CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENTS

6.1 CONCLUSION

In this study, a new framework namely, Integration of Clustering and Rule Induction Mining (ICRIM) Framework has been implemented to evaluate the performance of web usage knowledge discovery system. The proposed Miner framework is a primary attempt to patch up some of the drawbacks of the conventional web log file analyzers.

The web usage mining framework, ICRIM, initially performed clustering on web usage data for mining of web user’s pattern. Web data clusters have been determined and the web site visitor activities have analyzed and optimally segregated related user interests. Web usage data clustering is performed with expectation maximization (EM) and C-Fuzzy means cluster algorithms. Fuzzy C means clustering algorithm has determined web data clusters whereas Expected Maximization cluster system analyzed the Web site visitor trends in the clustered data. An EM algorithm is employed in statistics for identifying maximum likelihood estimates of parameters in probabilistic methods, where the method depends on unseen hidden variables.

After that ICRIM Framework has been built on Induction based decision rule model based on the web data clusters which is formed by the EM algorithm. Induction based decision rule model induced the knowledge
that extracts the decisive rules in web usage mining framework. By relating the Web logs with cookies and forms, it is more feasible to examine the visitor activities and profiles which facilitate an e-commerce site to deal several business questions. Based on this web traffic data the induction process is done. The number of decisive rules is extracted by inducing only rules that are related to customer’s requirement. Relevancy is focused by query predicates.

By integrating clustering model and Induction based decision rule model ICRIM framework has been implemented. ICRIM Framework efficiently discovered Web data clusters and analyzed web site visitor behavior trends. Whatever the number of clients has, ICRIM Framework has designed to provide the result in real time and generated inferences and implicit hidden behavioral aspects in the web usage mining to investigate at the web server and client logs.

Experimentation is carried out on ICRIM framework to evaluate the performance of web usage knowledge discovery system. ICRIM Framework is implemented by weka tool and experimentation is conducted with Weblog dataset from UCI Repository. The experimental results can be used for predicting user’s next request in the massive web sites. Results of EM represent that by decreasing the number of clusters, the log likelihood meets near lesser values and probability of the largest cluster will be decreased while the number of the clusters increases in each web usage pattern.

The performance results of ICRIM framework compared with existing methods, has given 5% to 8% better performance in cluster accuracy, 10% to 13% more in rule extraction, 15 to 26 % better result in Mean Absolute Error and 18 to 27 % better result in Root Mean Square Error.
6.2 FUTURE ENHANCEMENTS

The future works related to ICRIM Framework, can be summarized as follows,

In ICRIM Framework, preprocessing can be further extended by using site maps for page view and rule discovery. The site map could be considered with its various versions by using a versioning tool. Because, the site map can be used for recognizing more precisely the page views by determining which Web resources (for example, images) are not contained in the Web pages, but might be displayed explicitly based on the user request.

Further research direction focuses on the characteristics of how user visits Web page and usage pattern. Usage pattern could be abstracted from these classes using sequence mining that forecasts the coming behavior of users. Hence, Cache prefetching is to be added to ICRIM Framework which can improve the hit ratio and accelerate users visiting speed.

Future work on ICRIM framework can also inspect techniques for sampling data after sessionizing to decrease the overhead of the rule generation process. In addition, by classifying users to predefined clusters based on their present behavior for Web page prediction will be the next research direction.

To extend ICRIM framework, different clustering algorithms can also be examined in order to discover an optimal one. Furthermore, rules in different clusters can be studied that will be interesting to see if they can be used to recognize different user groups defined by user applications by developing some techniques. This type of information could be valuable for Web owners, particularly for e-commerce Website, to intend different Web pages to target different user groups.
The dynamic domains, for example practically any data generated on the Web, may need frequent costly updates of the clusters and, at any time new data records are updated to the dataset. The new upcoming data may be due to new user action on a website (e.g., click streams) or a search engine (queries), in the case of clustering, etc. Moreover, data records could result in a change of clustering in time. As a result, clusters may require to be updated, thus leading to the essential to mine dynamic clusters. Hence ICRIM framework requires dynamic clustering mechanism.

Future work on ICRIM framework in this area involves performing experiments with various types of transactions derived from user sessions, for instance, to segregate specific types of content pages in the recommendation process and also plan on integrating client-side agents to offer an additional level of personalization based on user preferences.