iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: VII Month of publication: July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45559

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Web Personalized Recommendation Model Using Temporal Fuzzy Association Rule Mining

Likhith B[1], B Praveen Nayak[2], Yeshwanth S P[3], Dr. Suneetha K R[4]
[1, 2, 3, 4]Department of Computer Science and Engineering, Bangalore Institute of Technology, Bangalore-560004, Karnataka, India

Abstract: A web access log file contains timely sequenced log entries which include essential fields to indicate user activities. Analysis of these patterns provides valuable information for web designer to quickly respond to their individual needs. Many industries are struggling to retain regular interested customers for the improvement of customer relationship. Retrieval of relevant information automatically from these log files for interested group of users is a difficult process, since acquiring interested user profiles which evolves continuously with respect to time are not so easy. The paper presents a novel Web Personalized Recommendation Model (WPRM) using temporal fuzzy association rule mining technique. Temporal Fuzzy Association Rule Mining (FTARM) technique is proposed and applied on a focused set of interested users to provide intelligent recommendations. The proposed model results in less execution time and reduced memory utilization with high accuracy.
Key words- Fuzzy Association Rule, Decision Trees, Time Stamp, Popular page pattern, Recommendations.

I. INTRODUCTION

The explosive growth of data available on the net has made the analysis and discovery of useful information more difficult. When browsing the web, without proper guidance, users often wander aimlessly without visiting the web pages of their interests and then leave the web site soon after losing their interests (Jaideep, Prasanna & Vipin, 2002 ; Tsuyoshi & Kota, 2006). Thus, the web systems need to guess varied interests of different users. To satisfy different users, the web system should be able to distinguish between different users or groups of users and their needs to be able to predict the user's needs. Web personalization is necessary in order to solve the above problems. The currently available techniques for web personalization are not sufficient to extract relevant information for web users since the cookies and other mechanisms used by current search engines are not providing accurate algorithms for mining web user profiles. Hence the paper presents a new model WPRM, which proposes an algorithm called Fuzzy-Temporal Association Rule Mining Algorithm (FTARM) to classify the interested web user profiles dataset periodically to know the users behaviors and interests based on temporal pattern analysis. The proposed model consists of two phases. In first phase, the web server log data is preprocessed and is classified into focused set of interested users using enhanced version of Decision tree C4.5 algorithm (Robert, Mobasher & Srivastava, 1997; Suneetha & Krishnamoorti, 2010a; Suneetha & Krishnamoorti, 2010b)

In the second phase Fuzzy Association Rule Mining (FARM) technique is proposed to provide intelligent recommendations.

Association Rule Mining (ARM) is an important and well established data mining technique used to identify patterns expressed in the form of association rules from transactional data sets (Bodon, 2003; Coenen, Leng, & Goulbourne 2004; Agrawal, Imielinski & Swami, 1993). The attributes in ARM data sets are usually binary valued but ARM also can be applied to quantitative and categorical (non-binary) data (Gyenesei, 2001; Srikant & Agrawal, 1996; Ye & Keane 1997). With the latter, values can be split into linguistically labeled ranges (for example "low", "medium", "high" etc) such that each range represents a binary valued attribute. Values can be assigned to these range attributes using crisp or fuzzy boundaries. The application of ARM using the latter is referred to as Fuzzy Association Rule Mining (FARM) **(**Kuok, Fu & Wong, 1998)**.** FARM has been shown to produce more expressive association rules than the "crisp" methods. Fuzzy logic deals with approximate rather than precise modes of reasoning and proposes three different types of qualifications named as (i) Truth qualification, for example " not quite true" ,(ii) probability-qualification, something is "unlikely" (iii) Possibility-qualification, might be expressed by "almost impossible". Fuzzy association rules is an implication of the form: if *A, X* then *B, Y,* where A and B is disjoint itemsets and X and Y are fuzzy sets. Fuzzy sets are generalized sets which allow for a graded membership of their elements.

The paper provides suitable experimental analysis for the proposed fuzzy logic based temporal association rule mining approach in which fuzzy logic is used for intelligent classification. This reduces the search space of the web user profiles dataset. These rules play an important role in prediction of users' next access more precisely. In this algorithm, the temporal constraints are used because the different users group accessing the internet are in different time periods.

Therefore, the users temporal data is stored classified, analyzed and the relevant rules are extracted. To access relevant web pages, relevancy factor is computed using the term frequency. For this purpose, the query words given by the user while searching are compared with each string present in the document and the words which have high matches based on a threshold value are considered for the retrieval of top ten pages. Then these pages are shown to the user to get his relevance feedback. Once user is satisfied with the pages, ontology is created with semantic to improve the performance of the semantic analysis process.

The rest of the paper is organized as follows. Section 2 briefs the related work. The proposed approach and its details are presented in section 3. Section 4 provides Fuzzy Temporal Association Rule Mining algorithm. Section 5 discusses the results and finally, conclusions are drawn in section 6.

