AN IMPROVED CUCKOO SEARCH BASED ROBUST ENSEMBLE CO-CLUSTERING ALGORITHM (ICS - RECCA) FOR GENE DATA CLUSTERING

Ms.R.Rajeswari
PhD Research scholar,
Department of Computer Applications
St. Peter's University,Chennai, India.
Email: rajeswari1520@gmail.com

Dr. G.GunaSekaran
Principal,
Meenakshi College of Engineering,
Chennai, India.
Email: gunaguru@yahoo.com

***Abstract* - This research work intends to propose a system with improved cuckoo searchbased robust ensemble co-clustering algorithm (ICS - RECCA) for enzyme clustering. The cuckoo search algorithm has been inspired by the obligate brood parasitism of some cuckoo species by laying their eggs in the nests of other host birds (of other species). Some host birds can engage direct conflict with the intruding cuckoos. Based on the improved cuckoo search optimization technique it has been applied for spectral co-clustering ensemble with constructive mathematical modeling. The proposed algorithm (ICS-RECCA) is capable enough to perform co-clustering with the objective function as the primary component. Simulation results proved that the proposed mechanism ICS-RECCA performs better in terms of accuracy and computation time.**

Keywords - Bioinformatics, DNA, Clustering, Coclustering, SS-NMF, NMF, CMRF, SS-CMRF, SRC, TSVM, RECCA, Accuracy, Computation Time, Text, Gene Expression, Image High Order Coclustering, High Order Coclustering, Supervised, Semisupervised, Optimization, Cuckoo Search

I. INTRODUCTION

Nature propelled reckoning procedures are those processing systems that definitely are gotten from the investigation of particular characteristic framework, which mostly is quite significant. Competitor answers for the enhancement issue definitely assume the part of people in a populace, and the wellness capacity decides the nature of the arrangements in a sort of big way. The usually utilized nature propelled algorithms literally are Genetic Algorithm, Firefly, Artificial Bee Colony, Ant Colony Optimization, Cuckoo Search, Particle Swarm Optimization, Bat, Cat, Dolphins, Elephant and some more, definitely contrary to popular belief.

The clustering problem comprises in requesting an arrangement of information into groups, in view of the components of the data samples. Cluster analysis is an unsupervised learning technique that is utilized for the investigation of interrelationships among an organizing so as to collection of samples, by organizing them into homogenous groups. In the most recent decades, ant colonies (and other social creepy crawlies) have propelled clustering algorithms that mimic the ants' capacities for isolating and clustering larvae and dead bodies. While numerous methodologies utilize the essential idea of picking and dropping data vectors [Shutin and Kubin.,2004], others depend on distinctive properties of social creepy crawlies, for example, concoction correspondence between the ants or building mechanical structures without anyone else's input amassing conduct [Molisch.,2005].

The remarkable contributions of this paper are:

- The knowledge of whether or not adding information from external sources to the database is able to improve the clustering quality for this application;
- The lateral way for the collected information to be transformed into constraint sets for the meticulous biological problem;
- To perform optimization using Improved Cuckoo Search (ICS) for Robust Ensemble Co-Clustering Algorithm (ICS-RECCA) For Enzyme Clustering

- To perform co-clustering in order to improve the performance by reducing the computation time and increasing average accuracy value.

II. BACKGROUND STUDY

Carlos Cobos et al.,2014 proposed a new description-centric algorithm for the clustering of web results, called WDC-CSK, which is based on the cuckoo search meta-heuristic algorithm, k-means algorithm, Balanced Bayesian Information Criterion, split and merge methods on clusters, and frequent phrases approach for cluster labeling. The objective of Ahmed Elkeran.,2013 was to minimize the length of the sheet while having all polygons inside the sheet without overlap. Their methodology hybridizes cuckoo search and guided local search optimization techniques is proposed.Elyasigomari et al.,2015 developed a new hybrid optimization algorithm, COA-GA synergizing recently invented Cuckoo Optimization Algorithm (COA) with a more traditional genetic algorithm (GA) for data clustering to select the most dominant genes using shuffling. For gene classification, Support Vector Machine (SVM) and Multilayer Perceptron (MLP) artificial neural networks are used. Mohapatra et al.,2015 proposed a improved cuckoo search based extreme learning machine (ICSELM) classify binary medical datasets. Extreme learning machine (ELM) is widely used as a learning algorithm for training single layer feed forward neural networks (SLFN) in the field of classification. Ameryan et al.,2014 presented a four novel clustering methods based on a recent powerful evolutionary algorithm called Cuckoo Optimization Algorithm (COA) inspired by nesting behavior and immigration of cuckoo birds. To take advantage of COA in clustering, here, an individual cuckoo represents a candidate solution consisting of clusters' centroids.

