ISSN 0976 – 1411

Available online at www.internationaleJournals.com

International eJournals

International eJournal of Mathematics and Engineering 126 (2011) 1145 - 1154

INTERNATIONAL eJOURNAL OF MATHEMATICS AND ENGINEERING

www.InternationaleJournals.com

A STUDY ON RECOMMENDER SYSTEMS

L Smitha, G Narayanamma Institute of Technology, Hyderabad, India
Hameed Ahmed, OUCE, Osmania University, Hyderabad, India <sampathsreekanth@yahoo.com>
S Sameen Fatima, OUCE, Osmania University, Hyderabad, India

Abstract: Personalization is the process of customizing a Web site to users' specific requirements using the knowledge acquired from the analysis of users' navigational behavior in correlation with other information collected in the Web context as well as other related individual intellectual, mental, emotional and social context elements .Web Personalization is viewed as an application of data mining and machine learning techniques to build models of user behavior that can be applied to the task of predicting user needs and adapting future interactions with the ultimate goal of improved user satisfaction. In this paper different Personalization paradigms are discussed like Collaborative filtering, in which users are invited to rate the objects or reveal their preferences and then return information that is predicted to be of interest to them, but Content-based filtering is based on the individual user's preferences. Hybrid approach integrates some techniques

1. INTRODUCTION

During the past few years the World Wide Web has become the biggest and most popular way of communication and information dissemination. An abundant amount of information is transferred over electronic media. It serves as a platform for exchanging various kinds of information, ranging from research papers, and educational content, to multimedia content, software and blogs. Every day, the web grows by roughly a million electronic pages, adding to the hundreds of millions pages already on-line. Because of its rapid and chaotic growth, the resulting network of information lacks of organization and structure. Computer Users often feel confused and get lost in that information overloads that continue to expand. Users risk becoming overwhelmed by the flow of information, and the users lack adequate tools to help them manage the situation. Information filtering is one of the methods that are rapidly evolving to manage large information flows. The effort of Information filtering is to expose users to only information that is relevant to them. Many Information filtering systems have been developed in recent years for various application domains. Information filtering systems can facilitate users by eliminating the irrelevant information and by bringing the related information to the user's attention. Filters are intermediaries between the sources of information and their end-users.

The system is based on a user modeling component [21], intended for building and maintaining long term models of individual Internet users. Presently the system acts as an edge for the Web search engines. The experimental results we have obtained are encouraging and support the choice of adaptive Information Filtering. Its main goal is the management of the excess information. In order to achieve this, user’s profile is compared to some reference characteristics. These features may originate from the information item (the content-based approach) or the user's social environment (the collaborative filtering approach).Whereas in information transmission electronic filters are used against syntax troublesome noise on the bit-level, the methods employed in information filtering act on the semantic level. The range of machine methods in use builds on the same principles as those for information extraction [1]. A prominent application can be found in the field of email spam filters. Thus, it is not only the information explosion that necessitates some form of filters, but also unintentionally or maliciously introduced pseudo-information. The different systems use various methods, and techniques from diverse research areas like: Information Retrieval, Artificial Intelligence, or Behavioral Science. There are many systems of widely varying philosophies, but all shares the goal of automatically directing the most valuable information to users in accordance with their User Model, and of helping them use their limited reading time most optimally. When a user interacts with the system for the first time, the user model needs to be made from scratch. In order to quickly build a reliable model an interview is proposed to the user, expressing an interest score for each of the domain categories. The user sets a query to the system that will do the job of posting it to the external Web search engine, obtaining documents that are filtered and returned to the user. In the filtering process the systems works using two different levels of refinement, a first, crude one, and a more detailed step that takes place only if the first stage succeeds. During the usual usage the system offers a series of panels, being the first the filtering panel [19]. Here shown [19] the list of documents retrieved by the search engine given the user query.

For our usage the system automatically sorts the document lists so to help the user locating the best documents. The user browses the needed documents by double-clicking on them, and then he can express a simple feedback [15] among three different values: very good, good or bad, in order to ease the burden on the user as recommended. In this way the system can modify the user model accordingly to user's preferences [3,4]. Furthermore, a system objects browser has been provided in order to allow the user to inspect all the system's data structures with an effective graphical interface to shorten the semantic gap between the user and the system.

