COLLABORATIVE WEB RECOMMENDATION SYSTEMS BASED ON AN EFFECTIVE FUZZY ASSOCIATION RULE MINING ALGORITHM (FARM)

A.KUMAR

Research Scholar, Department of CSE,
Sathyabama University, Chennai.

Dr. P. THAMBIDURAI

Principal, P K Institute of Engineering and Technology,
Karaikal.

Abstract

With increasing popularity of the web-based systems that are applied in many different areas, they tend to deliver customized information for their users by means of utilization of recommendation methods. This recommendation system is mainly classified into two groups: Content-based recommendation and collaborative recommendation system. Content based recommendation tries to recommend web sites similar to those web sites the user has liked, whereas collaborative recommendation tries to find some users who share similar tastes with the given user and recommends web sites they like to that user. Based on web usage data in adoptive association rule based web mining the association rules were applied to personalization. The technique utilize apriori algorithm to generate association rules. Even this method has some disadvantages. To overcome those disadvantages, the author proposed a new algorithm for web recommendation system known as an effective Fuzzy Association Rule Mining Algorithm (FARM). This proposed Fuzzy ARM algorithm for association rule mining in web recommendation system results in better quality and performance.

***Keywords*:** Association Rules, Apriori Algorithm, Fuzzy Healthy Association Rule Mining, Collaborative Recommender.

1. Introduction

The extent of the Internet is getting larger and larger in modern years. Therefore it is obligatory that a user need to expend much time to select indispensable information from large amount of web pages created every day. Addressing this problem, several web page recommender systems are constructed which automatically selects and recommends web pages suitable for user's support. The majority of web page recommender systems that was proposed earlier utilized collaborative filtering [Balabanovic *et al*, (1997)], [Jon Herlocker *et al*, (1999)], and [Sarwar *et al*, (2001)]. Collaborative filtering is often used in general product recommender systems, and consists of the following stages. The foremost stage in collaborative filtering is to analyze users purchase histories in order to extract user groups which have similar purchase patterns. Then recommend the products that are commonly preferred in the user's group.

In general the Recommender Systems (RS) [Hill *et al*, (1995)] uses the opinions of members of a community to facilitate individuals identify the information most likely to be interesting to them or pertinent to their needs. This can be achieved by drawing on user preferences and filtering the set of feasible options to a more manageable subset. Every Web Recommendation Systems have its own advantages and limitations. Moreover the assignment of recommender systems is to recommend items that fit a user's taste, in order to help the user in selecting/purchasing items from a devastating set of choices. Such systems have immense importance in applications such as e-commerce, subscription based services, information filtering, web services etc.

There are two fundamental approaches that can be applied when generating recommendations. Content based approaches profile users and items by identifying their characteristic features, such us demographic data for user profiling, and product information/descriptions for item profiling. The profiles are used by algorithms to unite user interests and item descriptions when generating recommendations [Takacs *et al,*]. Web Content Recommendation is an active application area for Information Filtering, Web Mining and Machine Learning research. The another approach is collaborative recommendation which tries to find some users who share similar tastes with the given user and recommends web sites they like to that user. This proposed paper gives a method of developing a collaborative web recommendation system using Fuzzy Association Rule Mining. Fuzzy Association Rule Mining is the problem of discovering frequent itemsets using fuzzy sets in order to handle the quantitative attributes in transactional and relational databases. It also explains some of the baseline algorithms that are used in developing the web recommendation systems. The proposed approach involves usage of fuzzy healthy association rule mining algorithm for Association Rule Mining. This approach helps the user to obtain the web sites which are most relevant to them.

The remainder section of this paper is organized as follows. Section 2 discusses various collaborative web recommendation systems that were earlier proposed in literature. Section 3 explains the proposed work of developing a web recommendation system using an effective Fuzzy Healthy Association Rule Mining. Section 4 illustrates the results for experiments conducted on sample dataset in evaluating the performance of the proposed web recommendation system. Section 5 concludes the paper with fewer discussions.

