International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 25-30

Enhanced Web Page Recommendation System Using Associative Rule Mining and Clustering Model

A.Roseline Immaculate[1], Dr.R.Suguna[2]
[1](*Department of Computer Science, Theivanai Ammal College for Women, India)*
[2](*Department of Computer Science, Theivanai Ammal College for Women, India)*

Abstract:*Association Rule Mining (ARM) is one of the most popular data mining techniques. Recently, a huge amount of data are being generated and received at the World Wide Web (WWW). To further enhance the growth of web users, the generated data has to be stored and processed efficiently to reduce the searching time complexity. Thus, a web recommendation model has been suggested by several researchers. This paper concentrates on developing a recommendation model that shares the common interest of different web users using association rule mining process. By the deployment of ARM, the common patterns of data have discovered. These common patterns are further clustered in a hierarchical form. A hierarchical clustering model recommends the upcoming users for easier data searching and retrieval process. Experimental results have shown the efficiency of the proposed system. And lesser computation time with reliable recommendation model have achieved.*

Keywords: *Association Rules, Clustering, Data Mining, Data Searching, and Web recommendation*

I. Introduction

With the rapid growth of internet technologies, Web has become a huge repository of information and keeps growing exponentially under no editorial control. However the human capability to read, access and understand content remains constant. Hence it became more challenging to the Website owners to selectively provide relevant information to the people with diverse needs. Modeling and analyzing Web navigational behavior is helpful in understanding the type of information online user's demand. This motivated researchers to provide Web personalized online services such as Web recommendations to alleviate the information overload problem and provide tailored Web experience to the Web users. In recent times, Web Usage Mining has emerged as a popular approach in providing Web personalization [1]. However conventional Web usage based recommender systems are limited in their ability to use the domain knowledge of the Web application and their focus is only on Web usage data. As a consequence thequality of the discovered patterns is low. These patterns do not provide explicit insight into the user's underlying interests and preferences, thus limiting the effectiveness of recommendations as well as the ability of the system to interpret and explainthe recommendations.

Association rule learning is a popular and well-researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness [1]. Based on the concept of strong rules, Rakesh Agrawal et al.[2] introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets.Following the original definition by Agrawal et al.[2] the problem of association rule mining is defined as: Let $I = \{i_1, i_2, \ldots, i_n\}$be a set of n binary attributes called items. Let $D = \{t_1, t_2, \ldots t_m\}$be a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I. A rule is defined as an implicationof the form $X \Rightarrow Y$ where $X, Y \subseteq I$and $X \cap Y = \emptyset$.The sets of items X and Y are called antecedent (left-hand-side or LHS) and consequent (right-hand-side or RHS) of the rule respectively.Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. Association rule generation is usually split up into two separate steps:

- First, minimum support is applied to find all frequent itemsets in a database.
- Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

While the second step is straightforward, the first step needs more attention.Many algorithms for generating association rules were presented over time.
Some well known algorithms are Apriori, Eclat and FP-Growth, but they only do half the job, since they are algorithms for mining frequent itemsets. Another step needs to be done after to generate rules from frequent itemsets found in a database.

The contributions made in this paper are:

a) About the current issues prevails in Web Recommendation model have studied.
b) Since, data searching time is a challenging task in this model, have learnt about the primitives of ARM and the clustering techniques.
c) An enhanced web recommendation model has designed which reduces the searching time complexity under clustering technique.
d) Experimental studies have shown the effectiveness of the approach in terms of time complexity.

The paper is organized as follows: Section 2 describes the related work in ARM, Section 3 describes the proposed techniques; Section 4 portrays the experimental analysis and concludes in Section 5.