II. RELATED WORK

In recent years abundant work has been carried out in the area of web mining, specifically on analysis of web log data. There are many works carried out on web usage mining (Olfa, Maha,Esin,Antonio & Richard, 2008; Yuefeng & Ning, 2006) which deal with various data mining or machine learning techniques to model and understand web user activity. The clustering technique proposed in (Hofgesang, 2009; Yan, Jacobsen,Garcia & Dayal,1996), is used to segment user sessions into clusters or profiles that can later form the basis for personalization. The notion of an adaptive web site was proposed in (Perkowitz & Etzioni,1997); where the user's access pattern is used to automatically synthesize index pages. Based on association rule mining discovery of web user activity model is proposed in (Srivastava, Cooley, Deshpande & Tan, 2000); whereas the approach proposed in (Ma, Pant & Sheng, 2007) uses probabilistic grammars to model web navigation patterns for the purpose of prediction. Web utilization miner presented in (Spiliopoulou & Faulstich, 1998) discovers navigation patterns with user-specified characteristics over an aggregated materialized view of the web log. New fuzzy relational clustering techniques are used in (Dimitrios, & Georgios 2010; Mangesh, Dr. Bharat & Ramprasad, 2008) to discover user profiles that are resistant to noise that are present in click stream data. A robust density-based evolutionary clustering technique was proposed in (Castellano, Fanelli & Torsello, 2008) to discover an optimal number of multi resolution and robust user profiles. Most researchers (Desikan & Srivastava, 2004; Nasraoui, Rojas & Cardona, 2006; Mofreh, Miroslav & Pawan 2003) in the data mining community have focused their efforts on finding efficient algorithms for analyzing huge amounts of data. Temporal usage mining involves application of data mining techniques on web usage data to discover temporal patterns which describe the temporal behavior of web users.

A number of research work have been concentrated on applying data mining techniques on to preprocessed web access log data to identify behavior of frequent users (Suneetha & Krishnamoorti, 2009c). But the proposed WPRM model tries to form well focused data of interested users using decision trees (Zidrina & Pabarskait, 2003; Zuhoor, Swamy, Muna & Haider 2005)and then fuzzy temporal frequent pattern mining algorithm is applied on this group, which inturn improves the performance. However, determining useful and interesting patterns is still an open problem. Comparing with all the works present in the literature, the work presented in this paper is different in many ways. First, it uses fuzzy logic for efficient decision making. Second, it uses temporal constraints for validating frequent relevant document. Third, it uses web data for effective classification. Finally, it provides recommendations for site re-organization through interested /popular page pattern identification.

III. PROPOSED WPRM MODEL

Fuzzy Logic is a problem-solving control system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, workstation-based data acquisition and control systems. Fuzzy logic provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, noisy, or missing input information.

Fuzzy C-Means clustering algorithm is used to create fuzzy partitions in, (Ashish & Vikram 2009) compared to this logic, the FTARM algorithm proposed in this work classifies the web user profiles dataset periodically. The temporal data stored in the database follows interval stamping of tuples where the start-time and end-time for the temporal attributes are provided as two separate attributes. Moreover, the data set used in this work follows transaction time since there is no difference between the transaction time and valid time in this web log data. Each tuples in the database is uniquely identified by a composite key in which the temporal start-time is one of the attributes. Moreover, this work provides suitable experimental analysis for the proposed fuzzy logic based temporal association rule mining approach in which relevancy is increased by enhancing semantics in addition to the relevancy measures provided by the conventional syntax based approaches.

The architecture of the proposed WPRM model in which FTARM used for intelligent classification is shown in Fig. 1. It has two phases.