2. Related work

In general, the collaborative recommendation systems can be grouped into four categories. On the basis of its temporal and spatial characteristics, each system can be either synchronous or asynchronous, and either local or remote. Conversely, the most significant difference between these different collaborative Web recommendation systems is the method used to extort user preferences from personal information. This section of the paper discusses various methods proposed earlier in literature for a collaborative web recommendation system.

[Chen *et al*, (2009)] proposed a Gradual Adaption Model for a Web recommender system. The model is used to track users' center of attention and its transition by analyzing their information access behaviors, and recommend appropriate information. The web pages admittance by users are classified by the concept classes, and grouped into three terms of short, medium and long periods, and two categories of significant and incomparable for each concept class, which are used to describe users' focus of interests, and to institute reprocess probability of each concept class in each term for each user by Full Bayesian Estimation as well. Based on the reuse probability and period, the information that a user is likely to be interested in is recommended. They proposed a new approach by which short and medium periods are determined based on dynamic sampling of user information access behaviors.

[Nima *et al*, (2008)] described a web page recommender system based on Folksonomy mining. They projected a way to assemble a new type of web page recommender system covering all over the Internet, by using Folksonomy and Social Bookmark which are getting very well-liked in these days. A new way to express users' preference to web pages was formulated by mining tag data of Folksonomy. Folksonomy is a new classification technique which may take place of past taxonomy. Social Bookmark (SBM) is a variety of web services on which users can divide up their bookmarks. Anyone can see anyone's bookmark on SBM. In order to solve some problems faced by conventional recommender systems, they expressed users' web page preference by "affinity level between each user and each tag." By this approach, users' preferences are abstracted and it becomes easier to find similar users. Clustering can also solve the problem of "tag redundancy in Folksonomy."

A hybrid web recommender system was described by [Taghipour *et al,* (2008)] They exploit an idea of combining the conceptual and usage information to enhance a reinforcement learning framework, primarily devised for web recommendations based on web usage data. Moreover the combination can improve the quality of web recommendations. A hybrid web recommendation method is proposed by utilizing the conceptual relationships among web resources to derive a novel model of the problem, enriched with semantic knowledge about the usage behavior. With their proposed hybrid model for the web page recommendation problem they revealed the pertinent and flexibility of the reinforcement learning framework in the web recommendation domain, and demonstrated how it can be extended in order to incorporate various sources of information. Their test results suggested that the method can improve the overall quality of web recommendations.

An intelligent recommender system was projected by kavitha [Devi *et al,* (2009)]. They designed and implemented an Intelligent Collaborative Recommender System (ICRS) to map users' needs to the items that can persuade them. A methodology is used to animatedly modernize the accuracy factor based on user intelligence. The diverse approaches for recommendation are categorized as memory-based and model-based approaches. Memory-based systems suffer from problems like data sparsity and scalability, whereas model-based approaches are liable to bind the range of users. Therefore they integrated these approaches to overcome their limitations. They applied the collaborative filtering approach for recommendations. Recommendations are made with high accuracy by applying regression to weighted aggregated predictions. Mean Absolute Error Metrics was considered for evaluating the performance of their proposed system. This approach thus improves scalability and sparsity problems and offers accurate recommendations.

[Cheng *et al*, (2009)] developed a two stage collaborative recommender system. They proposed a chronological pattern based collaborative recommender system that predicts the customer's time-variant acquisition behavior in an e-commerce environment where the customer's purchase patterns may change gradually. A new two-stage recommendation process is developed to envisage customer behavior for the selection of different categories, as well as for product items. Their study is the first to recommend time-decaying sequential patterns within a collaborative recommender system. Their experimental results revealed that the proposed system outperforms the traditional collaborative system.