II. Association Rule Mining Approaches

This section reveals the existing studies suggested by other researchers. The author in [3] studied pattern mining concept in form of market based analysis for finding association between items bought in a market. It focuses on improving the quality of databases together with necessary functionality to process decision support queries. The author in [4] suggested apriori algorithm which is a straightforward approach that requires many passes over the database, generating many candidate itemsets and storing counters of each candidate while most of them turn out to be not frequent. The author in [5] discussed about the Multiple dimensional association rule mining is to discover the correlation between different predicts/attributes. Each attribute/predict is called a dimension, such as: age, occupation and buys in this example. At the same time multiple dimensional association rule mining concerns all types of data such as Boolean data, categorical data and numerical data.The giant amount of data poses a challenge of maintaining and updating the discovered rules while the data may change from time to time in different ways.The FUP (Fast UPdate) algorithm [6] was introduced to deal with insertion of new transaction data.

The author in [7] discussed Multiple level association rule mining is trying to mine strong association rules among intra and inter different levels of abstraction. For example, besides the association rules between milk and ham, it can generalize those rules to relation between drink and meat, at the same time it can also specify relation between certain brand of milk and ham.In order to improve the efficiency of existing mining algorithms, constraints were applied during the mining process to generate only those association rules that are interesting to users instead of all the association rules [8]. The author in [9] suggested frequent itemsets are generated with only two passes over the database and without any candidate generation process.By avoiding the candidate generation process and less passes over the database, FP-Tree is an order of magnitude faster than the Apriori algorithm. In [10], RARM is claimed to be much faster than FP-Tree algorithm with the experiments result shown in the original paper. By using the SOTrieIT structure RARM can generate large 1-itemsets and 2-itemsets quickly without scanning the database for the second time and candidate's generation.

A new algorithm named Inverted Hashing and Pruning (IHP) [11] for mining association rules between items in transaction databases. The performance of the IHP algorithm was evaluated for various cases and compared with those of two well-known mining algorithms, Apriori algorithm. It has been shown that the IHP algorithm has better performance for databases with long transactions.Fuzzy grids based rules mining algorithm (FGBRMA) is introduced by [12] to generate fuzzy association rules from a relational database. The proposed algorithm consists of two phases: one to generate the large fuzzy grids, and the other to generate the fuzzy association rules. A numerical example is presented to illustrate a detailed process for finding the fuzzy association rules from a specified database, demonstrating the effectiveness of the proposed algorithm.The author in [13] suggested a new clustering method, called HBM (Hierarchical Bisecting Medoids Algorithm) to cluster users based on the time-framed navigation sessions. Those navigation sessions of the same group are analyzed using the association-mining method to establish a recommendation model for similar students in the future. Finally, an application of this recommendation method to an e-learning web site is presented, including plans of recommendation policies and proposal of new efficiency measures. The effectiveness of the recommendation methods, with and without time-framed user clustering, are investigated and compared.

The author in [14] presented an efficient algorithm named cluster-based association rule (CBAR). The CBAR method is to create cluster tables by scanning the database once, and then clustering the transaction records to the *k*-th cluster table, where the length of a record is *k*. Moreover, the large itemsets are generated by contrasts with the partial cluster tables. This not only prunes considerable amounts of data reducing the time needed to perform data scans and requiring less contrast, but also ensures the correctness of the mined results. The author in [15] presented a new approach for constructing a classifier, based on an extended association rule mining technique in the context of classification. The characteristic of this approach is threefold: first, applying the information gain measure to the generation of candidate itemsets; second, integrating the process of frequent itemsets generation with the process of rule generation; third, incorporating strategies for avoiding rule redundancy and conflicts into the mining process. Classification Association Rule Mining (CARM) [16]

systems operate by applying an Association Rule Mining (ARM) method to obtain classification rules from a training set of previously classified data. The rules thus generated will be influenced by the choice of ARM parameters employed by the algorithm (typically support and confidence threshold values). In this paper examine the effect that this choice has on the predictive accuracy of CARM methods.