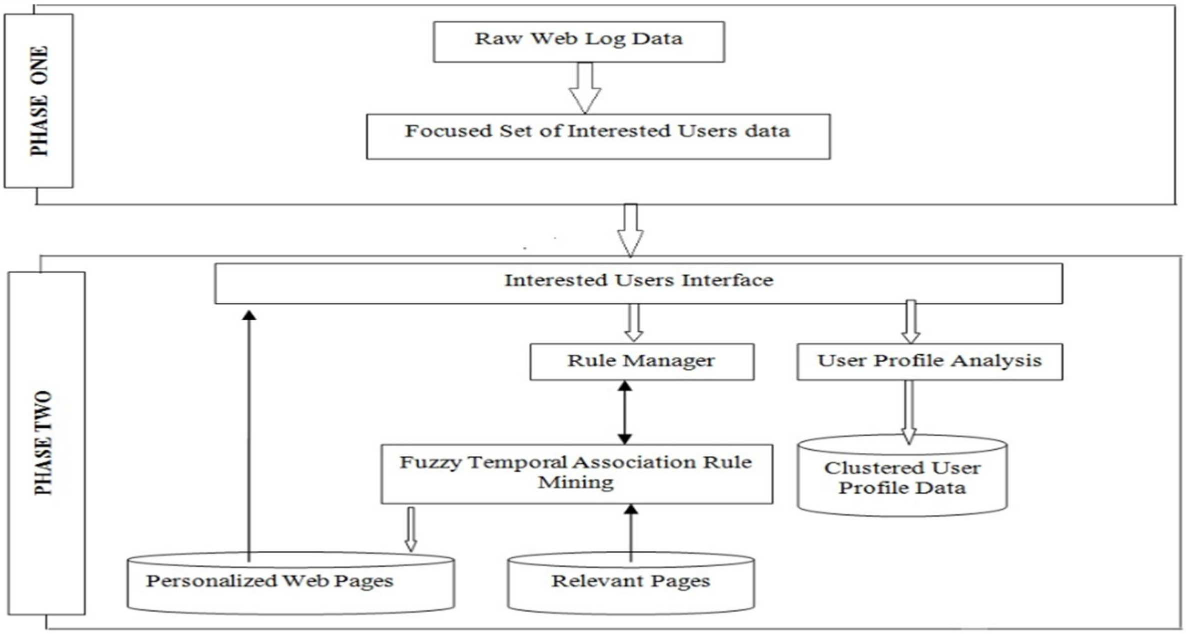


Fig. 1. System Architecture of Web Personalized Recommendation Model

Phase 1, collects raw web server log file and process the log data to avoid erroneous data. The preprocessed data is classified to identify interested users (Suneetha & Krishnamoorti, 2010a; Suneetha & Krishnamoorti, 2010b) . In Phase 2, FTARM algorithm is proposed and applied on this classified set to provide recommendations. The mined patterns are expressed in the form of fuzzy temporal association rules which satisfy the temporal requirements specified by the user. These rules are used to provide recommendations for site re-organization through interested /popular page pattern identification. In our previous work (Suneetha & Krishnamoorti, 2011d) the pages are identified as popular pages in the sequence of frequent patterns using set of attributes. Comparing with the previous work the present proposed system works faster in identification of interested/ popular pages by use of two attributes time-spent and count. The pages just used for navigational purpose may be eliminated by creating direct link between these pages which yields popular page patterns.

IV. PROPOSED FUZZY TEMPORAL ASSOCIATION RULE MINING ALGORITHM

The proposed algorithm uses a partition-approach to generate fuzzy temporal association rules. The dataset is logically divided into p disjoint horizontal partitions P_1, $P_{2,\dots\dots}$ Pn. Each partition is as large as can fit in available main memory. For ease of exposition, it is assumed that the partitions are equal- sized, though each partition could be of any arbitrary size as well.

The following notations are used in this work

- E = Focused fuzzy dataset
- Set of partitions P = $\{P_1, P_2, \dots, P_n\}$
- *tid (it)* = *tidlist* of item set *(it)*,
- t_1, t_2 Start timing and End timing of events
- μ = fuzzy membership of item set, with reference to time spent.
- time-spent *(it)* = total time spent by an item set *(it)* in the current transaction T
- count *(it)* = number of occurrences of the item set *(it)* in the current transaction T
- d= number of partitions

The byte-vector-like data structure is used in which each cell of the byte-vector stores μ of the j^{th} item set of i^{th} transaction corresponding to the cell index of the *tid* to which the μ pertains. Thus, the j^{th} cell of the byte-vector contains the μ for the j^{th} *tid* in i^{th} transaction. If a particular transaction does not contain the item set under consideration, the cell corresponding to that transaction is assigned a value of 0. When the byte-vector is initialized, each cell by default has value 0.

The fuzzy membership μ of an item set is defined in the Equ.1, by considering time spent on individual page evaluated with difference between end time and start time.

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\mu_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} t_{S_{ij}^{e}} - t_{S_{ij}^{s}} \tag{1}$$

Where μ_{ij} fuzzy membership of j^{th} singleton S in the i^{th} transaction

$t_{S_{ij}^{e}}$ is the end time of j^{th} singleton S in the i^{th} transaction

$t_{S_{ij}^{s}}$ is the start time of j^{th} singleton S in the i^{th} transaction

Time-spent of each item set is defined in the Equ. 2.

$$\sum_{i=1}^{n}\sum_{j=1}^{m} time-spent\left[S_{ij}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} time-spent(S_{ij}) + \mu_{ij} \tag{2}$$

Where,

S_{ij} is the j^{th} Singleton S in i^{th} transaction $\forall i \in T, \forall T \in D$

μ_{ij} fuzzy membership of j^{th} singleton S in the i^{th} transaction

Equ. 3 and Equ. 4 define a t-norm and the cardinality of a fuzzy set in a finite universe D (31), (32), (33) (De Cock, Cornelis, & Kerre, 2003; Yan, Chen, Cornelis & Kerre 2004; Verlinde, De, & Boute, 2006). Fuzzy sets S and $S`$ in $D$ are mapped as $D \rightarrow (0, 1)$, with $S$ ($i$) and $S`$ (i) being the degrees to which attributes S and $S`$ are present in a transaction $i$ respectively. Thus, using fuzzy partitions $S$ and $S`$ and a t-norm, we can define fuzzy support Equ. 5 and confidence Equ. 6. The more generally used t-norms are listed in Table 1, with TM t-norm being the most popular one. The same t-norm is used in the proposed algorithm.