In RuiTang et al.,2012, the constructs of the integration of bio-inspired optimization methods into K-means clustering were presented. The extended versions of clustering algorithms integrated with bio-inspired optimization methods produce improved results. Nature inspired, unsupervised classification method, based on the most recent metaheuristic algorithm, stirred by the breeding strategy of the parasitic bird, the cuckoo, was introduced in Goel et al.,2011. Emary et al.,2014presented an approach to automatic vessel segmentation in retinal images that utilisespossibilistic fuzzy c-means (PFCM) clustering to overcome the problems of the conventional fuzzy c-means objective function. In order to obtain optimised clustering results using PFCM, a cuckoo search method was used. Chifu et al.,2014 presented a method for clustering food offers based on the cuckoo search algorithm. InKarthikeyan and Venkatalakshmi.,2012suggested PSO incorporated cuckoo search optimization algorithm for clustering in energy aware way and compared it with cuckoo search algorithm. Azizipanah-Abarghooee et al.,2014proposed a multi-objective optimal static and dynamic scheduling of thermoelectric power systems considering the conflicting environmental and economical objectives. A novel multi-objective θ-improved cuckoo optimisation algorithm was projected to solve the optimisation problems by defining a set of nondominated points as the solutions.

III. CUCKOO SEARCH ALGORITHM

The Cuckoo Search algorithm was developed by imitating both the obligate brood parasitic behavior of the cuckoo species and the Lévy flight of certain birds and fruit flies. To describe the new Cuckoo Search algorithm, three rules must be followed: the cuckoo must lay one egg at a time and randomly dump it in a nest, the nest with the high-quality eggs represents the solution that carries over to the next generation, and the number of available host nest is fixed [Yang and Deb.,2009].The search is performed in a probability-based manner in which the host bird can either throw the cuckoo egg away or abandon it and relocate to a new nest, which can be assumed as a new random solution at a new location in a cuckoo search. The Cuckoo Search algorithm can generate new solutions through Lévy flights, where the random steps in the random walk are drawn from a Lévy distribution for large steps. Through the cuckoo search, the fraction (represented by p_a) of the worst nest is abandoned, and a new nest at a newlocations is constructed through Lévy flights [Yang and Deb.,2010].

When the Cuckoo Search algorithm is implemented through Lévy flights, it can be assumed that the Cuckoo Search algorithm adopts the flight behavior of the animal or insect that uses generalized random walks through a probability distribution and stochastic processes. The Lévy flights behavior can also be applied to optimization and optimal searches [Yang and Deb.,2009].In the Cuckoo Search algorithm, the eggs in the nest represent new solutions. Because each egg represents a new solution, new and potentially better solutions replace the solutions in the nest that are not as good as those obtained using the cuckoo search algorithm. A set of solutions are generated for more complicated cases [Yang and Deb.,2009]. Furthermore, Cuckoo Search algorithm solutions may be generated through a Lévy walk for a more rapid local search. New solutions for the Cuckoo Search algorithms are generated through field randomization because the Cuckoo Search algorithm is a population-based algorithm that can be extended to a meta-population algorithm. Thus, the CS can be summarized through the pseudo code.

Algorithm 1: Cuckoo Search (CS)

```
Begin
Objective function f(x), $x = (x_1, \dots, x_d)^T$;
 Generate initial population of n host nests $x_i$ (i = 1, 2,...,n);
While (t <MaxGeneration) or (stop criterion)
Get a cuckoo randomly by L´evy flights;
Evaluate its quality/fitness $f_i$;
Choose a nest among n (say, j) randomly;
If ($f_i$>$f_j$)
Replace j by the new solution;
End
A fraction ( $p_a$ ) of worse nests;
Are abandoned and new ones are built;
Keep the best solutions;
(Or nests with quality solutions);
Rank the solutions and find the current best;
End while
Post-process results and visualization;
End
```