The weighted association rules (WARs) [17] mining are made because importance of the items is different. Negative association rules (NARs) play important roles in decision-making. But the misleading rules occur and some rules are uninteresting when discovering positive and negativeweighted association rules (PNWARs) simultaneously. The author in [18] studied about the negative association rules become a focus in the field of data mining. Negative association rules are useful in market-basket analysis to identify products that conflict with each other or products that complement each other. The negative association rules often consist in the infrequent items. The experiment proves that the number of the negative association rules from the infrequent items is larger than those from the frequent.

Association Rules Mining based Alarm Correlation Analysis System (ARM-ACAS) was suggested by [19] to find interesting association rules between alarm events. In order to mine some infrequent but important items, ARM-ACAS first uses neural network to classify the alarms with different levels. In addition, ARM-ACAS also exploits an optimization technique with the weighted frequent pattern tree structure to improve the mining efficiency. The author in [20] suggested a fuzzy association rules to address the first limitation. In this they put forward a discovery algorithm for mining both direct and indirect fuzzy association rules with multiple minimum supports to resolve these three limitations. Then, a new approach (PNAR_IMLMS) [21] for mining both negative and positive association rules from the interesting frequent and infrequent item sets mined by the IMLMS model. The experimental results show that the PNAR_IMLMS model provides significantly better results than the previous model.

The traditional algorithms for mining association rules are built on binary attributes databases, which has two imitations [22]. Firstly, it cannot concern quantitative attributes; secondly, it treats each item with the same significance although different item may have different significance.A variable neighbourhood search (VNS) [23] algorithm is developed to solve the problem with near-optimal solutions. Computational experiments are performed to test the VNS algorithm against a benchmark problem set. The results show that the VNS algorithm is an effective approach for solving the MTFWS problem, capable of discovering many large-one frequent itemset with time-windows (FITW) with a larger time-coverage rate than the lower bounds, thus laying a good foundation for mining ARTW. The author in [24] numerical ARM problem using a multi-objective perspective by proposing a multi-objective particle swarm optimization algorithm (i.e., MOPAR) for numerical ARM that discovers numerical association rules (ARs) in only one single step. To identify more efficient ARs, several objectives are defined in the proposed multi-objectiveoptimization approach, including confidence, comprehensibility, and interestingness. Finally, by using the Pareto optimality, the best ARs are extracted.

III. Proposed Technique

This section describes the proposed technique employed for achieving better web recommendation model. Weblog is the most important entity which depicts the interesting patterns of the web users. It helps to navigate on other web pages under a defined server. Association Rule Mining is widely used in web transactional process which narrates the relationship based on the hitting of page views. The following are the steps involved in proposed system:

a) Input: Weblogs chosen from the server, www.cs.depaul.edu site
b) The logs mostly compose of redundant data that seeks data mining tasks, association rule process.
c) Data cleaning is done on the weblogs which removes the redundant data.
d) Data selection is done to select the relevant features required for designing recommendation model.
e) The selected features are then stored onto the web database.
f) Hierarchical clustering model is applied which segments the data and assign with an index number for further clarification.
g) The clustered data are used for estimating minimum support and minimum confidence values.
h) Thus, the output of ARM presents the similar patterns and thus web pages of the client are then evaluated.

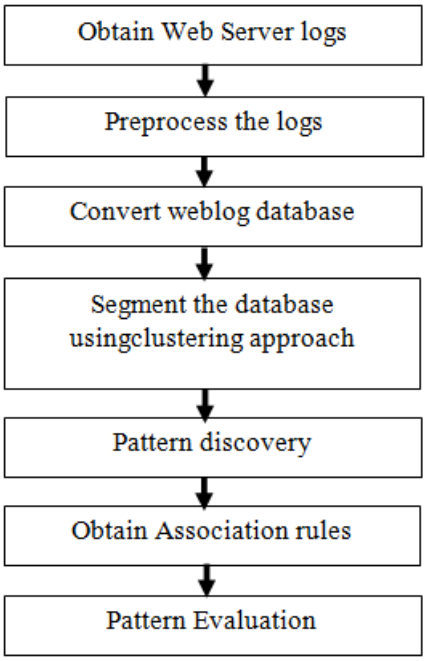

Fig.1. Proposed workflow

IV. Experimental Analysis

This section presents the experimental analysis of proposed web recommendation model. The following are the design goals of model,