$$S \bigcap\nolimits_T S'(i) = T(S(i), S`(i)) \tag{3}$$

$$|S| = \sum\nolimits_{(i \in D)} A(i) \tag{4}$$

$$\text{Supp(S=>S`)} = \sum\nolimits_{(i \in D)} \left(S \bigcap\nolimits_T S`\right)(i) \tag{5}$$

$$\text{Conf(S =>S`)} = \frac{\sum_{(i \in D)} \left(S \bigcap_T S`\right)(i)}{\sum_{(i \in D)} S(i)} \tag{6}$$

Table 1 t-norms in fuzzy sets

t-norm
$TM(x, y) = \min(x, y)$
$TP(x, y) = xy$
$TW(x, y) = \max(x + y - 1, 0)$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

A. Proposed Algorithm

In stage1, each transaction in the current partition of the data set is scanned and a list is created for each singleton item found *(tidlist)*. The count of each itemset *(it)* is maintained in count *(it)*, total time spent by an item set *(it)* is maintained in time-spent *(it)*, in the current transaction T and by checking the support count value of singletons which are not frequent are dropped. To generate larger itemsets, Breadth-First Search (BFS) technique is used which is similar to the one used in Apriori. At the k^{th} level, each k-item set *(itk)* is combined with another k-item set *(itk')* to generate a *(k+1)*-item set *(itk+1)*; if the two k-itemsets differ by just one singleton. The *(tidlist) td(itk+1)* for each *(k+1)*-item set *(itk+1)* is generated by intersecting the *tidlists* of its parent k-itemsets, *td(itk)* and *td (itk')*. If *(itk+1)* is not frequent, then *td (itk+1)* is discarded. Additionally, the count of each *(k+1)*-item set *(itk+1)* is maintained in count *(itk+1)*. Then the next partition is traversed in a similar manner, till all partitions have been processed.

The main advantage of computing count is that, with the generation of association rules we are able to predict users' next action. By observing count and time spent on particular page or item set in the sequence of rule, one can able to decide whether the pages are really user interested one or used just for navigational purpose. This information is beneficial for the service provider to create direct link between interested pages which results in less time consumption to reach destination pages.

The steps of the algorithm includes two stages as stage1 and stage 2 are explained below with pseudo codes.
Algorithm: FTARM Algorithm
Input: set of disjoint partitions $p_1, p_2, \ldots p_n$.
Output: Fuzzy Temporal Association Rules.

1) Stage 1.

```
for each partition P
    for each transaction T in current partition P
        for each singleton S in current Transaction T
            compute count(S) // count total number of time S appears in T
                while(count(S) < 0 )
                    calculate  μ = t_S^e − t_S^s  for each S
                    time-spent[S] +=  μ //add each μ value of S
                    count(S) -- //decrement count
                end while
          end for
                add T, count for S and corresponding time-spent[S] to tidlist td(S)
    end for
    for each singleton S
        if S is not d-frequent
            delete td(S)
        end if
    end for
    while no. of d-freq k-itemsets at each k-level ≥ 2
        for each possible pair of itemsets it_k and it'_k
         If it_k and it'_k differ exactly by 1 singleton
           combine itk and itk' to get itk+1
           td (it_{k+1}) = td (it_k) ∩ td (it'_k) /* use t-norm TM for intersection */
                compute count (it_{k+1})
                while(count(S) < 0 )
```

```
                    calculate μ for itk+1 using td (it_{k+1})
                    time-spent (it_{k+1}) += μ //add each value of (it_{k+1})
                    count (it_{k+1}) -- //decrement count
                    end while
                    if it_{k+1} is not d-frequent
                       delete td (it_{k+1})
                    end if
                end if
            end for
        end while
end for
```

An itemset is d-frequent if the ratio of total number of times *(it)* appears over the total number of partitions should be equal or exceeds the support value. In stage2, each partition is scanned for the remaining non frequent items. For each remaining item set *(it)*, identify its constituent singletons S_1, S_2, …, S_m and then obtain the *tidlist* of *tid (it)* by intersecting the *tidlists* of all the constituent singletons. Additionally, the count of each singleton *(it)* is updated in count *(it)* and time spent an item set is updated in time-spent*(it)*. The process is continued till there are no more itemsets.

2) Stage 2:// for remaining itemsets

```
iterate each partition P
     for each item set (it)  to P in 1st stage
        if (it)  is frequent (based on count (it)) over the whole dataset E
              output (it)
              remove (it)
         end if
     for each remaining item set (it)
         identify constituent singletons
        S1, S2, …, Sm of (it)  ∀ (it) = S1 ∩ S2 ∩ … ∩ Sm
         tidlist td(it)  = intersect tidlists of all constituent singletons
         /* tidlist intersection is same as in stage 1 */
         compute count(it)
         while(count(it) < 0 )
              calculate μ for (it)  using td (it)
               time-spent (it) += μ
               count(it)--//decrement count
         end while
     end for
       if no itemsets remain to be enumerated //exit
       end if
end iterate
```

V. EXPERIMENTAL RESULTS

In this section, the performance of proposed algorithm with respect to fuzzy Apriori is presented with implementation details. The main data sources used for experimental purpose are of type server access log files. In this work raw web log data is collected from the various resources. One such standard source is from NASA (Kennedy space center NASA) server over the months of July 1995 (195 MB) and August 1995 (160 MB).

