, D., LAWLESS, S. & WADE, V. 2012. Improving search via personalized query expansion using social media. Information Retrieval**,** 1-25.

[14] CHIRITA, P.-A., FIRAN, C., S. & NEJDL, W. 2007. Personalized Query Expansion for the Web. 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007). Amsterdam, The Netherlands: ACM, pp. 7-14.

[15] LEVELING, J. & JONES, G. J. F. 2010. Classifying and Filtering Blind Feedback Terms to Improve Information Retrieval Effectiveness. Adaptivity, Personalization and Fusion of Heterogeneous Information (RIAO 2010). Paris, France: LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, pp. 156-163.

[16] YIN, Z., SHOKOUHI, M. & CRASWELL, N. 2009. Query Expansion Using External Evidence. Lecture Notes In Computer Science. 31st European Conference on Information Retrieval (ECIR 2009). Toulouse, France: Springer, pp. 362-374.

[17] CAO, G., NIE, J.-Y., GAO, J. & ROBERTSON, S. 2008. Selecting Good Expansion Terms for Pseudo-Relevance Feedback. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008). Singapore, Singapore: ACM, pp. 243-250.

[18] OGILVIE, P., VOORHEES, E. & CALLAN, J. 2009. On the Number of Terms Used in Automatic Query Expansion. Information Retrieval, 12**,** 666-679.

[19] CAO, G., NIE, J.-Y., GAO, J. & ROBERTSON, S. 2008. Selecting Good Expansion Terms for Pseudo-Relevance Feedback. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008). Singapore, Singapore: ACM, pp. 243-250.

[20] DE LUCA, E. W. & NÜRNBERGER, A. 2006. Adaptive Support for Cross-Language Text Retrieval. Lecture Notes in Computer Science. 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2006). Dublin, Ireland: Springer, pp. 425-429

[21] CALLAN, J. P., CROFT, W. B. & BROGLIO, J. 1995. TREC and TIPSTER Experiments with INQUERY. Information Processing & Management, 31**,** 327-343.

[22] XU, Y., ZHANG, B., CHEN, Z. & WANG, K. 2007. Privacy-Enhancing Personalized Web Search. 16th International Conference on World Wide Web (WWW 2007). Banff, Alberta, Canada: ACM, pp. 591-600.

[23] NGUYEN, D., OVERWIJK, A., HAUFF, C., TRIESCHNIGG, D., HIEMSTRA, D. & DE JONG, F. 2008. WikiTranslate: Query Translation for Cross-Lingual Information Retrieval Using Only Wikipedia. Lecture Notes in Computer Science. Cross-Language Evaluation Forum (CLEF 2008). Aarhus, Denmark: Springer, pp. 58-65.

[24] RUTHVEN, I. 2003. Re-examining the Potential Effectiveness of Interactive Query Expansion. 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003). Toronto, Canada: ACM, pp. 213-220.

[25] HARMAN, D. 1992b. Relevance Feedback Revisited. 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1992). Copenhagen, Denmark: ACM, pp. 1-10.

[26] SALTON, G. & BUCKLEY, C. 1990. Improving Retrieval Performance by Relevance Feedback. Journal of the American Society for Information Science, 41**,** 288-297.

[27] BAST, H., MAJUMDAR, D. & WEBER, I. 2007. Efficient Interactive Query Expansion with Complete Search. 16th ACM Conference on Information and Knowledge Management (CIKM 2007). Lisbon, Portugal: ACM, pp. 857-860.

[28] SPERETTA, M. & GAUCH, S. 2005. Personalized Search based on User Search Histories. IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005). Compiegne University of Technology, France, pp. 622-628.