a) Collection of webserver logs: It's a file administered by the web server which is in log format.
b) Preprocessing the logs: Generally, it composes of variant web requests which are in incomplete form. Preprocessing helps to remove the duplicate data to obtain better accuracy.
c) Conversion of log file into database: The logs are not directly used as input to ARM process. Thus, it is converted into MYSQL database table.
d) Partitioning the database: Relied upon the support count, the records are clustered in hierarchical form. It helps to find out the user's interest.
e) Discovery of patterns: Within the cluster group, the similar patterns are recognized.
f) Association rules: It describes the relationship between different itemsets.
g) Evaluating the patterns: The extracted patterns are again evaluated and thus interpretation is done for developing recommendation model.

Since, Association rule mining is the basic concept of the proposed web recommendation model. It is measured from support and confidence values of the predicted rules.

a) Support: The support of the web page (P) is estimated as the proportion of the transaction that contain relevant web page.
b) Confidence: The confidence of the rule is defined from the eqn (1):

$$Conf(P_i \longrightarrow P_n) = \frac{supp\ (P_i \cup P_n)}{supp\ (P_i)} \quad (1)$$

Table 1: Pattern discovery

Rule No.	Rule	Confidence	Predicted subsequent page
R1	P16^P17^P15^P2	0.217	P1
R2	P16∧P17∧P15∧P	0.219	P6
R3	P16∧P17∧P15∧P3	0.312	P6
R4	P16∧P17∧P5∧P7	0.314	P1
R5	P16∧P17∧P8∧P4	0.215	P1

From the analyzed associated rules, the performance metrics studied are the precision and coverage.

a) Precision

Precision is defined as the prediction of accurate recommendation for all the test users. It depicts the quality of the each individual recommendation. It is given as in eqn (2):

$$Precision = \frac{(T(p) \cap R(p))}{R(p)} \quad (2)$$

Where R(p) is the set of recommendation and T(p) is the session. It generally varies based on the pages recommended.

b) Coverage

Coverage is the proportion of relevantrecommendations to the all pages that should be recommended. It given as in eqn (3): $Coverage\ (P_n \rightarrow P_i) = Support\ (P_i)$ (3)

Table 2: Precision & Coverage analysis

No.of recommended pages	Precision analysis	Coverage analysis
1	98.7	63.10
2	89.36	66.66
3	85.71	62.57
4	86.36	69.58
5	96.31	67.23
1	98.7	61.39

V. Conclusion

Web Mining is defined as an application of data mining techniques on the navigational traces of the users to extract knowledge about their preferences and behavior to develop recommendation model for further assistances. The knowledge discovered from web mining can be useful in many Web applications such as Web caching, Web prefetching, intelligent online advertisements, in addition to Web recommendation systems. Most of the research efforts in Web personalization correspond to the evolution of extensive research in Web Mining.In this paper, the existing association rules mining in data mining applicationsbriefly reviewed. This review would be helpful to researchers to focus on the various issues of web mining system. An enhanced web recommendation model has developed which reduces the searching complexity. With the ARM and clustering techniques as base, novel web recommendation systems seek out similar patterns of the web users. This similar pattern are then clustered for designing the system and tested on the web metrics. Experimental analysis has shown the efficiency of proposed technique.

References

[1]. William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus, Knowledge Discovery in Databases: An Overview, *AI Magazine,13(3)* ,1992.

[2]. Jochen Hipp, Ulrich Guntzer, and Gholamreza Nakhaeizadeh,Algorithms for Association Rule Mining – A General Survey and Comparison, *Volume 2, Issue 1* ,2000,pp58 .

[3]. Agrawal.R, Imielinski.T, and Swami.A.N, Mining Association Rules between Sets of Items in Large databases, *ACM New York, NY, USA*, 1993, 207-216.