The second data set is from the educational web site www.enggresources.com of month August 2010 (41MB). This web site focuses on engineering education and provides information related to engineering subjects, syllabus, courses, teaching guide lines, question banks, etc. Various experiments are conducted with the proposed framework to predict, to analyze interested user behavior and to provide recommendations. We have implemented the system components -in JAVA SDK 6.0 and simulated on NVIDIA GFORCE GT 630 + i3 processor with 4GB of physical RAM and 465GB of free disk space with Windows 7 operating system. Time taken for data cleaning is 1 minute 24 seconds, for user identification 19 seconds and session identification 35 seconds. Overall time consumption for preprocessing is 2 minutes 16 seconds. Summarized information of data preprocessing is given in Table 2.

Table 2 Summarized details of server log file

Website	Duration	Original Size	Size after Preprocessing	% Reduction in Size	No. of Sessions	No. of Users
NASA	Jul-95	195MB	37MB	81.2%	38714	26938
NASA	Aug-95	160MB	30MB	72.98%	16821	15421
enggresources.com	August-10	41.1MB	8.97MB	78.18%	3858	2633
mynews.com	25-Aug-10	395MB	57MB	72.22%	16810	12525
mynews .com	26-Aug-10	681MB	8.99MB	92%	46314	32135
Academic Site	12-28th May 2010	209MB	51MB	72.5%	1645	936

Generation of focused set of interested users using decision rules reflects in the database size reduction for further analysis. The database size is reduced by 40% after classification. The main attributes used to identify users interest are, total time a user stays at the site, total number of accessed pages, access methods used (GET, POST). The main aim is to avoid the transactions which includes set of pages just scanned through, which reduces the number of partitions as much as possible during further step. Less number of partitions means faster processing and less consumption of resources like main memory and processor.

A. FTARM Analysis

We have used transaction time attribute present in the dataset to form partitions and then generated the fuzzy version of the dataset (using a threshold for membership function μ as 0.1). Using various values of support ranging from 0.1 to 0.9, it is clearly observed that FTARM performs 10-15 times faster than fuzzy Apriori, depending on the support used. Fig. 2 shows comparison on execution time variation for FARM and FTARM. For any dataset there is a particular support value for which optimal number of itemsets is generated and for supports less than this value, we get too many itemsets which are of no practical use. From the experimental analysis it has been observed that the proposed algorithm performs most efficiently.

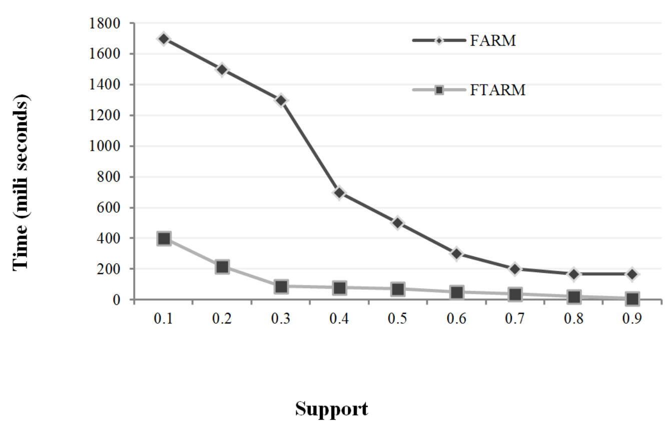


Fig. 2. Comparison on Execution Time Variation

The experiment results shows that the new FTARM approach gives interesting rules and produces less frequent itemsets from useful interested users patterns. Fig. 3 and Fig. 4 shows number of interesting rules and frequent itemsets using user specified fuzzy confidence and support.

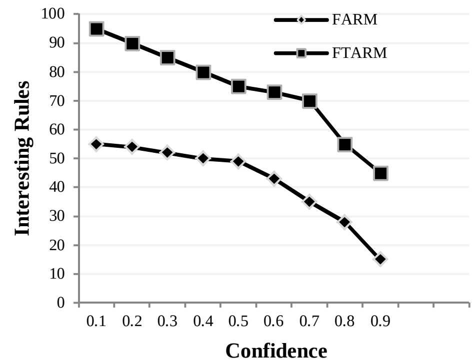


Fig. 3. Comparison of Interesting Rules

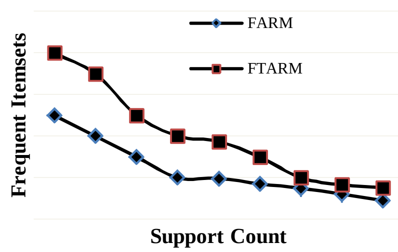


Fig. 4. Comparison on Frequent Itemsets

In this proposed work, interested pages are identified by use of time-spent and count attributes in the sequence of generated association rules obtained from set of partitions.