[29] VALLET, D., CANTADOR, I. & JOSE, J. 2010. Personalizing Web Search with Folksonomy-Based User and Document Profiles. In: GURRIN, C., HE, Y., KAZAI, G., KRUSCHWITZ, U., LITTLE, S., ROELLEKE, T., RÜGER, S. & VAN RIJSBERGEN, K. (eds.) 32nd European

Conference on Information Retrieval (ECIR 2010). Milton Keynes, UK: Springer Berlin / Heidelberg, pp. 420-431.
[30] STAMOU, S. & NTOULAS, A. 2009. Search Personalization Through Query and Page Topical Analysis. User Modeling and User-Adapted Interaction, 19**,** 5-33.
[31] NOLL, M. & MEINEL, C. 2007. Web Search Personalization Via Social Bookmarking and Tagging. In: ABERER, K., CHOI, K.-S., NOY, N., ALLEMANG, D., LEE, K.-I., NIXON, L., GOLBECK, J., MIKA, P., MAYNARD, D., MIZOGUCHI, R., SCHREIBER, G. & CUDRÉ-MAUROUX, P. (eds.) 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC 2007). South Korea: Springer Berlin / Heidelberg, pp. 367-380.
[32] AGICHTEIN, E., BRILL, E. & DUMAIS, S. 2006a. Improving Web Search Ranking by Incorporating User Behavior Information. 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2006). Seattle, Washington, USA: ACM, pp. 19-26.
[33] HAVELIWALA, T. H. 2002. Topic-sensitive PageRank. 11th International Conference on World Wide Web (WWW 2002). Honolulu, Hawaii, USA: ACM, pp. 517-526.
[34] Liu F., Yu C., and Meng W. Personalized web search by mapping user queries to categories. In Proc. Int. Conf. on Information and Knowledge Management, 2002, pp. 558–565
[35] Pretschner A. and Gauch S. Ontology based personalized search. In Proc. 11th IEEE Int. Conf. on Tools with Artificial Intelli- gence, 1999, pp. 391–398.
[36] Qiu F. and Cho J. Automatic identification of user interest for personalized search. InProc. 15th Int. World Wide Web Conference, 2006, pp. 727–736.
[37] Shen X., Tan B., and Zhai C. Implicit user modeling for personalized search. In Proc.Int. Conf. on Information and Knowledge Management, 2005, pp. 824–831.
[38] Sugiyama K., Hatano K., and Yoshikawa M. Adaptive web search based on user profile constructed without any effort from users. In Proc. 12th Int. World Wide Web Conference, 2004, pp. 675–684.
[39] Sun J.-T., Zeng H.-J., Liu H., Lu Y., and Chen Z. CubeSVD: a novel approach to personalized web search. In Proc. 14th Int. World Wide Web Conference, 2005, pp. 382–390
[40] http://en.wikipedia.org/wiki/User_modeling
[41] Chen, M., Han, J., and Yu, P. (1996) Data Mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering. 8, 866-883.
[42] Yu, P. (1999) Data Mining and Personalization Technologies. In: Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA99). 1-22.
[43] Wang, T., & Ren, Y. (2009). Research on personalized recommendation based on web usage mining using collaborative filtering technique. WSEAS Transactions on Information Science and Applications , 6 (1), 62-72.
[44] Amatriain, X., Jaimes, A., Oliver, N., & Pujol, J. M. (2011). Data Mining Methods for Recommender Systems. In Recommender Systems Handbook (pp. 39-71). Springer US.
[45] Chen, M., Han, J., and Yu, P. (1996) Data Mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering. 8, 866-883.
[46] Yu, P. (1999) Data Mining and Personalization Technologies. In: Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA99). 1-22.
[47] Chen, M., Han, J., and Yu, P. (1996) Data Mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering. 8, 866-883.