[4]. R.Agrawal, and R.Srikant, Fast algorithms for mining association rules, 1994, 487-499.

[5]. R.Srikant, and R.Agrawal, Mining quantitative association rules in large relational tables, *ACM Press*, 1996, 1-12.

[6]. D. W.Cheung, S. D.Lee, and B .Kao, A general incremental technique for maintaining discovered association rules. In Database Systems for Advanced Applications, 1997, 185-194.

[7]. J. Han, and M.Kamber, *Data Mining Concepts and Techniques* (Morgan Kanufmann, 2000).

[8]. J.Pei, and J.Han, Can we push more constraints into frequent pattern mining?,*ACM Press*, 2000,350-354.

[9]. J.Han, and J.Pei, Mining frequent patterns by pattern-growth: methodology and implications, *ACM SIGKDD Explorations Newsletter 2, 2*, 2000, 14-20.

[10]. A.Das, W.-K.Ng, and Y.-K.Woon, Rapid association rule mining, ACM Press, 2001, 474 - 481.

[11]. John D. Holt, and Soon M. Chung, M ining association rules using inverted hashing and pruning, *Elsevier, Volume 83, Issue 4*, 31 August 2002, 211-220.

[12]. Yi-Chung Hu, Ruey-Shun Chen, and Gwo-Hshiung Tzeng, Discovering fuzzy association rules using fuzzy partition methods, Elsevier, Volume 16, Issue 3, April 2003, 137-147.

[13]. Feng-Hsu Wang, and Hsiu-Mei Shao, Effective personalized recommendation based on time-framed navigation clustering and association mining, Elsevier ,Volume 27, Issue 3, October 2004, 365-377.

[14]. Yuh-Jiuan Tsay, and Jiunn-Yann Chiang, BAR: an efficient method for mining association rules, Elsevier, Volume 18, Issues 2–3, April 2005, 99-10.

[15]. Guoqing Chen, Hongyan Liu, Lan Yu, Qiang Wei, and Xing Zhang, A new approach to classification based on association rule mining,Elsevier, Volume 42, Issue 2, November 2006, 674-689.

[16]. Frans Coenen, and Paul Leng, The effect of threshold values on association rule based classification accuracy, *Elsevier, Volume 60, Issue 2*, February 2007, 345-360.
[17]. He Jiang, Yuanyuan Zhao, and Xiangjun Dong, Mining Positive and Negative Weighted Association Rules from Frequent Itemsets Based on Interest, *IEEE ,vol.2*, 2008, 242,245.
[18]. YuanyuanZhao, He Jiang, RunianGeng, and Xiangjun Dong, Mining Weighted Negative Association Rules Based on Correlation from Infrequent Items, *IEEE*, 2009, 270-273.
[19]. Tongyan Li, and Xingming Li, Novel alarm correlation analysis system based on association rules mining in telecommunication networks, *Elsevier, Volume 180, Issue 16*, 15 August 2010, 2960-2978.
[20]. WeiminOuyang, and Qinhua Huang, Mining direct and indirect fuzzy association rules with multiple minimum supports in large transaction databases, *IEEE, vol.2*, 2011,947-951.
[21]. WeiminOuyang, Mining Positive and Negative Fuzzy Association Rules with Multiple Minimum Supports, *IEEE*, 2012.
[22]. Anjana Gosain, and Maneela Bhugra, A Comprehensive Survey of Association Rules On Quantitative Data In Data Mining, *IEEE*, 2013.
[23]. Yiyong Xiao, Yun Tian, and Qiuhong Zhao, Optimizing frequent time-window selection for association rules mining in a temporal database using a variable neighbourhood search, *Elsevier Volume 52*, December 2014, 241-250.
[24]. Vahid Beiranvand, Mohamad Mobasher-Kashani, and Azuraliza Abu Bakar, Multi-objective PSO algorithm for mining numerical association rules without a priori discretization, *Elsevier, Volume 41, Issue 9*, July 2014, 4259-4273.