[Lin *et al*, (2004)] described an efficient Adaptive-Support Association Rule Mining for Recommender Systems. They investigated the utilization of association rule mining as an underlying technology for collaborative recommender systems. Association rules have been used with sensation in other domains. Nevertheless, most currently existing association rule mining algorithms were designed with market basket analysis in mind. They described a collaborative recommendation technique based on a novel algorithm distinctively designed to excavate association rules for this rationale. The main advantage of their proposed approach is that their algorithm does not require the minimum support to be specified in advance. Rather, a target range is given for the number of rules, and the algorithm adjusts the minimum support for each user in order to obtain a rule set whose size is in the desired range. Moreover they employed associations between users as well as associations between items in making recommendations. The experimental evaluation of a system based on their algorithm revealed that its performance is significantly better than that of traditional correlation-based approaches.

Jung described a user-support method based on the distribution of knowledge with other users through the collaborative Web browsing, focusing exclusively on the user's interests extracted from users' own bookmarks. More prominently, they focused on those items of information which are associated to the user's interests. In collaborative Web browsing, they considered that recognizing the user's interests is an extremely essential mission. Furthermore, asking applicable information for other users, filtering the query results, and recommending them are additional most important tasks that have to be unreservedly conducted by them in. Based on the personalized TF-IDF proposal they introduced the extended application of a BIS Agent, which is a bookmark sharing agent system. Moreover they implemented an ontological supervisor which can carry out the semantic analysis of the Web sites pointed to by these bookmarks. They also designed a multi-agent system that consists of a facilitator agent and many personal agents. The main limitation of this system is that it does not consider the privacy problems related with sharing personal information of the user.

A novel recommender system was formulated by [Marko *et al*,. (1997)]. Their approach named, "Fab" is a recommendation system designed to help users sift through the mammoth amount of information obtainable in the World Wide Web. Their proposed approach is the combination of content-based filtering and collaborative filtering methods. The combination exploits the advantages of both the methods thereby avoiding the shortcomings. Fab's hybrid structure can be used for automatic recognition of emergent issues relevant to various groups of users. It also enables two scaling problems, pertaining to the rising number of users and documents, to be addressed. In general the content-based approach to recommendation has its pedigree in the information retrieval (IR) community, and utilizes many of the same techniques. The collaborative approach computed the similarity of the users rather than computing the similarity of the items. They maintained user profiles content analysis and directly compared these profiles to determine similar users for collaborative recommendation. The process of recommendation can be divided into two stages: collection of items to form a manageable database or index, and subsequently selection of items from this database for particular users. The experimental results using the hybrid Fab system achieved higher accuracy.

3. Methodology

3.1. *Recommendation using association rules*

Recommendation using association rules [Lee *et al,* (2003)] is to predict preference for item k when the user preferred item i and j, by adding confidence of the association rules that have k in the result part and i or j in the condition part. This association rules can be generated using apriori algorithm. Association rules capture relationships among items based on patterns of co-occurrence across transactions. In [Nakagawa et al, (2003)], association rules were applied to personalization based on web usage data. This approach was adapted to the context of collaborative filtering.

A set of user profiles U is given and a set of item sets $I = \{I1, I2, , Ik\}$, the support of an item set $Ii \in I$ is defined as $\sigma(Ii) = |\{u \in U : Ii \subseteq u\}| / |U|$. Item sets that satisfy a minimum support threshold are used to generate association rules. These groups of items are said as frequent item sets.

An association rule r represents an expression of the form $X \Rightarrow Y$ (σ r,σ r), where X and Y are item sets, σ r is the support of $X \cup Y$, and σr is the confidence for the rule r given by $\sigma (X \cup Y)/ \sigma (X)$. In addition, association rules that do not satisfy a minimum lift threshold are pruned, where lift is defined as $\sigma r/\sigma (Y)$. If there is no enough support for a particular item that item will never appear in any frequent item set. The implication is that such an item will never be recommended. The issue of coverage is a tradeoff. Lowering the support threshold will ensure that more items can be recommended, but at the risk of recommending an item without sufficient evidence of a pattern.