For new solution generation, $x^{(t+1)}$indicates the Cuckoo i, and the L'evy flight is generated as follows:

$$x_i^{(t+1)} = x_i^{t} + a \oplus L'evy(\lambda) \tag{1}$$

where $\alpha > 0$: the step size in accordance with the investigated problem scales. Moreover, $\alpha = O$ is mostly used for $\oplus$to represent the entry-wise multiplication. The random steps are generated through large steps with L'evy distribution using the following relation:

$$L'evy \sim u = t^{-\lambda} \ (1 < \lambda \leq 3) \tag{2}$$

This equation has infinite variance and an infinite mean. Notably, in real terms, if the host bird's eggs are the same as the Cuckoo eggs, the latter eggs are less likely to be seen; thus, the fitness is related to the solution variations.

Many studies have used the same implementation and validation methods for the Cuckoo Search algorithm. Certain parameters are used when studying Cuckoo Search algorithms. It is easier to implement the Cuckoo Search algorithm using benchmark testing functions [Yang and Deb.,2010]; after the algorithm has been implemented, it must be validated using benchmark testing functions with known analytical solutions [Yang and Deb.,2009].An example of the many test functions that validate the algorithm functions is the bivariate Michalewicz's function.Other test functions designed to test the Cuckoo Search algorithm optimization performance though with slight variations from different researchers include De Jong's first function, Rosenbrock's function, Schwefel's function (multimodal), Rastrigin's function, Easom's function and Griewangk's function (local minima). However, [Yang and Deb.,2009] enumerated De Jong's first function, Easom's function (unimodal), Shubert's bivariate function, Ackley's function (multimodal), Schwefel's function (multimodal), Rastrigin's function, and Michalewicz's function (local optima) as benchmark testing functions for Cuckoo Search (CS) algorithm. The literatures on stochastic test functions include deterministic functions, which are typically more difficult for algorithms. Many deterministic functions can be turned into stochastic test functions [Yang and Deb.,2010].

However, other studies have used the benchmark testing functions to evaluate the optimization performance of the Cuckoo Search and other algorithms, such as the Genetic Algorithm (GA) and the Particle

Swarm Optimization (PSO), and found that the Cuckoo Search was more efficient at finding the global optima and presented higher success rates. Secondly, among stochastic functions, the Cuckoo Search algorithms outperformed both the Genetic Algorithm (GA) and particle swarm optimization (PSO) [Yang and Deb.,2010, Yang and Deb.,2009].These results were attributed to randomization, intensification and using the lowest number of control parameters as a good balance strategy for intensive local searching. Having efficient exploration throughout the entire search space through metaheuristics, such as the Cuckoo Search algorithms, could thus generate an efficient algorithm [Yang and Deb.,2009].

Yang and Deb.,2009 noted certain similarities and significant differences between Cuckoo Search and hill-climbing with some large scale randomization. The first similarity is that Cuckoo Search (CS) is a population-based algorithm that is similar to the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), but it differs because it uses elitism and/or selection, which is more similar to harmony searches. Second, the Cuckoo Search algorithm's randomization is more efficient because the step length is heavy-tailed, so any large step is possible. Third, the number of tuneable parameters is lower than for the GA and PSO; thus, it may be more generic and adaptable to a wider class of optimization problems.

Algorithm 2: Improved Cuckoo Search

```
Step 01:  Begin
Step 02:  Sorting
          According to value-to-weight ratio $p_i/w_i (i = 1,2,3,\dots,n)$ in descending order, a queue $\{s_1, s_2, \dots, s_n\}$ of length $n$ is formed.
Step 03:  Initialization.
          Set the generation counter $G = 1$; Set probability of mutation $p_m = 0.15$.
          Generate $P$ cuckoo nests randomly $\{\langle X_1, Y_1\rangle, \langle X_2, Y_2\rangle, \dots, \langle X_p, Y_p\rangle\}$. Divide the whole population into $M$ memeplexes, and each memeplex contains $N$ $(i.e. P/M)$ cuckoos; Calculate the fitness for each individual, $f(Y_i), 1 \le i \le P$, determine the global optimal individual $\langle {X_g}^{best}, {Y_g}^{best}\rangle$ and the best individual of each memeplex $\langle {X_k}^{best}, {Y_k}^{best}\rangle, 1 \le k \le M$.
Step 04:  While the stopping criterion is not satisfied
          do
          For $i$ = 1 to $P$
          $k = i$ mod$M$
          select uniform randomly $p_1 \neq i$
          For $j = 1$ to D
          $X_1(j) = X_1(j) + \alpha \oplus Levy\ (\lambda)$
          If $r_1 \ge 0.5$ then
          $Temp = B_g(j) + r_2 \times (B_k(j) - X_{p1}(j))$
          Else
          $Temp = B_g(j) - r_2 \times (B_k(j) - X_{p1}(j))$
          End if
          End for
          If $f(Y_{temp}) > f(Y_i)$ then
          $X_i = Temp$
          Else if $r_3 \le FS$ then
          $X_i = L + r_4 \times (U - L)$
          End if
          End if   where $r_1, r_2, r_3, r_4 \sim U(0,1)$
          Repair the illegal individuals and optimize the legal individuals by performing GTM method
          End for
          Keep best solutions.
          Rank the solutions in descending order and find the current best $(Y^{best}, f(Y^{best}))$.
          $G = G + 1$
Step 05:  Shuffle all the memeplexes
Step 06:  End while
```

Step 07: **End**

The final ensemble step can be formulated as a partition problem on a bipartite graph. For convenience of discussion, we use small-bold letters such as u, v as vectors. Capital-bold letters such as **M, E, L** will denote matrices, and capital letters such as V, R will denote vertex sets. Denote the bipartite graph $G = (V_r, V_c, \boldsymbol{E})$ containing two sets of vertices including row labeling vertices V_r and column labeling vertices V_c respectively. It is easy to verify that the adjacency matrix M of the bipartite graph can be written as

$$M = \begin{bmatrix} O & E \\ E^T & O \end{bmatrix} \tag{3}$$

where

$$E = \begin{bmatrix} C_{rr} & C_{rc} \\ C_{cr} & C_{cc} \end{bmatrix} \tag{4}$$

C_{rr} denotes the edge-weights between row labeling vertices that are both in V_r .C_{rc} denotes the edge-weights between labeling vertices with one in V_r and the other in V_c .C_{cc}; C_{cr} are defined similarly and $C_{rc} = C_{cr}^T$. Let $|E|_{ij}$ denote the (i,j)th element of E. $|E|_{ij}$ is the edge weight between two vertices. More specifically,

$$|E|_{ij} = \frac{\sum_{\alpha=1}^{k(i)} \sum_{\beta=1}^{k(j)} O_{\alpha,\beta} \log\left(\frac{|O|.O_{\alpha,\beta}}{O_\alpha^i O_\beta^j}\right)}{\sqrt{\left(\sum_{\alpha=1}^{k(i)} O_\alpha^i log \frac{O_\alpha^i}{|O|}\right)\left(\sum_{\beta=1}^{k(j)} O_\beta^i log \frac{O_\beta^i}{|O|}\right)}} \tag{5}$$

if the ith and jth vertices are both the row labeling vertices for enzyme clusters;

$$|E|_{ij} = \frac{\sum_{\alpha=1}^{\ell(i)} \sum_{\beta=1}^{\ell(j)} O_{\alpha,\beta} \log\left(\frac{|F|.F_{\alpha,\beta}}{F_\alpha^i F_\beta^j}\right)}{\sqrt{\left(\sum_{\alpha=1}^{\ell(i)} F_\alpha^i log \frac{F_\alpha^i}{|F|}\right)\left(\sum_{\beta=1}^{\ell(j)} F_\beta^i log \frac{F_\beta^i}{|F|}\right)}} \tag{6}$$

if the ith and jth vertices are both the column labeling vertices for enzyme clusters. Otherwise $|E|_{ij} = 0$

According to the bipartite graph $G = (V_r, V_c, \boldsymbol{E})$ given above, now we define the co-clustering partition matrix **Y** as

$$Y = \begin{bmatrix} Y_r \\ Y_c \end{bmatrix} \tag{7}$$

where$\boldsymbol{Y_r}$ is the partition on row labeling vertex set V_r and $\boldsymbol{Y_c}$ is the partition on column labeling vertex set V_c . Thus, the laplacian matrix L can be defined as

$$L = D - M \tag{8}$$

where

$$D = \begin{bmatrix} D_r & O \\ O & D_c \end{bmatrix} \tag{9}$$

$D_r$and$D_c$are diagonal matrices such that $|D_r|_{ii} = \sum_j E_{ij}$, $|D_c|_{jj} = \sum_i E_{ij}$. Note that the key step is to find the minimum cut vertex partitions on the bipartite graph. The normalized-cut objective function can be expressed as

$$\min_{\boldsymbol{Y}} \boldsymbol{tr}(\boldsymbol{Y^T L Y}) \tag{10}$$

One way to solve the partition problem of the bipartite graph is to compute the left and right eigenvectors of the matrix A defined as

$$A = D_r^{-1/2} E D_c^{-1/2} \tag{11}$$

After the left and right eigenvectors of matrix **A** are obtained, the left and right eigenvectors of the second to the $(\omega+1)th$ eigenvalues are selected as $U=[u_2,u_3,\dots,u_{\omega+1}]$ and $V=[v_2,v_3,\dots,v_{\omega+1}]$ respectively. Here, the $\omega = log_2 k$ singular vectors $u_2,u_3,\dots,u_{\omega+1}$, and $v_2,v_3,\dots,v_{\omega+1}$ often contain k-modal information about the original co-clustering labeling. Thus, the k-dimensional data matrix can be written as

$$X=\begin{bmatrix} D_r^{-1/2} & U \\ D_c^{-1/2} & V \end{bmatrix} \tag{12}$$

At last, ICS-RECCA is preformed on X, and the final consensus co-clustering result is obtained.

IV. ABOUT THE DATASET

Several datasets[Yanhua Chen et al.,2010] have been taken for the performance analysis. Here a single dataset is consider, the datasets for gene expression pairwise (Condition-Gene) coclustering is given in Table 1.

TABLE 1. Data Sets for Gene Expression Pairwise (Condition-Gene) Coclustering

Name	Datasets	Data Structure	No. of clusters	No. of documents
BT1	ALL/AML	ALL, AML	2	72
BT2	Breast Cancer	Relapse, Non-relapse	2	97
BT3	Central Nervous	Class1, Class2	2	60
BT4	Colon Tumor	Positive, Negative	2	62
BT5	Lung Cancer	MPM, ADCA	2	181
BT6	Ovarian Cancer	Cancer, Normal	2	253
BT7	ALL/MLL/AML	ALL,MLL,AML	3	72

V. RESULTS AND DISCUSSIONS

Performance of ICS-RECCA is made a comparison with Semisupervised Non-negative Matrix Factorization (SS-NMF) [Yanhua Chen et al.,2010], Non-negative Matrix Factorization (NMF) [Xu et al.,2003], Combinatorial Markov Random Field (CMRF) [Bekkerman and Jeon.,2007], Semisupervised Combinatorial Markov Random Field (SS-CMRF) [Bekkerman and Sahami.,2006], Spectral Relational Clustering (SRC) [Long et al.,2006] and Transductive Support Vector Machines (TSVM) [Joachims.,1999] in terms of accuracy and computation time. Figure 1 uses the Gene Expression Pairwise (Condition-Gene) Coclustering datasets depicted in Table2.

The experiments are performed on a Windows 8.1 machine with Intel Core i3 processors and 4 GB DDR III RAM. The experiments on algorithms are evaluated using MATLAB R2012a.

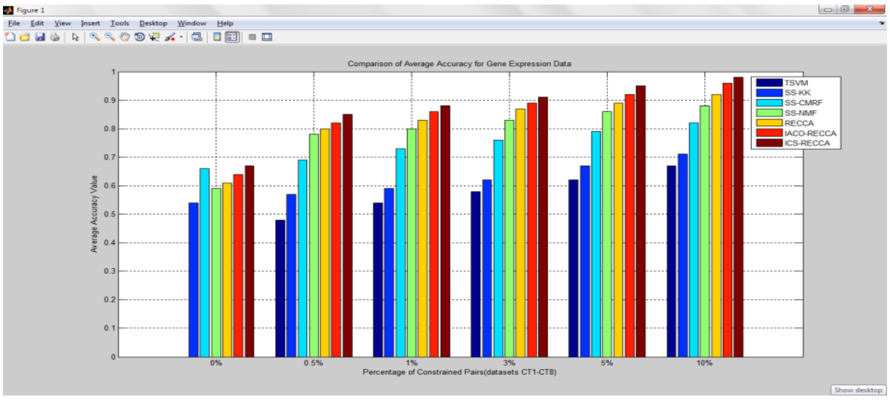

Figure 1. Comparison of Average Accuracy for Gene Expression Data

Figure 1 presents the performance evaluation of average accuracy for gene expression data. It is most visible that the proposed ICS-RECCA mechanism outperforms other mechanisms in terms of increasing percentage of pairwise constraints for semisupervised condition coclustering. The performance values are depicted in Table 2.

