CHAPTER 5

WEB PERSONALIZATION AND RECOMMENDATION

When browsing the web, without proper guidance, users often wander aimlessly without visiting the web pages of their interests; then leave the web site sooner after losing their interests. The users typically do not have such a complete priori-information when they are searching or browsing the Web. Thus, the web systems need to guess what information the user wants. Different users or group of users have different interests. In order to satisfy different users, the web system should be able to distinguish between different users or groups of users and needs to be able to predict the user’s needs. Web personalization is necessary in order to solve the above problems.

![Figure 5.1 System Architecture of Web Personalization and Recommendation Module](image-url)

Figure 5.1 System Architecture of Web Personalization and Recommendation Module
The Web Personalization and Recommendation module designed and implemented in this work as shown in Figure 5.1 provides customized WebPages to the user. This Web Personalization and Recommendation module proposed in this work uses the Fuzzy temporal association rule mining (FTARM) to classify the Web user profiles dataset periodically to know the users behaviors and their interests based on temporal pattern analysis.

5.1 Fuzzy Temporal Association Rule Mining

The Fuzzy-Temporal Association Rule Mining Algorithm (FTARM) proposed in this work is to classify the Web user profiles dataset periodically and to know the users behaviors as well as their interests based on temporal pattern analysis.

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 in this web log data which is used in this work, where is no difference between the transaction time and valid time. Each tuple in the database is uniquely identified by a composite key in which the temporal start-time is one of the attributes. In the fuzzy logic, we convert the qualitative information into quantative information using a test score semantics and fuzzy rules.

In this work, FTARM is used for reducing the search space of the Web user profiles dataset in where Fuzzy logic is used for intelligent classification. 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 relevancy of a web page is computed using the term frequencies initially. 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. Now the pages shown to the user to get his relevance feedback. After the user is satisfied with the pages, ontology is created and the semantic makes along with ontology are used to perform the semantic analysis.

In this algorithm, the temporal constraints are used because the different user groups accessing the internet are in different time periods. Therefore, the user temporal data is stored and are also analyzed classified and the relevant rules are extracted. Using this, the relevant WebPages are retrieved after matching the pages with user’s interest even though the user’s accessing time varies.

5.1.1 Proposed Algorithm

The steps of this proposed algorithm are as follows.

Step1: Read the data records

Step2: Compute fuzzy temporal memberships.

Step3: Compute support and confidence based on fuzzy temporal attributes.

Step4: Perform user analysis using these rules.
Step5: Find the relevant pages.

Step6: Perform recommendation using these rules and the user profile analysis result.

5.1.2 Proposed Fuzzy Temporal Association Rule Mining Algorithm

Algorithm: - The Proposed algorithm uses a partition-approach to generate FTA rules.

Input: - The dataset is logically divided into p disjoint horizontal partitions P1, P2, …, Pp. 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.

We use the following notations are used in this work

- E = Fuzzy dataset generated after pre-processing
- Set of partitions P = {P1, P2, …, Pp}
- td[it] = tid list of item set it,
- t1, t2 Start timing and end timing of events
- μ = fuzzy membership of any item set• count [it] = cumulative μ of item set it over all partitions in which it has been processed• d = number of partitions (for any particular item set.

Output: Fuzzy Temporal Association Rules.

The Proposed FTA Rule mining algorithm has good Performance.
The steps of the algorithm are explained below using pseudo codes.
5.1.3 Pseudo Code for FTA Rule Mining

1. Traverse each partition ‘P’
2. Traverse each transaction ‘t’ in current partition ‘P’
3. For each singleton attribute s in current transaction ‘t’
4. Calculate ‘µ’ for each ‘s’ in the interval $[t_1, t_2]$ 
5. Count[s] = count[s] + µ;
6. Add t and the corresponding ‘µ’ for ‘s’ to tid list td[s]
7. For each singleton ‘s’
8. If (s=d-frequent)
   
   {
   
   While (no. of d-freq k-item sets at each k-level $\geq 2$)
   
   Check the time interval for the item-sets
   
   If temporal constraint is satisfied add s to frequent item set
   
   Else Exit.
   
   }

   Else (s$\not\in$ d-frequent) delete td[s];

9. For each possible pair of item sets itk and itk’ then
   
   If itk and itk’ differ exactly by 1 singleton then
   
   Combine “itk” and itk’ to get itk+1 and compute td
   
   td[itk+1] = td[itk] / td[itk’]

10. Increment the counter using the relation
   
   Count [itk+1] = count [itk+1] + µ

11. If ([itk+1] = d-frequent) then
For each remaining item set ‘it’

Identify constituent singletons $s_1$, $s_2$, …, $s_m$ of it $\forall$ it = $s_1$, $s_2$ …, $s_n$

12. Set tid list $td[it]$ = intersect tid lists of $s_i$

13. Calculate the total time ‘T’ for the entire process time using the relation

$$T = (t_2-t_1)+(t_4-t_3)+…+(t_n-t_{n-1})$$

14. Calculate ‘$\mu$’ for it using $td[it]$

15. Set Count[it] += $\mu$

16. If no item sets remain to be enumerated then exit

17. Else

18. Go to step 11.

5.1.4 Results and Discussion

The results obtained by applying the FTRA are used to get the precision graphs. Validation has been performed using precision and recall analysis.

![Figure 5.2 Performance Analysis of Proposed Recommendation System](image)
In figure 5.2, the proposed Fuzzy T ARM algorithm is compared with existing Fuzzy ARM 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. The data set consist of search information about journals and books. The data set contains 1500 records with 8 features.

![Figure 5.3 Relevancy Measurement](image)

In Figure 5.3, 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.

This work uses a fuzzy T ARM which is not used by any other researchers found in the literature. The existing ARM base page ranking has been compared with FTARM.