[48] Yu, P. (1999) Data Mining and Personalization Technologies. In: Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA99). 1-22.
[49] Cho, Y. H., Kim, J. K., & Kim, S. H. (2002). A Personalized Recommender System Based on Web Usage Mining and Decision Tree Induction. Expert Systems with Applications , 23 (3), 329-342.
[50] Cooley, R., Mobasher, B., & Srivastava, J. (2000). Automatic personalization based on Web usage mining. Communications of the ACM 43.8., (pp. 142-151).
[51] Van Meteren, R., & Van Someren, M. (2000). Using content-based filtering for recommendation. Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000.
[52] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Analysis of Recommendation Algorithms for E-Commerce. ACM Conference on Electronic Commerce (pp. 158 - 167). New York: ACM.
[53] Labroche, N., Lesot, M.J. and Yaffi, L., "A New Web Usage Mining and Visualization Tool", 19th IEEE International Conference on Tools with Artificial Intelligence, Vol. 1, Pp. 321-328, 2007.
[54] Xidong Wang, Yiming Ouyang, Xuegang Hu and Yan Zhang, "Discovery of user frequent access patterns on Web usage mining", The 8th International Conference on Computer Supported Cooperative Work in Design, Vol. 1, Pp. 765-769, 2004.
[55] Girardi, R. and Marinho, L. B. (2007), 'A domain model of web recommender systems based on usage mining and collaborative filtering',Requirements Eng12(1), 23–40
[56] Eirinaki, M. and Vazirgiannis, M. (2003), 'Web mining for web personalization', ACM Transactions on Internet Technology 3(1), 1–27
[57] Agichtein, E., Brill, E., Dumais, S. T. and Ragno, R. (2006), 'Learning user interaction models for predicting web search result preferences', Personal Information Management pp. 251-259.
[58] http://en.wikipedia.org/ontology_%28informationscience%29.
[59] http://www.unicist.org/what-is-an-ontology.pdf
[60] http://en.wikipedia.org/wiki/Ontology
[61] Bhaganagare Ravishankar, Dharmadhikari Dipa, "Web Personalization Using Ontology: A Survey",IOSR Journal of Computer Engineering (IOSRJCE) ISSN : 2278-0661 Volume 1, Issue 3 (May-June 2012), PP 37-45 www.iosrjournals.org.
[62] Gruber, T., "Toward Principles for the Design of Ontologies Used for Knowledge sharing". 1993.
[63] K. Curran, C. Murphy, and S. Annesley, " Web intelligence in information retrieval". In Proc. of WI' 03, pages 409 – 412, 2003.
[64] S. Staab and S. R., editors. "Handbook on Ontologies"
[65] E. Rich, "Users are individuals: individualizing user models". Int. J. Man-Machine Studies (1983) 18, 199-214
[66] Pablo CASTELLS, Miriam FERNÁNDEZ, and David VALLET, "An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval"
[67] P. Kearney, S. S. Anand, "Employing a Domain Ontology to gain insights into user behaviour", in Proc. of the 3rd Workshop on Intelligent Techniques for Web Personalization (ITWP 2005), Endiburgh, Scotland, August 2005.
[68] S. Acharyya, J. Ghosh, "Context-Sensitive Modeling of Web Surfing Behaviour Using Concept Trees", in Proc. of the 5th WEBKDD Workshop, Washington, August 2003.
[69] Mehrdad Dianati, Insop Song, and Mark Treiber. An introduction to genetic algorithms and evolution strategies. Technical report, University of Waterloo,Ontario, N2L 3G1, Canada, July 2002.
[70] Gareth Jones. Genetic and evolutionary algorithms. In Paul von Rague, editor,Encyclopedia of Computational Chemistry. John Wiley and Sons, 1998.

[71] Melanie Mitchell. An Introduction to Genetic Algorithms. MIT Press, Cambridge,MA, 1996.
[72] Ulrich Bodenhofer. Genetic Algorithms: Theory and Applications. Lecture notes Fuzzy Logic Laboratorium Linz-Hagenberg, Winter 2003/2004.
[73] Diane Kelly,School of Information & Library Science, IPAM | 04 October 2007.
[74] BAEZA-YATES, R. & RIBEIRO-NETO, B. 2011. Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition), Addison-Wesley.
[75] BROOKE, J. 1996. SUS-A quick and dirty usability scale. In: JORDAN, P. W., THOMAS, B., WEERDMEESTER, B. A. & MCCLELLAND, A. L. (eds.) Usability Evaluation in Industry, pp. 189-194.
[76] SMYTH, B. & BALFE, E. 2006. Anonymous Personalization in Collaborative Web Search. Information Retrieval, 9**,** 165-190.
[77] STAMOU, S. & NTOULAS, A. 2009. Search Personalization Through Query and Page Topical Analysis. User Modeling and User-Adapted Interaction, 19**,** 5-33.
[78] STEFANI, A. & STRAPPARAVA, C. 1999. Exploiting NLP Techniques to Build User Model for Web Sites: the Use of WordNet in SiteIF Project. 2nd Workshop on Adaptive Systems and User Modeling on the World Wide Web. Toronto, Canada.
[79] SPERETTA, M. & GAUCH, S. 2005. Personalized Search based on User Search Histories. IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005). Compiegne University of Technology, France, pp. 622-628.
[80] AGICHTEIN, E., BRILL, E. & DUMAIS, S. 2006a. Improving Web Search Ranking by Incorporating User Behavior Information. 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2006). Seattle, Washington, USA: ACM, pp. 19-26.
[81] STEICHEN, B., LAWLESS, S., O'CONNOR, A. & WADE, V. 2009. Dynamic Hypertext Generation for Reusing Open Corpus Content. 20th ACM Conference on Hypertext and Hypermedia (Hypertext 2009). Torino, Italy: ACM, pp. 119-128.
[82] CALLAN, J. P., CROFT, W. B. & BROGLIO, J. 1995. TREC and TIPSTER Experiments with INQUERY. Information Processing & Management, 31**,** 327-343.
[83] MICARELLI, A., GASPARETTI, F., SCIARRONE, F. & GAUCH, S. 2007. Personalized Search on the World Wide Web. In: BRUSILOVSKY, P., KOBSA, A. & NEJDL, W. (eds.) The Adaptive Web. 1 ed.: Springer, pp. 195-230.
[84] PITKOW, J., SCHUTZE, H., CASS, T., COOLEY, R., TURNBULL, D., EDMONDS, A., ADAR, E. & BREUEL, T. 2002. Personalized Search. Communications of the ACM, 45**,** 50-55.
[85] Cross Language Evaluation Forum: http://www.clef-campaign.org/
[86] Text REtrieval Conference: http://trec.nist.gov/
[87] NII Test Collection for IR Systems: http://research.nii.ac.jp/ntcir
[88] TEEVAN, J., DUMAIS, S. T. & LIEBLING, D. J. 2008. To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008). Singapore, Singapore: ACM, pp. 163-170.
[89] ARGUELLO, J., DIAZ, F., CALLAN, J. & CARTERETTE, B. 2011. A Methodology for Evaluating Aggregated Search Results. 33rd European Conference on Information Retrieval (ECIR 2011). Dublin, Ireland, pp. 141-152.
[90] DIAZ, F. & ARGUELLO, J. 2009. Adaptation of Offline Vertical Selection Predictions in the Presence of User Feedback. 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2009). Boston, MA, USA: ACM, pp. 323-330.
[91] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, “Itembased collaborative filtering recommendation algorithms,” in Proceedings of the 10th International Conference on World Wide Web (WWW ’01), pp. 285–295, May 2001.
[92] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, An algorithmic framework for performing collaborative filtering,” in Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR ’99), pp. 230–237, 1999.
[93] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “Grouplens: an open architecture for collaborative filtering of netnews,” in Proceedings of the ACM Conference on Computer Supported Cooperative Work, pp. 175–186, New York, NY, USA, 1994.
[94] G. Salton andM.McGill, Introduction to Modern Information Retrieval, McGraw-Hill, New York, NY, USA, 1983.
[95] G. Karypis, “Evaluation of item-based top-N recommendation algorithms,” in Proceedings of the International Conference on Information and Knowledge Management (CIKM ’01), pp. 247–254, Atlanta, Ga, USA, ACM Transactions on Information Systems, vol. 22, no. 1, pp. 143–177, 2004.
[97] J. Breese, D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” in Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI ’98), 1998.
[98] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, “Analysis of recommendation algorithms November 2001.
[96] M.Deshpande and G.Karypis, “Item-based top-N recommendation algorithms,” for E-commerce,” in Proceedings of the ACM E-Commerce, pp. 158–167, Minneapolis, Minn, USA, 2000.
[99] J.Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Francisco, Calif, USA, 1988.
[100] J.Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, Calif, USA, 2001.
[101] X. Su, M. Kubat, M. A. Tapia, and C. Hu, “Query size estimation using clustering techniques,” in Proceedings of the 17th International Conference on Tools with Artificial Intelligence (ICTAI ’05), pp. 185–189, Hong Kong, November 2005.
[102] J. B. MacQueen, “Some methods for classification and analysis of multivariate observations,” in Proceedings of the 5th Symposium on Math, Statistics, and Probability, pp. 281–297, Berkeley, Calif, USA, 1967.
[103] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD ’96), 1996.
[104] G. Shani, D. Heckerman, and R. I. Brafman, “AnMDP-based recommender system,” Journal ofMachine Learning Research, vol. 6, pp. 1265–1295, 2005.
[105] T. Hofmann, “Unsupervised learning by probabilistic latent semantic analysis,” Machine Learning, vol. 42, no. 1-2, pp. 177–196, 2001.
[106] Wouter IJntema, Frank Goossen ,Flavius Frasincar, Frederik Hogenboom ,” Ontology-Based News Recommendation” EDBT 2010, March 22–26, 2010, Lausanne, Switzerland.
[107] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, “Analysis of recommendation algorithms for E-commerce,” in

 ISSN: 2321-8169
1385 – 1395

Proceedings of the ACM E-Commerce, pp. 158–167, Minneapolis, Minn, USA, 2000.

[108] D. Nikovski and V. Kulev, "Induction of compact decision trees for personalized recommendation," in Proceedings of the ACM Symposium on Applied Computing, vol. 1, pp. 575–581, Dijon, France, April 2006.

[109] C. C. Aggarwal, J. L. Wolf, K. Wu, and P. S. Yu, "Horting hatches an egg: a new graph-theoretic approach to collaborative filtering," in Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '99), pp. 201–212, 1999.

[110] B. Marlin and R. S. Zemel, "The multiple multiplicative factor model for collaborative filtering," in Proceedings of the 21st International Conference on Machine Learning (ICML '04), pp. 576–583, Banff, Canada, July 2004.

[111] J. Canny, "Collaborative filtering with privacy via factor analysis," in Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 238–245, Tampere, Finland, August 2002.

[112] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," Journal of the Royal Statistical Society. Series B, vol. 21, no. 3, pp. 611–622, 1999.

[113] P. Melville, R. J. Mooney, and R. Nagarajan, "Contentboosted collaborative filtering for improved recommendations," in Proceedings of the 18th National Conference on Artificial Intelligence (AAAI '02), pp. 187–192, Edmonton, Canada, 2002.

[114] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, and H.-P. Kriegel, "Probabilistic memory-based collaborative filtering," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 1, pp. 56–69, 2004.

[115] D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles, "Collaborative filtering by personality diagnosis: a hybrid memory- and model-based approach," in Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI '00), pp. 473–480, 2000.

[116] H. Al Tair, M. J. Zemerly, M. Al-Qutayri , M. Leida "Architecture for Context-Aware Pro-Active Recommender System" International Journal of Multimedia and Image Processing (IJMIP), Volume 2, Issues 1/2, March/June 2012.

[117] Lorraine McGinty ,Barry Smyth, "Improving the Performance of Recommender
Systems That Use Critiquing" Lecture Notes.

[118] Park, Y. J. and Chang, K. N., "Individual and group behavior-based customer profile model for personalized product recommendation," *Expert Systems with Applications*, vol. 36, pp. 1932-1939, 2009.

[119] www.recsyswiki.com/wiki/**MovieLens**

[120] Boi Faltings, Florent Garcin, "PEN recsys: a PersonalizedNews Recommender Systems Framework" News Recommender Systems Workshop (NRS), October 13, 2013.

[121] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2-3, pp. 131–163, 1997.

.