B. Recommendations and Analysis

Recommendations assist the web site designers to improve the performance by giving preference to the interested regular users' patterns to improve customer loyalty. Recall indicates what proportion of all the relevant documents has been retrieved from the collection. Precision indicates what proportion of the retrieved documents is relevant. Since the collection of documents is from the web, the total number of relevant documents in the collection is usually unknown. Precision is calculated from the retrieved set of documents and hence, only the precision measure is considered. Relevance-based measure of recall and precision are most widely used to test the performance of an information retrieval system. Using this, the relevant webpage's are retrieved after matching the pages with user's interest even though the user's accessing time varies. The recall and precision are defined in the following equations.

The metrics precision, coverage and F-measure is used to evaluate the recommendaton system. Precision and coverage are defned in Equ.7 and Equ. 8 as follows:

$$\text{Precision} = \frac{\text{Recommended pages}}{\text{Pages retrieved from relevant documents}} \qquad (7)$$

$$\text{Coverage} = \frac{\text{Number of relevant documents}}{\text{Total numberof documents retrieved}} \qquad (8)$$

Precsion measures the degree of accurate recommended pages. Coverage is the probability of arrival of recommended pages in the sequence of relavent documents. F-measure considers both as in Equ. 9

$$F = \frac{2 * \text{Precision} * \text{Coverage}}{\text{Precision} + \text{Coverage}} \qquad (9)$$

The results obtained by applying the FTRAM is used to get the precision graphs

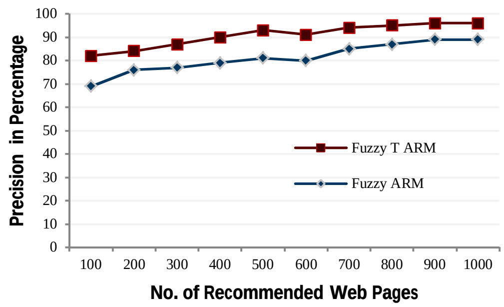


Fig. 5. Performance Analysis of Proposed Recommendation System

In Fig. 5, the proposed FTARM algorithm is compared with existing FARM algorithm. From this figure, it has been observed that the performance of the proposed algorithm is improved 12 % with respect to precision. This improvement helps to retrieve customized web pages.

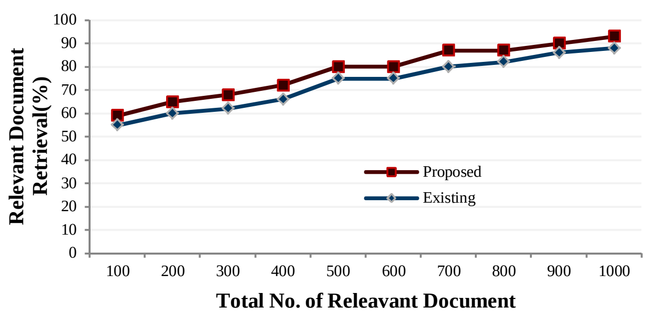


Fig. 6. Relevancy Measurement

In Fig. 6, the relevancy of the proposed algorithm is analyzed by comparing it with the existing algorithm. From this graph, it can be seen that the proposed algorithm improves the relevancy by 10% when it is compared with existing algorithm. This helps to retrieve relevant and personalized web pages to the user. Further works in this direction could be the inclusion of semantics for effective relevant information retrieved. In our previous work (Suneetha & Krishnamoorti, 2011d) the pages are identified as popular pages in the sequence of frequent patterns using set of attributes. Comparing with the previous work the present proposed system works faster in identification of interested/ popular pages by use of two attributes time-spent and count. The pages just used for navigational purpose may be eliminated by creating direct link between these pages which yields popular page patterns. Popular page patterns are condensed in size compared to frequent patterns as the navigation pages are eliminated.

Using popular page patterns, the user reaches the destination with less number of hops and also in reduced time. This imposes less burden on computation. Recommendations are drawn from this well focused set to assist service provider for restructuring web site and web personalization. This satisfies the demanding requirements of today's applications such as web personalization, site modifications and business intelligence for the success of e-commerce.

