ABSTRACT

With the increasing popularity of World Wide Web (WWW), huge size of data such as user’s addresses or URLs which have been demanded are repeatedly collected by Web servers and collected in access log files. Investigating a server access data can offer considerable and valuable information for performance improvement, reorganizing a website for improving efficiency, and user intending in electronic commerce. Determining for relationships and universal patterns that survive in huge access log files is generally named as Web usage mining.

Web usage mining tries to find out helpful knowledge from the secondary data attained from the communications of the users with the web. Web usage mining has turned into very important for efficient website management, constructing adaptive websites, commerce and support services, personalization; network traffic stream analysis etc., Website under research is branch of a non-profit association that does not trade any products. It was critical to recognize who the users were, what they glanced at, and how their interests varied with time. To attain this, one of the hopeful techniques is web usage mining, which extracts web logs for user models and suggestions.

This usage data provides the paths leading to retrieve web pages. This information is often collected automatically into access logs via the Web server. In this area, the concerned techniques mostly focus on the customer behavioral patterns discovery from a web server log file in order to mine relationships within collected data. Web usage mining methods have been
broadly utilized for modeling user web navigation performance. This research develops a model for extracting of user’s navigation pattern.

The proposal of the work proceeds in the direction of building a robust web usage knowledge discovery system, which discovers the web user profiles at the web server, application server and interior application level. For this purpose, a novel framework, Integration of Clustering and Rule Induction Mining (ICRIM) is presented to evaluate the performance of web usage knowledge discovery system.

Initially, ICRIM framework optimizes the usage mining framework with fuzzy C means clustering algorithm (to discover web data clusters) and compare with Expected Maximization (EM) cluster system to analyze the website visitor trends. The evolutionary clustering algorithm, EM algorithm is proposed to optimally segregate similar user interests. The clustered data is then employed to examine the trends using EM model. By connecting the web logs with cookies and forms, it is further probable to examine the visitor behavior and profiles which could assist an e-commerce site to address several business questions.

Afterward ICRIM framework build the induction based decision rule model which is based on the web data clusters that are formed by the EM algorithm. Decision rule model is used for generating inferences and implicit hidden behavioral aspects in the web usage mining which investigates at the web server and client logs. The decision based rule induction mining combines a fast decision rule induction algorithm and a method for converting
a decision tree to a simplified rule set. Finally, ICRIM framework evaluates the performance of web usage knowledge discovery system.

Proposed ICRIM framework is implemented by using weka tool to evaluate the performance of web usage knowledge discovery system. Experimentations are conducted with CFuzzy means and Expected Maximization (EM) clusters on web log dataset from UCI Repository.

Performance of ICRIM framework is evaluated by number of performance metrics such as Log likelihoods, True and false positive rate, Cluster precision, Decisive rules, Execution time, Root mean square error and Mean absolute error. Performance results shows that ICRIM framework shows better performance compared with conventional methods.