CHAPTER 6

CONCLUSIONS AND SCOPE FOR FURTHER STUDY

6.1. Conclusions

This research work aims to propose a mechanism with improved preprocessing technique for data clustering. A robust ensemble mechanism was proposed in this research work, which deal with the Enhanced weighted version of Principal Component Analysis (EPCA). EPCA uses a Pearson correlation coefficient as objective function to find the missing values that gives higher weights, which are considered to be more significant. Then the objective function for the co-clustering ensemble towards the application to data clustering was presented. In order to measure the statistical information between two co-clustering, mutual information is considered as a symmetric measure which is used in this work. The final ensemble step was formulated as a partition problem on a bipartite graph. Simulation results proved that the proposed Robust Ensemble Co-Clustering Algorithm (RECCA) mechanism performs better in terms of accuracy and computation time. In the first stage of the work co-clustering ensembles algorithm required more time to complete the task. To reduce the computation time and to increase the clustering accuracy this research work intended to propose a system with Improved Ant Colony Optimization (IACO) for data clustering.

Improved Ant Colony Optimization (IACO) based RECCA (IACO-RECCA) was proposed for data clustering. The overall search space is divided into two parts; "Class Hierarchy sub-graph" and "Antecedent Construction sub-graph". In the same way, an ant is equipped with two types of memories. First type of memory named "class memory" is used to save the classes selected during the tour of the ant in the class hierarchy sub-graph. The second type of memory named "antecedent memory" saves antecedent

part of the rule during the ant tour in the antecedent construction subgraph. At that point the target function for the co-clustering troupe towards application to data clustering is presented. An optimization technique for spectral co-clustering ensemble algorithm is described with constructive mathematical modeling. The proposed IACO-RECCA algorithm is capable enough to perform co-clustering with the objective function as the primary component.

At the final stage of the work propose a system with Improved Cuckoo Search based RECCA (ICS - RECCA) for data clustering. The CSA has been applied for spectral co-clustering ensemble with constructive mathematical modeling. The major objective of the ICS is to speed up convergence rate and reduces computational cost significantly. The modification involves the additional step of information exchange between the top eggs. Proposed ICS was performed based on the behavior of quantum-inspired cuckoo search through convincing results. The proposed ICS-RECCA algorithm was capable enough to perform co-clustering with the objective function as the primary component. Simulation results proved that the proposed RECCA, IACO-RECCA and ICS-RECCA perform better in terms of accuracy and computation time.

6.2. Scope for Further Study

Regarding the future direction of this work, RECCA can be hybrid with optimization techniques for the much better performance of accuracy and computation time.

Some issues could be included in the future work. In this research work majorly discusses about the certain type of smoothing by random walk, while the proposed method could be extended by using other types of smoothing, e.g., diffusion kernels, where scalable optimization could also be developed by using a similar iterative subroutine. Moreover, the smoothing of

noisy datas brings improved clustering accuracy but at the cost of increased running time. Algorithms that are more efficient in both time and space should be further investigated. In addition, the approximated matrix could also be learnable. In current experiments only used constant K-NN graphs as input for fair comparison, which could be replaced by more comprehensive graph construction methods.

In this new approach, the data clustering process can be carried out by incorporating both previous domain knowledge of biomedical datasets and text datasets in the form of pair-wise constraints and category knowledge of features into the ICS-RECCA co-clustering framework. Under this ICS - RECCA co-clustering framework, the clustering problem can be formulated as the problem of finding the local minimizer of objective function, taking into description the dual prior information. In this ICS-RECCA co-clustering framework, rules are derived and updated with an iterative optimization algorithm can be designed for the co-clustering ensemble process.

As further work, new semi-supervised co-clustering algorithm can be performed based on the orthogonal property. Orthogonal property that depends on a strategy reversal, instead of performing Non-negativity of the iterates at every step and striving to attain orthogonality at the limit, NMF enforces orthogonality of its iterates while obtaining Non-negativity at the limit. This increases the clustering efficiency.