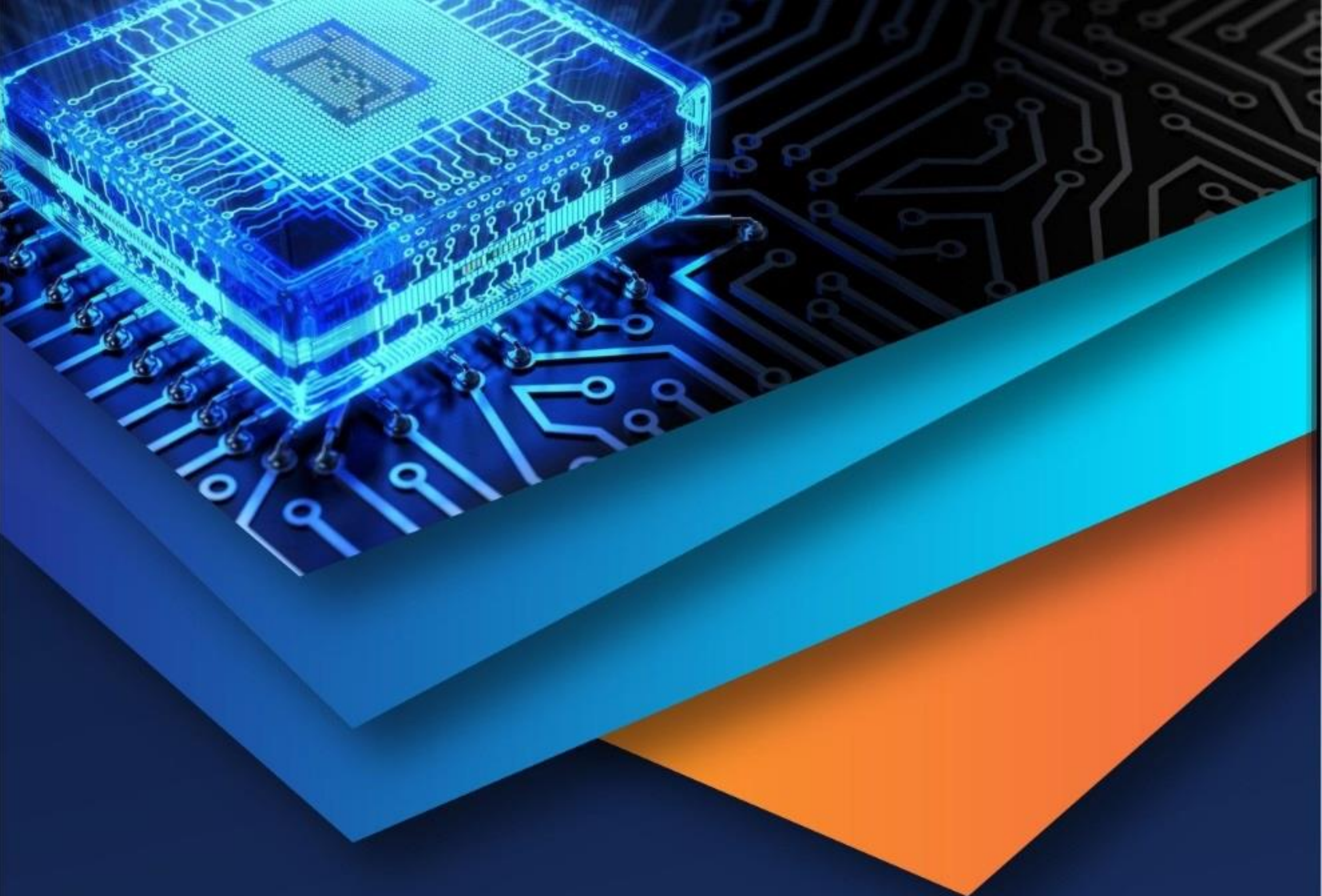
VI. CONCLUSION

A number of research work concentrated on applying data mining techniques on to preprocessed web log data to identify frequent patterns. But the proposed WPRM model tries to form well focused data of interested users in Phase 1 and then FTARM algorithm is applied on this focused group in Phase 2. The advantage here is instead of considering overall entries which include patterns of interested as well as uninterested, importance is given to a focused set of interested users / customers in order to retain regular interested customers. This in turn improves the performance by reducing the size of the database, execution time, memory usage and the algorithm FTARM helps to retrieve relevant and personalized web pages to the user. Interested pages are identified in the sequence of association rules using time-spent and count attributes. Trying to remove the pages that are used just for navigational purpose by forming direct link in the sequence of path will yields popular page pattern and also with a smaller number of hops the user will be able to reach the destination. This retains the regular customers and also attracts new customers in support for extraction of relevant pages within minimum time.