2. CLASSIFICATION OF WEB MINING TECHNIQUES

Web Content Mining: The technique, involving mining web data contents, describe the automatic search of information resource available online

Web-Structure Mining: It tries to discover the link structure of the hyperlinks at the inter-document level to generate structural summary about the Website and Web page

Web-Usage Mining: his usage mining is the application of data mining techniques to find out usage patterns from

the secondary data derived from the interactions of the users while surfing on the web, in order to understand and better serve the needs of Web based applications.

3. RECOMMENDATION SYSTEMS

Recommender systems are active information filtering systems that attempt to present to the user information items (movies, music, books, news, web pages) the user is interested in. These systems add information items to the information flowing towards the user, Typically, a recommender system compares the user's profile to some reference characteristics, and seeks to predict the rating that a user would give to an item they had not yet considered [20]. Recommender systems use collaborative filtering approaches or a combination of the collaborative filtering and content-based filtering approaches, although content-based recommender systems do exist [7]. Web-based Recommender Systems (RS) are recently applied to provide different type of customized information for their users. The Recommender Systems are applied in many areas such as: web-browsing, information filtering, net-news or movie recommender and e-Commerce. The central element of all recommender systems is the user model that contains knowledge about the individual preferences which determine his or her behavior in a complex environment of web-based systems.

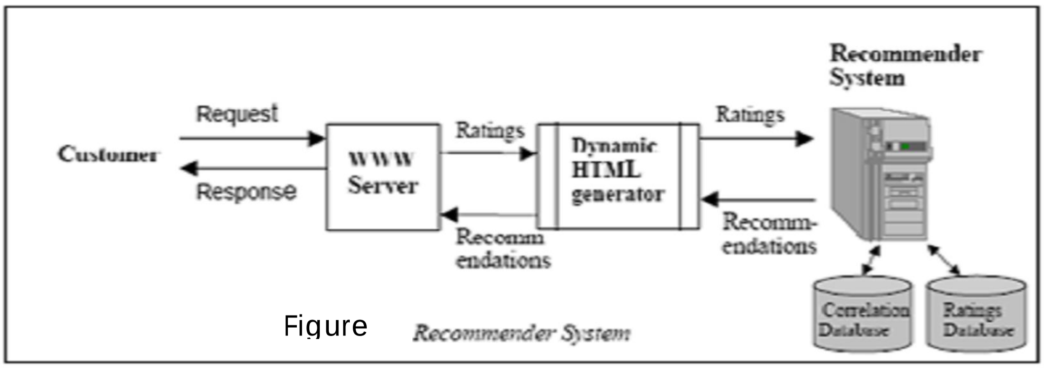

Figure *Recommender System*

Recommender systems must: (i) choose which (of the items) should be shown to the user, (ii) decide when and how the recommendations must be shown. Then, metrics emerges naturally from the framework[41]. User modeling as well as RS are characterized by cross-fertilization of various research fields such as: Information Retrieval, Artificial Intelligence, Knowledge Representation, Discovery and Data/Text Mining, Computational Learning and Intelligent and Adaptive Agents. The alternating information environment that is combined of various users, their needs and contexts of use as well as different system platforms necessitates application of recommender systems.

The growth of the e-Commerce in the global economy increases the value of Recommender systems. RS systems are developed by different domains such as personal agents and adaptive hypermedia. The personalized hypermedia application is defined as a hypermedia system that adapts: the content, structure, and/or presentation of the web objects to each individual user's model. RS's are applied in many different areas from web browsing for purchase recommendation. Montaner et. al in their work present comprehensive taxonomy of the recommender. agents. In this taxonomy the following two dimensions are considered: profile generation and maintenance, and profile exploitation. The dimension of profile generation and maintenance considers the following elements: user profile representation, initial profile generation, profile learning technique and relevance feedback [22].

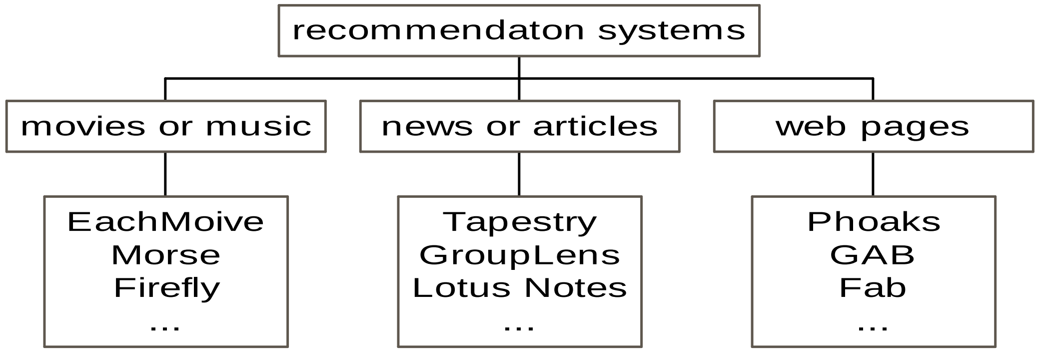

Figure 2. The recommendation systems' categories

3.1 CLASSIFICATION OF RECOMMENDATION SYSTEM

Many research groups have built various types of systems that recommend pages to web users. Here some are summarized and discuss how they differ from each other. The objective of collecting user information is to create a profile that describes user characteristics.

The more common techniques are explicit profiling, implicit profiling, and use of legacy data
3.1.1Explicit profiling:
Each user is asked to fill in a form when visiting the web site. This method has the advantage of charter users specify directly their interests.
3.1.2.Implicit profiling:
The user's behavior is studied automatically by the system. This method is generally clear to the user. Often, user registration is saved as cookies, which is kept at the browser and updated for every visit.
A user profile can be divided into two varieties static and dynamic.
A user profile can be either static, when the information it contains is never or rarely altered, or dynamic when the user profile's data change frequently. Behavior information is generally stored in a log file. Legacy data: The Legacy data provides a rich source of profile information for known users.

3.2. Web Personalization paradigm Comparison

3.2.1Content-based filtering:
Systems which are implementing these kinds of techniques are solely based on individual users' preferences. The system tracks each user's behavior and recommends items that are similar to items the user liked in the past. It is based on description analysis of the items rated by the user and correlations between the content of these items and user's preferences. It is an alternative paradigm that has been used mainly in the context of recommending items such as books, Web pages, news, etc. for which informative content descriptors exist [37, 38,39].

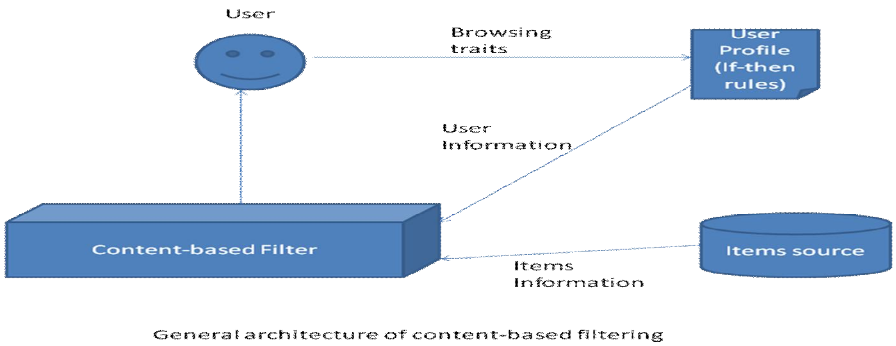

Figure 3.

This technique is primarily characterized by two weaknesses, content Limitations and over-Specialization. There are content limitations like IR methods that can only be applied to a few kinds of content, such as text and image, and the extent aspects can only capture certain aspects of the content. On the other hand content-based recommendation systems provide recommendations merely based on user profiles, therefore, users have no chance of exploring new items that are not similar to those items included in their profiles and thus leading to over-specialization.

Some drawbacks that have been identified in time are [37, 36, and 38]:

(a) Search-based models built with keywords, categories, fail to provide recommendations with interesting, targeted titles.
(b) They also scale poorly for customers with frequent purchases and ratings.
(c) User input may be subjective and prone to bias
(d) Explicit user ratings may not be available
(e) Profiles may be static and can become outdated quickly.
(f) Semantic relationships between objects are missed.

3.2.2. Rule-Based Techniques:

The users are asked to answer a set of questions. These questions are consequent from a decision tree, so as the user proceeds to answer them. What he finally receives is a result (e.g. list of products) modified to his needs. Some of the rule-based filtering drawbacks are: User input may be subjective and prone to bias, explicit (and non-binary) user ratings may not be available, profiles may be static and can become outdated quickly, and for large systems it becomes burdensome to manage. Related interesting systems include Dell, Apple Computer, Amazon.com, and Broadvision [33, 34, 35, and 36].

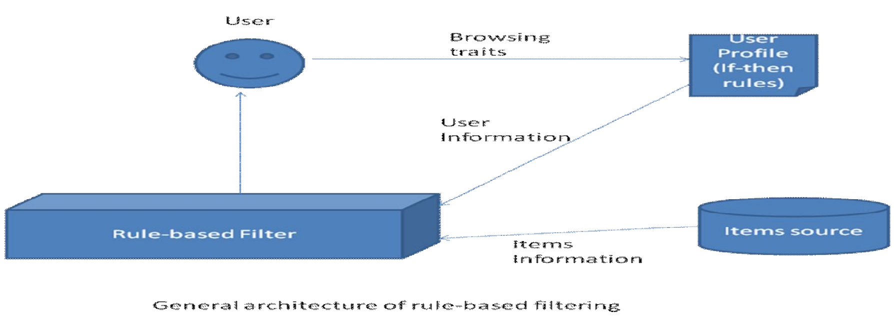

Figure 4.

Rule-based techniques exploit a set of rules specified in the system in order to drive personalization. Cross-selling is an e-business example of the rule-based technique: a rule could be specified to offer product X to a customer who has just bought product Y. For example, a customer bought a product might be interested in similar products.

3.2.3. Collaborative filtering:

Systems invite users to rate the objects or divulge their preferences and interests and then return information that is predicted to be of interest to them. This is based on the assumption that users with similar behavior (e.g. users that are rate similar objects) have analogous interests. Moreover, the goals in a collaborative filtering system are basically focused upon the reduction of computation time, the increase of the extent in which predictions can be computed in parallel, and the raise of prediction accuracy. Collaborative filtering can further refine the process of giving each individual personal recommendation compared to rule-based filtering. It overcomes the drawbacks of the content-based filtering because it typically does not use the actual content of the items for recommendation, which is usually based on assumptions. With this algorithm the similarity between the users is evaluated based on their ratings of products, and the recommendation is generated considering the items visited by nearest neighbors of the user. In its original form, the nearest-neighbor algorithm uses a two-dimensional user-item matrix to represent the user profiles. This original form suffers from three problems, scalability, sparsity, and synonymy [37, 40].

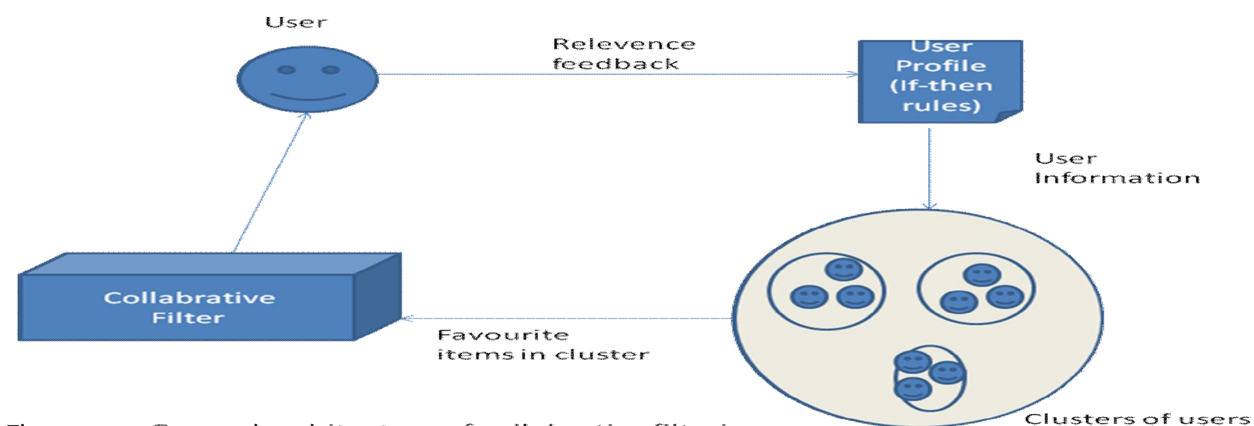

Figure General architecture of collabrative filtering

Some more drawbacks of collaborative filtering are:
(a) Collaborative-filtering techniques are often based in matching in real-time the current user's profile against similar records obtained by the systems over time from other users. It becomes hard to scale collaborative filtering techniques to a large number of items, while maintaining reasonable prediction performance and accuracy. Part of this is due to the increasing sparsity in the data.One Possible solution is as follows: In the context of Web personalization ,it involves clustering user transactions identified in the preprocessing stage.
(b) Traditional collaborative filtering does little or no offline computation, and its online computation scales with the number of customers and catalog items. The algorithm is impractical on large data sets, unless it uses dimensionality reduction, sampling, or partitioning – all of which reduce recommendation quality.

(c) User input may be subjective and prone to bias;
(d) Explicit user ratings may not be available;
(e) Profiles may be static and can become outdated quickly;
(f) They are not able to recommend new items that have not already been rated by other users.

3.2.4. Item –based collabrativeFiltering:
Based on item relations, not on user relations, as in classic Collaborative Filtering. Bardul M. Sarwar projected a different approach in the area of filtering algorithms, that was suggested recently[42] [28] [29]. In the Item-based Collaborative Filtering algorithm, we look into the set of items, that the active user, has rated, compute how similar they are to the target item and then select the k most similar items {i1, i2, ..., ik}, based on their corresponding similarities {si1, si2, ..., sik}. The predictions can then be computed by taking a weighted average of the active user□s ratings on these similar items. The first step in this new approach is the Representation. Its purpose is the same as with the classic Collaborative Filtering algorithm: represent the data in an organized manner. The Item Similarity Computation should be calculated. The basic idea in that step is to first isolate the users who have rated two items ij and ik and then apply a similarity computation technique to determine their similarity. Various ways to compute that similarity have been proposed.

3.2.5. Hybrid Filtering
Filtering defines two distinct filtering components. The first component implements plain Collaborative Filtering, while the second component implements Content based Filtering. The final rating prediction is calculated as a weighted sum of those components, where the applied weights are decided by how close is the prediction of each component to the actual rating. Content-based, Rule-based, and collaborative filtering may also be used in combination, for deducing more accurate results [42].Such methods are used in order to achieve The benefits form two or more methods and will also minimizes their disadvantages.

4. CHALLENGING ISSUES IN RECOMMENDATION SYSTEM
The several current challenges of the recommender systems are considered in this section. The first set of challenges concerns issues of bringing people together into communities of interest. A major concern here is privacy.

- The second challenge is to create recommendation algorithms that combine multiple types of information, probably acquired from different sources at different times. Establishing the user tasks to be supported by a system, and selecting a data set on which performance enables empirical experimentation – scientifically repeatable evaluations of recommender system utility.
- A majority of the published empirical evaluations of recommender systems to date has focused on the evaluation of recommender system□s accuracy. We assume that if a user could examine all items available, they could place those items in a ordering of preference. Accuracy metric empirically measures how close a recommender system□s predicted ranking of items for a user differs from the user□s true ranking of preference. Accuracy measures may also measure how well a system can predict an exact rating value for a specific item. Researchers who want to quantitatively compare the accuracy of different recommender systems must first select one or more metrics.

In selecting a metric, researchers face a range of questions.
Will a given metric measure the effectiveness of a system with respect to the user tasks for which it was designed? Are results with the chosen metric comparable to other published research worked in the field?
Are the assumptions that a metric is based on true? Will a metric be sensitive enough to detect real differences that exist?
How large a difference does there have to be in the value of a metric for a statistically significant difference to exist?

- Recommender systems are a powerful new technology for extracting additional value for a business from its customer databases. The systems help customers find products they want to buy from a business. Recommender systems benefit customers by enabling them to find products they like. Conversely, they help the business by generating more sales. Recommender systems are rapidly becoming a crucial tool in E-commerce on the Web. Recommender systems are being stressed by the huge volume of customer data in existing corporate databases, and will be stressed even more by the increasing volume of customer data available on the Web. New technologies are needed that can dramatically improve the scalability of recommender systems. Web recommendation system is seen as a fully automated process, powered by operational Knowledge. In addition to the various improvements to the Web recommendation system process, there are a number of other issues, which need to be addressed in order to develop effective Web personalization systems. The treatment of time in the user models can be distinguished as being particularly difficult. The main source of difficulty is that the manner in which the behavior of users changes over time varies significantly with the application and possibly the type of the user. Therefore, any solution to this problem should be sufficiently parametric to cater for the requirements of different applications.

REFERENCES

[1] Han, J., Kamber, M. (2001). Data Mining: Concepts and Techniques. Morgan-Kaufmann Academic Press, San Francisco.
[2] R. Agrawal; T. Imielinski; A. Swami: Mining Association Rules Between Sets of Items in Large Databases", SIGMOD Conference 1993: 207-216
[3] Peetr Hajek, Tomas Feglar, Jan Rauch, David Coufal. The GUHA method, data preprocessing bvand mining. Database Support for Data Mining Applications, ISBN 978-3-540-22479-2, Springer, 2004
[4] Hajek P., Havel I., Chytil M.: The GUHA method of automatic hypotheses determination, Computing 1(1966) 293-308.
[5] Piatetsky-Shapiro, G. (1991), Discovery, analysis, and presentation of strong rules, in G. Piatetsky-Shapiro & W. J. Frawley, eds, “Knowledge Discovery in Databases”, AAAI/MIT Press, Cambridge, MA.
[6] “Web Site Personalisation”. IBM High-Volume Web site team, January 2000,http://www-4.ibm.com/software/developer/library/personalization/index.html.
[7] Bilgic and R. J. Mooney. Explaining recommendations: Satisfaction vs. promotion. In Beyond Personalization Workshop, IUI, 2005.

[8] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In ACM conference on Computer supported cooperative work, 2000.
[9] Mcsherry. Explanation in recommender systems. Artificial Intelligence Review, 24(2):179 – 197, 2005.
[10] Chi, P. Pirolli, K. Chen, and J. Pitkow. Using information scent to model user information needs and actions on the web. In ACM CHI 2001 Conference on Human Factors in Computing Systems, pages 490-497, Seattle WA, 2001.
[11] Billsus and M. Pazzani. A hybrid user model for news story classification. In Proceedings of the Seventh International Conference on User Modeling (UM '99), Banff, Canada, 1999.
[12] Jennings and H. Higuchi. A user model neural network for a personal news service. User Modeling and User-Adapted Interaction, 3(1):1-25, 1993.
[13] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In Proc. of the 20th Int'l Conference on Very Large Databases (VLDB'94), Santiago, Chile, Sep 1994.
[14] John and P. Langley. Estimating continuous distributions in bayesian classifiers. In Eleventh Annual Conference on Uncertainty in Artificial Intelligence, pages 338-345, San Francisco, 1995. Morgan Kaufmann Publishers. .
[15] McCarthy, J. Reilly, L. McGinty, and B. Smyth. Thinking positively - explanatory feedback for conversational recommender systems. In Proceedings of the European Conference on Case-Based Reasoning (ECCBR-04) Explanation Workshop,, 2004.
[16] R. Rafter and B. Smyth. Conversational collaborative recommendation-an experimental analysis. Artif. Intel. Rev, 24(3-4):301–318, 2005.
[17] Badrul M. Sarwar et al., on “Recommender Systems for Large-scale E-Commerce: Scalable Neighborhood Formation Using Clustering” Department of Computer Science and Engineering University of Minnesota, Minneapolis, MN 55455, USA
[18] Resnick, P., N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In Proceedings of the 1994 Computer Supported Cooperative Work Conference, 1994.
[19] Shardanand, U. and P. Maes. Social information filtering: Algorithms for automating „word of mouth. In Proc. of the Conf. on Human Factors in Computing Systems, 1995.
[20] Hill, W., L. Stead, M. Rosenstein, and G. Furnas. Recommending and evaluating choices in a virtual community of use. In Proceedings of CHI 95.
[21] Hanani U, Shapira B and Shoval P. Source: User Modeling and User-Adapted Interaction, Volume 11, Number 3, August 2001 , pp. 203-259(57)
[22] Oard, D., Kim, J. Implicit Feedback for Recommender Systems. In Proceedings of AAAI Workshops on Recommender Systems. July 1998.
[23] Claypool, M., Le, P., Waseda, M., Brown, D. Implicit Interest Indicator. Computer Science Technical Report Series, WPI Computer Science Department, Worcester, Massachussets, July 2000.
[24] Billsus, D., and Pazzani, M. J. (1998). Learning Collaborative Information Filters. In Proceedings of ICML 98. pp. 46-53.
[25] Swearingen, k., sinha, R. 2001. Beyond Algorithms: An HCI Perspective on Recommender Systems. In SIGIR 2001 Workshop on Recommender Systems.
[26] Cosley, d., lam, s. K., Albert, i., konstan, j. A. and riedl, j. 2003. Is Seeing Believing? How Recommender Interfaces Affect Users' Opinions. CHI Letters 5. TURPIN, A., HERSH, W. 2001. Why Batch and User Evaluations Do Not Give the Same Results. In Proceedings of the

24th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, 17-24

[27] NEWMAN, W., "Better or just different? On the benefits of designing interactive systems in terms of critical parameters. In Designing Interactive Systems" (DIS97), ACM Press, New York, 2001 NY. 239-246.

[28] Bardul M. Sarwar, George Karypis, Joseph A. Konstan, and John T. Riedl, "Item-based collaborative filtering recommendation algorithms," in 10th International World Wide Web Conference (WWW10), Hong Kong, 2001

[29] George Karypis, "Evaluation of item-based top-n recommendation algorithms," in CIKM 2001, 2001.

[30] Eirinaki M., & Vazirgiannis M. (2003). Web Mining for Web Personalization, ACM Transactions on Internet Technology, Vol. 3, No. 1, Pages 1-27.

[31] Bardul M. Sarwar, "Sparsity, Scalability, and Distribution in Recommender Systems", Ph.D. thesis, University of Minnesota, 2001.

[32] Emmanouil G. Vozalis, Konstantinos G. Margaritis, "Recommender Systems: An Experimental Comparison of two Filtering Algorithms", 2003, In Proceedings of the Ninth Panhellenic Conference in Informatics.

[33]Wang J., & Lin J. (2002). Are personalization systems really personal? – Effects of conformity in reducing information overload, Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS'03), 0-7695-1874-5/03, © 2002 IEEE.

[34] Mobasher B., Dai H., Luo T., Nakagawa M., & Wiltshire J. (2000). Discovery of aggregate usage profiles for Web personalization, [online], http://maya.cs.depaul.edu/ ~mobasher/papers/webkdd2000.pdf.

[35] Sharbanand U., & Maes P. (1998). Social information filtering: Algorithms for automating "Word of Mouth", [online], http://citeseer.ist.psu.edu/ shardanand95social.html.

[36] Eirinaki M., & Vazirgiannis M. (2003). Web Mining for Web Personalization, ACM Transactions on Internet Technology, Vol. 3, No. 1, Pages 1-27.

[37] Pazzani J. M. (2005). A framework for collaborative, content-based and demographic filtering, [online], http://www.ics.uci.edu/~pazzani/Publications/ AIREVIEW.pdf.

[38] Basilico J., & Hofmann T. (2004). Unifying collaborative and content-based filtering, Proceedings of the 21st International Conference on Machine Learning, Banff, Canada.

[39] Sharbanand U., & Maes P. (1998). Social information filtering: Algorithms for automating "Word of Mouth", [online], http://citeseer.ist.psu.edu/ shardanand95social.html.

[40] Shahabi C., & Chen Y. (2003). Web information personalization:

[41] Félix Hernández del Olmo and Elena Gaudioso[a]. Madrid, Spain.Evaluation of recommender systems: A new approach

[42] Emmanouil Vozalis, Konstantinos G. Margaritis. Analysis of Recommender Systems' Algorithms