3.2. *FARM Algorithm*

For fuzzy association rule mining standard ARM algorithms can be used or at least adopted after some modifications [Muyeba *et al,* (2006)]. Less attention has been given to developing efficient algorithms for fuzzy association rule mining [Dubois *et al,* (2006)] but still; there are some contributions in this area [Gyenesei et al, (2008), Kuok *et al,* (1998) and Xu *et al,* (2003)]. An efficient algorithm was needed for the FARM methodology because a lot of pre-processing (filtration, conversions, normalization) and mining steps are involved in the web recommendation system. For this problem, ordinary Boolean ARM algorithms are inappropriate. Careful attention is needed in attribute partitioning because any clustering technique is not employed; and did this manually from input data.

The FARM algorithm is presented below and the notations used are given as follows: Lists of items
List may be presented with each item marked by bullets and numbers.

Algorithm Notations

$I_{websites}$	Available websites
I_{wsa}	Converted web system attributes
D	weblog database
T	set of ordinary logs
$T_{websites}$	logs with available websites
T_{wsa}	WSA converted logs
D_{wsa}	WSA logs database
T_{fuzzy}	fuzzy logs
D_{fuzzy}	fuzzy logs database
F_k	set of frequent k-itemsets
C_k	set of candidates k-itemsets
I	set of complete item sets
mincorr	minimum correlation value
minsupp	minimum support
minconf	minimum confidence

FARM algorithm consists of four major steps:
1. Filtration and transformation of ordinary web logs database into a database with average (Web System Attributes) WSA logins.
2. Appropriate and accurate transformation of WSA logs' into a database containing fuzzy extensions. Normalization is performed for this database.
3. Candidate generation and search for all fuzzy frequent itemsets within candidates that have fuzzy support higher than user specified minimum support.
4. Use of frequent itemsets to generate the desired most relevant websites by calculating the fuzzy confidence and correlation values.

FARM comprises of the following algorithmic components:

$\boldsymbol{WSAConverter}(\boldsymbol{T}, \boldsymbol{WSA}, \boldsymbol{I_{websites}})$

1. $\forall T \ in\ D$

2. $\forall I \ in\ T$

3. $\boldsymbol{if}\ (websites\ ==\ check(I_j, I_{websites})$

4. $T_{websites} = T_{websites} \cup (I_j)$

5. $T_{wsa} = averageWSA(T_{websites})$

6. $D_{wsa} = wrirte(T_{websites})$

7. $\boldsymbol{end}$;

$\boldsymbol{WSA - FuzzyConverter(T_{wsa}, FuzzyLogs)}$

1. $\forall T_{wsa}\ in\ D_{wsa}$

2. $\forall I_{wsa}\ in\ T_{wsa}$

3. $\boldsymbol{fuzzyattr} = getFuzzyAttr(I_{wsa}, Fuzzylogs)$

4. $T_{fuzzy} = T_{fuzzy} \cup (fuzzyattr)$

5. $D_{fuzzy} = wrirte(T_{fuzzy})$

7. $\boldsymbol{end}$;

$\boldsymbol{FHARM(minsupp, minconf, mincorr, T_{fuzzy})}$

1. $\boldsymbol{K} = 0;\ C_k = \varnothing;\ F_k = \varnothing$

2. $\boldsymbol{Do}$

3. $\boldsymbol{k} = k + 1$

4. $\boldsymbol{if}(k == 1)$

5. $\boldsymbol{C}_k = GenerateFirstCandidates\ (T_{Fuzzy})$

6. $\boldsymbol{else}$

7. $\boldsymbol{C}_k = Generatecandidates(F_{k-1})$

8. $\forall Ck$

9. $\boldsymbol{count} = CountSupport(C_k)$

10. $\boldsymbol{C}_k = PruneCandidates(C_k, count, minsupp)$

11. $\boldsymbol{C}_k = CalcSignificance(C_k, minconf)$

12. $\boldsymbol{F}_k = GenerateFrequentItemsets(C_k, minconf)$

13. $\boldsymbol{F} = F \cup F_k$

14. $\boldsymbol{while}(C_k.count > k)$

15. $\boldsymbol{cfactor} = CalcCertainty(F, mincorr)$

16. $\boldsymbol{Output}\ (Rules(F, mincorr, cfactor))$

The Fuzzy ARM Algorithm belongs to the breadth first traversal family of ARM algorithms, developed using tree data structures and it works in a fashion quite similar to the Apriori algorithm [Jon Herlocker et al, (1999)]. Also, implementation of this proposed approach is different from [Gyenesei et al, (2008), Kuok et al, (1998)] by avoiding an extra database scan to find correlation values, thus increasing efficiency.

4. Experimental results

Set of experiments are performed in this section for evaluating the impact of proposed system on the predication process. Overall these experiments have verified the effective of the proposed techniques in web page recommendation. This experiment result shows how the new fuzzy HARM approach gives performs for the DePaul CTI Web server (http://www.cs.depaul.edu) data set. The data is obtained based on a random sample of users visiting this site for a 2 week period. The original (unfiltered) data contain a total of 20950 sessions from 5446 users. The filtered data files were produced by filtering low support page views, and eliminating sessions of size 1. The filtered data contains 13745 sessions and 683 page views. The data sets are split into two non-overlapping time windows to form training and a test data set. 70% of the data set (9745 sessions) was used as the training set and the remaining was used to test the system. For the evaluation, each user session is presented to the system, and recorded the recommendations it made after seeing each page the user had visited. The system was allowed to make n recommendations in each step with n < 10 and n < l , where l is the number of outgoing links of the last page visited by the user.

Quality Measures

This experiment shows how the new fuzzy HARM approach gives more interesting rules than the previous one using ARM algorithm. Figure 1 shows the difference between the number of large item sets generated from the previous method and the new FARM approach using different fuzzy support values. The number of large item sets increases as the minimum support decreases, naturally. The graph is plotted against fuzzy support and Frequent item sets.

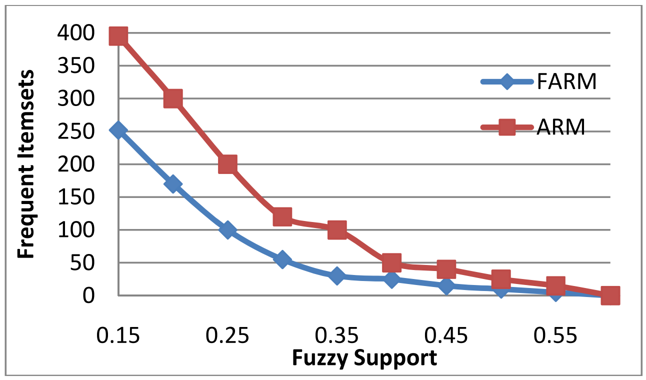

Figure 1: Number of frequent Item sets comparison

From the results, it is clear that the approach with normalization produces less frequent item sets (or even rules) than the converse. This is because during the normalization process, the fuzzy degree of fuzzy sets is averaged thus making the data more dissimilar and consequently fewer rules. The problem of producing many rules is easily handled by introducing the fuzzy interestingness measures.

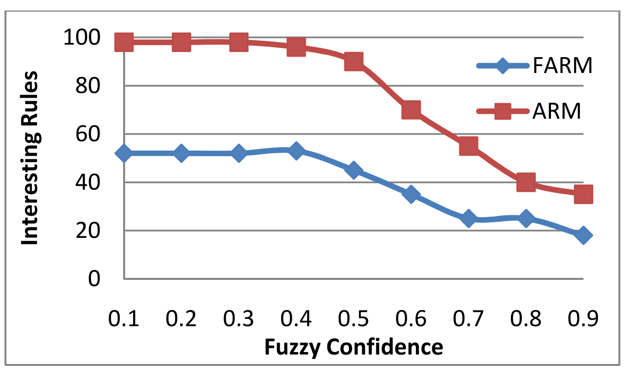

Figure 2: Number of Interesting Rules comparison

Figure 2 shows the number of interesting rules using user specified fuzzy confidence and fuzzy correlation values respectively. The graph is plotted against fuzzy confidence and interesting rules.

The number of interesting rules is less for the FARM when it is compared to the number of rules in figure1.

Performance Measures

In this experiment the performance measure of our new approach by varying the number of attributes and the size of data with and without normalization are shown. The support threshold is 0.20, confidence is 0.6 and correlation value is 0.5.

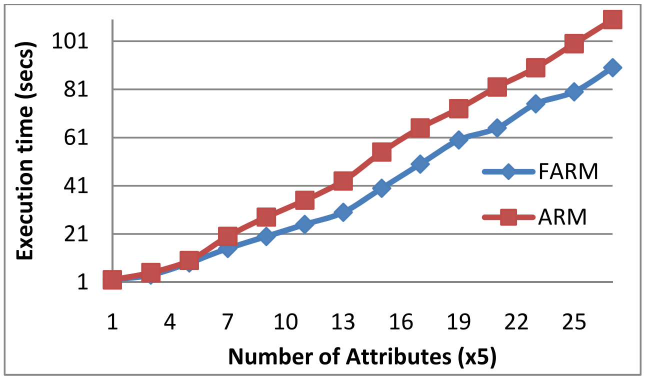

Figure 3: Performance Measures: Number of Attributes

Figure 3 shows the execution time of our proposed algorithm with different number of attributes. The graph is plotted against Number of Attributes and Execution time represented in seconds. Execution time increases as the number of attributes are increased. The FARM algorithms have lesser timings while comparing with the ARM algorithm execution time. When the number of attributes is increased then the number of rules also increases with more attributes but fixed transactions. It is intuitive that using more attributes increases the problem's dimension.

5. Conclusion

In this paper, new fast algorithms for mining association rule to provide web recommendation have been proposed. It can be concluded that the proposed approach is capable of making web recommendation more accurately and effectively against the conventional method. By combining similarity between rules and active user and confidence of the weighted rules, the recommendation engine has selected only the most relevant pages. Therefore, it increases the efficiency of the recommendation engine. The simulation results show that the performance of proposed FARM outperforms the existing collaborative web recommendation algorithm by means of time consumption, quality and performance. The proposed algorithm guides the society by providing the most related web sites that is needed.

References

[1] Balabanovic Y. and Y. Shoham, "Fab: content-based, collaborative recommendation", Communications of the ACM, Vol. 40, Issue 3, 1997, pp. 66-72.
[2] Cheng-Lung Huang, and Wei-Liang Huang, "Handling sequential pattern decay: Developing a two-stage collaborative recommender system," Elsevier, Electronic Commerce Research and Applications, vol. 8, no. 3, pp. 117-129, 2009.
[3] Dubois, D., Hüllermeier, E. and Prade, H., "A Systematic Approach to the Assessment of Fuzzy Association Rules," Data Mining and Knowledge Discovery Journal, Vol. 13, No. 2, (2006), 167 – 192.
[4] Gabor Takacs, Istvan Pilaszy, Bottyan Nemeth, and Domonkos Tikk, "Scalable Collaborative Filtering Approaches for Large Recommender Systems," The Journal of Machine Learning Research, vol. 10, pp. 623-656, 2009.
[5] Gery M. and H. Haddad, "Evaluation of Web Usage Mining Approaches for User's Next Request Prediction", Proceedings of the fifth ACM international workshop on Web information and data management, 2003, pp. 74–81.
[6] Gyenesei, A., "A Fuzzy Approach for Mining Quantitative Association Rules," Acta Cybernetical, Vol. 15, No. 2, (2001), 305-320.
[7] Hill, W., L. Stead, M. Rosenstein, and G. Furnas, "Recommending and evaluating choices in a virtual community of use", in Proceedings of CHI'95.
[8] Jason J. Jung, "Collaborative Web Browsing Based on Semantic Extraction of User Interests with Bookmarks," Journal of Universal Computer Science, vol. 11, no. 2, pp. 213-228, 2005.
[9] Jian Chen, Roman Y. Shtykh, and Qun Jin, "A Web Recommender System Based on Dynamic Sampling of User Information Access Behaviors," Ninth IEEE International Conference on Computer and Information Technology, vol. 2, pp. 172-177, 2009.
[10] Jon Herlocker, Joseph Konstan, Al Borchers, John Riedl, "An algorithmic framework for performing collaborative filtering", Proceedings of the 1999 Conference on Research and Development in Information Retrieval pp263-266, 1999.
[11] Kavitha Devi M.K., and P. Venkatesh, "ICRS: an intelligent collaborative recommender system for electronic purchasing," International Journal of Business Excellence 2009, vol. 2, no. 2, pp. 179-193, 2009.

[12] Kuok, C., Fu, A., & Wong, H., "Mining Fuzzy Association Rules in Databases". ACM SIGMOD Record, 27, (1998), 41-46.

[13] Lee, C.-H., Chen, M.-S., Lin, C.-R. , "Progressive Partition Miner, An Efficient Algorithm for Mining General Temporal Association Rules," IEEE Transactions on Knowledge and Data Engineering, Vol. 15, No. 4, (2003), 1004 - 1017.

[14] M. Nakagawa and B. Mobasher, " A hybrid web personalization model based on site connectivity", in WebKDD Workshop at the ACM SIGKKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, August 2003.

[15] Marko Balabanovic, and Yoav Shoham, "Fab: content-based, collaborative recommendation," Communications of the ACM, vol. 40, no. 3, pp. 66-72, 1997.

[16] Mobasher B. , R. Cooley, J. Srivastava, "Automatic Personalization based on Web Usage Mining", Communications of the ACM, 43 (8), 2000, pp. 142-151.

[17] Muyeba M, M. Sulaiman Khan, M. Malik, Z. and Tjortjis, C. "Towards Healthy Association Rule Mining (HARM), A Fuzzy Quantitative Approach", In Proc. 7th Conf. Intelligent Data Engineering Automated Learning, Lecture Notes in Computer Science Series, Vol. 4224, Springer-Verlag, (2006), 1014-1022.

[18] Nima Taghipour, and Ahmad Kardan, "A hybrid web recommender system based on Q-learning," Proceedings of the 2008 ACM symposium on Applied computing, pp. 1164-1168, 2008.

[19] Rakesh Agrawal and Ramakrishnan Srikant," Fast Algorithms for Mining Association Rules", IBM Almaden Research.

[20] Sarwar B.M. "Item-Based Collaborative Filtering Recommendation Algorithms," 10th International World Wide Web Conference, ACM Press, 2001, pp. 285-295.

[21] Satoshi Niwa, Takuo Doi, and Shinichi Honiden, "Web Page Recommender System based on Folksonomy Mining," Proceedings of the Third International Conference on Information Technology: New Generations, IEEE Computer Society, pp. 388-393, 2006.

[22] Weiyang Lin, Sergio A. Alvarez, and Carolina Ruiz, "Efficient Adaptive-Support Association Rule Mining for Recommender Systems," Journal on Data Mining and Knowledge Discovery, vol. 6, no. 1, pp. 83-105, 2004.

[23] Xu B., J. Lu, Y. Zhang, L. Xu, H. Chen and H. Yang. "Parallel Algorithm for Mining Fuzzy Association Rule", Int'l Conf. on Cuberworld, Singapore, 2003.