REFERENCES

[1] Agrawal, R., Imielinski, T., Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. Proceedings of ACM SIGMOD, International Conference on Management of Data, Washington, 207-216.
[2] Ashish Mangalampalli, and Vikram Pudi. (2009). Fuzzy Association Rule Mining Algorithm for Fast and Efficient Performance on Very Large Datasets. IEEE International Conference on Fuzzy Systems, Report No: IIIT/TR/2009/173.
[3] Bodon, F. (2003). A Fast Apriori implementation. Proceedings of IEEE ICDM Workshop on Frequent Itemset Mining Implementations, 90.
[4] Coenen, F., Leng, P., and Goulbourne, G. (2004). Tree Structures for Mining Association Rules. Data Mining and Knowledge Discovery, 8, 1, 25 -51.
[5] Castellano, G., Fanelli, A., M., and Torsello, M., A. (2008). Computational Intelligence techniques for Web Personalization. Web Intelligence and Agent Systems, 6, 3, 253-272.
[6] De Cock, M., Cornelis, C., and Kerre, E., E. (2003). Fuzzy Association Rules: A Two-Sided Approach, In: FIP, 385- 390.
[7] Desikan, P., and Srivastava, J., (2004). Mining Temporally Evolving Graphs. Proceedings of Workshop on Web Mining and Web Usage Analysis.
[8] Dimitrios Pierrakos, and Georgios Paliouras. (2010). Personalizing Web Directories with the Aid of Web Usage Data. IEEE Transactions on Knowledge and Data Engineering, 22, 9, 1331-1344.
[9] Gyenesei, A. (2001). A Fuzzy Approach for Mining Quantitative Association Rules. ActaCybernetical, 15, 2, 305-320.
[10] Hofgesang, P., I. (2009). Online Mining of Web Usage Data: An Over-view. Web Mining Applications in E-Commerce and E-Services, Springer, 1-24.
[11] Jaideep Srivastava, Prasanna Desikan, and Vipin Kumar. (2002). Web- Mining – Accomplishments & Future Directions. Technical Report Computer Science Department, University of Minnesota, Minneapolis, USA, 51-61.
[12] Kuok, C., Fu, A., and Wong, H. (1998). Mining Fuzzy Association Rules in Databases. ACM SIGMOD Record, 27, 1, 41-46.
[13] Mangesh Bedekar, Dr. Bharat Deshpande, Ramprasad Joshi. (2008). Web Search Personalization by User Profiling. First International Conference on Emerging Trends in Engineering and Technology.
[14] Ma, Z., Pant, G., and Sheng, O., R., L. (2007). Interest-Based Personalized Search. ACM Transaction Information Systems, 25, 1, 5.
[15] Mofreh Hogo, Miroslav Snorek, and Pawan Lingras. (2003). Temporal Web Usage Mining. Proceedings of the IEEE International Conference on Web Intelligence.
[16] Nasraoui, O., Rojas, C., and Cardona, C. (2006). A Framework for Mining Evolving Trends in Web Data Streams Using Dynamic Learning and Retrospective Validation. Computer Networks, special issue on Web dynamics, 50, 14.
[17] Olfa Nasraoui, Maha Soliman, Esin Saka, Antonio Badia, and Richard Germain. (2008). A web usage mining framework for mining evolving user profiles in dynamic web sites. IEEE Transactions on Knowledge and Data Engineering, 20, 202-215.
[18] Perkowitz M., and Etzioni, O. (1997) Adaptive Web Sites: Automatically Learning for User Access Pattern. Proceedings of Sixth International World Wide Web Conference.
[19] Robert Cooley, Mobasher B., and Srivastava. (1997). Data Preparation for Mining World Wide Web Browsing Patterns. Knowledge and Information System, 1, 1, 5-32.
[20] Spiliopoulou M., and Faulstich, L., C. (1998). WUM: A Web Utilization Miner. Proceedings of First International Workshop on Web and Databases.
[21] Srikant, R., Agrawal, R. (1996). Mining Quantitative Association Rules in Large Relational Tables. Proceedings of ACM SIGMOD Conference on Management of Data, ACM Press, Montreal, Quebec, 1-12.
[22] Srivastava, J., Cooley, R., Deshpande, M., and Tan, N.,P. (2000). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. SIGKDD Explorations, 1, 2, 1-12.
[23] Suneetha, K. R., and Krishnamoorti, R. (2010a). Classification of Web Log Data to Identify Interested Users Using Decision Trees. International Journal of Ubiquitous and Communication Journal, 5. (www.ubicc.org/files/pdf/Classn_439.pdf.)
[24] Suneetha, K. R., and Krishnamoorti, R. (2010b). Extracting Users Pattern from Web Log Data using Decision Tree and Association Rule. International Journal of Business Performance and Supply Chain Modeling, 2, 2, 125-133. (doi 10.1504/IJBPSCM.2010)

[25] Suneetha K., R., and Krishnamoorti, R. (2009c). Identifying User Behavior by Analyzing Web Server Access Log File. International Journal of Computer Science and Network Security, 9, 4, 327-332.
[26] Suneetha K., R., and Krishnamoorti. R. (2011d). IRS: Intelligent Recommendation System for Web personalization. European Journal of Scientific Research, 65, 2,175-186.
[27] Tsuyoshi Murata and Kota Saito. (2006). Extracting Users Interests from Web Log Data. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Hong Kong, China, 343-346.
[28] Verlinde, H., De Cock, M., Boute, R. (2006). Fuzzy Versus Quantitative Association Rules: A Fair Data-Driven Comparison. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 36, 679-683.
[29] Yan, P., Chen, G., Cornelis, C., De Cock, M., and Kerre, E.E. (2004). Mining Positive and Negative Fuzzy Association Rules. In: KES, Springer, 270-276.
[30] Yan, T., Jacobsen, M., Garcia-Molina, H., and Dayal, U. (1996). From User Access Patterns to Dynamic Hypertext Linking. Proceedings of Fifth International World Wide Web Conference.
[31] Ye, X., and Keane, J., A. (1997). Mining Composite Items in Association Rules. Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 1367-1372 .
[32] Yuefeng Li and Ning Zhong. (2006). Mining Ontology for Automatically Acquiring Web User Information Needs. IEEE Transactions on Knowledge and Data Engineering, 18, 4, 554-568.
[33] Zidrina, and Pabarskait, (2003). Decision trees for web log mining. Intelligent Data Analysis, 7, 2 141-154.
[34] Zuhoor A1-Khajri, Swamy Kutti, Muna Hatem and Haider ALKhajri. (2005). A Classification Technique for Web Usage Analysis. Journal of Computer Science, 413-418.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY