VI. CONCLUSION AND FUTURE WORK

This research is carried out to build a classification system to classify web site users into their analogous groups since it finds its relevance in many applications like personalization, recommendation, site improvement and in many other web related tasks. The dissertation mainly focused on discovery and classification of similar web navigation patterns in terms of task-oriented web user profiles and web user groups from web log to support classification. Analysis of user navigation patterns is a useful research area due to its applicability in various domains and in the present work it was carried out to know the users interests implicitly by creating sessions and these user sessions are matched with a new user’s sessions.

The research work is carried with three important techniques namely feature selection, clustering and classification. The log is preprocessed initially and as a final step in preprocessing it is characterized into a matrix model in which each row reflects the navigation pattern of web users. The work focused on the selection of relevant features from user session matrix and navigation pattern analysis using an ensemble of relevance vector machine classifiers with Bolzano–Weierstrass based clustering technique.

Preprocessing is carried out to remove unwanted entries and user’s browsing patterns were reconstructed into session matrix which consists of navigation patterns with pages visited and frequency of visit. As a final preprocessing step, feature selection is carried out. The proposed feature selection method selected relevant and less number of features. A Quantum \( \alpha \)-Skew Divergence is integrated with in the Independent Component Analysis to promote weight matrix generation. This approach gives more comprehensive results than the existing Independent Component Analysis and Principal
Component Analysis due to the skew divergence factor. This step also enhanced the clustering and accuracy of classification in the next steps and it was proven in the second and third technique.

The selected features were grouped into clusters by Bolzano-Weierstrass Fuzzy C Means method. The clusters are validated to check the compactness of patterns in inter and intra clustering. This novel clustering approach results in more accurate clustering results when compared with the existing clustering approaches such as FCM and PFCM in terms of measures such as Rand Index, precision, recall. And also the method finds its significance in the reduction of time complexity.

The resultant clusters were passed to the ensemble classification approach through vector machine learning algorithm. RVM algorithm has been used in this approach for classification. In order to improve the performance of the Relevance Vector Machine, a hybrid methodology of incorporating Fuzzy concepts to optimize is carried out in this work. Weight values were optimized for more accurate results and the methodology resulted in significant classification accuracy. But, among the three datasets considered, the system showed a variation in the e-commerce data set.

Due to the instability in output, an extension of Fuzzy RVM classification approach is carried out. In this work Particle Swarm Optimization with conditional entropy is amalgamated with Fuzzy RVM. The key parameter of RVM is the kernel which is optimized through Particle Swarm Optimization (PSO).

The volume of three data sets are different and as a forecast the system will be applied to any diversified data and scalable sizes, ensemble classification is found to be an useful approach. It was carried out by divide and conquer rule and the optimized
classifiers are applied to each partitioned training data sets and the results are evaluated and enhanced iteratively. Adaboost algorithm was used for combination due to its significance in other applications. The final results of the three data sets were measured in terms of accuracy, sensitivity, specificity. The unstable results of Fuzzy RVM were overcome in the Ensemble Relevance Vector Machine-Entropy PSO. The execution time for the approach is also lesser in comparison with previous method. This is mainly due to the influence of the PSO algorithm in optimizing the kernel parameters. The results obtained from all the methods were compared. EFMWRVM approach outperformed the other existing classification approaches such as SVM, RVM whereas ERVM-EPSO has showed its superiority in its performance.

The combination of the EFMWRVM-BWFCM-QSDICA approach outperformed the existing approaches whereas the second proposed classification model on the whole ERVM-EPSO-BWFCM-QSDICA gives enhanced results than the first proposed classification model.

The objectives formulated in this research were obtained and the findings are listed as follows

- Minimum numbers of features were selected with lesser time than the existing methods to enhance the clustering and classification process.
- A novel clustering methodology was developed to group patterns into similar groups with more accuracy.
A better ensemble multi type classification models was developed to classify a new user from the groups of navigation patterns which is more accurate and has less execution time.

The classification model is designed and it works on heterogeneous websites logs without the knowledge of the site map.

**Scope and Future Enhancements**

- The model can be extended for prediction, recommender and personalization systems for the resourceful browsing of users.
- The logs from multiple web sites can be merged and the proposed system is implemented and also from multiple servers in cloud environment.
- Semantic knowledge can be incorporated in the process for better understanding.
- With the emerging of new applications over the Internet, many new types of Web data, such as email traffic, web-blog, and wiki pages are available. These log data have produced a large amount of new knowledge resources, where the system can be remodeled a little for its use.