TABLE 2. Comparison of Average Accuracy for Gene Expression Data

Algorithms / Percentage of Constrained Pairs	TSVM	SS-KK	SS-CMRF	SS-NMF	RECCA	IACO-RECCA	ICS-RECCA
0%	0.00	0.54	0.66	0.59	0.61	0.64	0.67
0.5%	0.48	0.57	0.69	0.78	0.80	0.82	0.85
1%	0.54	0.59	0.73	0.8	0.83	0.86	0.88
3%	0.58	0.62	0.76	0.83	0.87	0.89	0.91
5%	0.62	0.67	0.79	0.86	0.89	0.92	0.95
10%	0.67	0.71	0.82	0.88	0.92	0.96	0.98

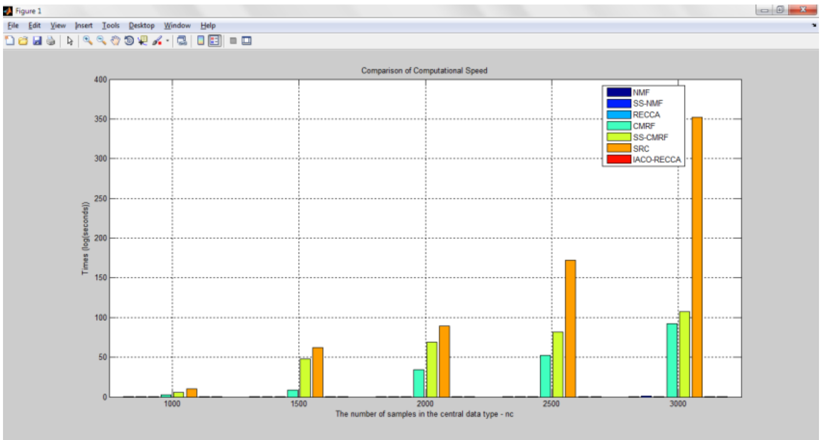

Figure 2. Comparison of Computational Speed - In Log(Seconds) For Increasing N_c

Figure 2 presents the performance of computational time (number of samples in the central data type - N_c) and the results proved that the proposed ICS-RECCA approach delivers significant better performance over other methods. The performance values are depicted in Table 3.

TABLE 3. Comparison of Computational Speed - In Log(Seconds) For Increasing Nc

Algorithms / Percentage of Constrained Pairs	NMF	SS-NMF	RECCA	CMRF	SS-CMRF	SRC	IACO-RECCA	ICS-RECCA
1000	0.05	0.23	0.21	3	6	10	0.19	0.15
1500	0.2	0.45	0.38	9	48	62	0.35	0.31
2000	0.1	0.56	0.52	34	69	89	0.48	0.43
2500	0.4	0.62	0.57	52	82	172	0.53	0.48
3000	0.52	0.84	0.74	92	107	352	0.70	0.65

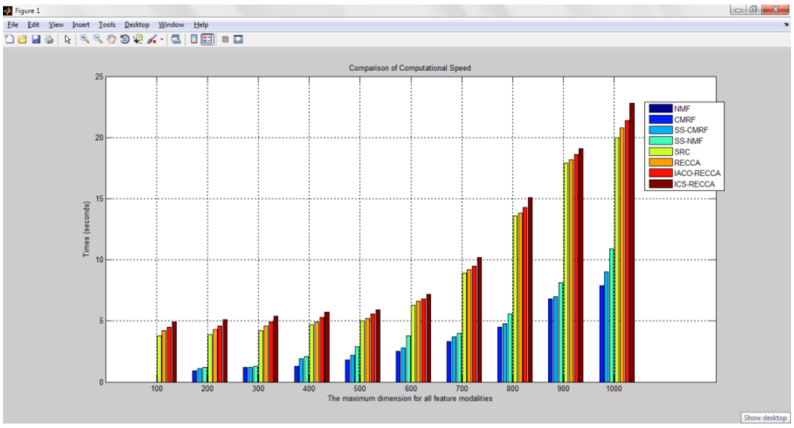

Figure 4. Comparison of Computational Speed - In Log(Seconds) For Increasing Np

Figure 4 presents the performance of computational time (the maximum feature dimension for all feature modalities - N_p) and the results proved that the proposed ICS-RECCAapproach delivers significant better performance over other methods. The performance values are depicted in Table 4.

TABLE 4.Comparison of Computational Speed - In Log(Seconds) For Increasing N_p

Algorithms / Percentage of Constrained Pairs	NMF	CMRF	SS-CMRF	SS-NMF	SRC	RECCA	IACO-RECCA	ICS-RECCA
100	0	0	0	0	3.8	4.2	4.5	4.9
200	0.2	0.9	1.1	1.2	3.9	4.3	4.6	5.1
300	0.2	1.2	1.22	1.3	4.2	4.6	4.9	5.4
400	0.1	1.3	1.9	2.1	4.7	4.9	5.3	5.7
500	0.2	1.8	2.2	2.9	5	5.2	5.6	5.9
600	0.3	2.5	2.8	3.8	6.3	6.6	6.8	7.2
700	0.4	3.3	3.7	4	8.9	9.2	9.5	10.2
800	0.2	4.5	4.8	5.6	13.6	13.8	14.3	15.1
900	0.4	6.8	7	8.1	17.9	18.2	18.6	19.5
1000	0.2	7.9	9	10.9	20	20.8	21.4	22.8

VI. CONCLUSION

This paper presented a mechanism with improved optimization technique namely ICS-RECCA for enzyme clustering. Initially the proposed work ICS-RECCA contains objective function for the co-clustering ensemble towards application to enzyme clustering is presented and also described. The objective function plays a major role which can perform co-clustering. Simulation results show that the proposed mechanism ICS-RECCA performs better in terms of accuracy and computation time. With the contribution of improved cuckoo search algorithm the co-clustering ensemble for enzyme clustering has been winded up for the doctoral research.

REFERENCES

[1] Ahmed Elkeran, A new approach for sheet nesting problem using guided cuckoo search andpairwise clustering, European Journal of Operational Research, Volume 231, Issue 3, 16 December 2013, Pages 757-769

[2] Ameryan, Totonchi, Mahdavi, "Clustering based on Cuckoo Optimization Algorithm," Intelligent Systems (ICIS), 2014 Iranian Conference on , vol., no., pp.1,6, 4-6 Feb. 2014

[3] Azizipanah-Abarghooee, Niknam, Zare, Gharibzadeh, "Multi-objective short-term scheduling of thermoelectric power systems using a novel multiobjective θ-improved cuckoo optimisation algorithm," Generation, Transmission & Distribution, IET , vol.8, no.5, pp.873,894, May 2014

[4] Bekkerman, Jeon, "Multi-Modal Clustering for Multimedia Collections," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[5] Bekkerman, Sahami, "Semi-Supervised Clustering Using Combinatorial MRFs," Proc. 23rd Int'l Conf. Machine Learning (ICML) Workshop Learning in Structured Output Spaces, 2006.

[6] Carlos Cobos, Henry Muñoz-Collazos, RicharUrbano-Muñoz, Martha Mendoza, Elizabeth León, Enrique Herrera-Viedma, Clustering of web search results based on the cuckoo search algorithm and Balanced Bayesian Information Criterion, Information Sciences, Volume 281, 10 October 2014, Pages 248-264

[7] Chifu, Salomie, Chifu, Izabella, Pop, Antal, M., "Cuckoo search algorithm for clustering food offers," Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on , vol., no., pp.17,22, 4-6 Sept. 2014

[8] Elyasigomari, Mirjafari, Screen, Shaheed, Cancer classification using a novel gene selection approach by means of shuffling based on data clustering with optimization, Applied Soft Computing, Volume 35, October 2015, Pages 43-51

[9] Emary, Zawbaa, Hassanien, Schaefer, Azar, "Retinal vessel segmentation based on possibilistic fuzzy c-means clustering optimised with cuckoo search," Neural Networks (IJCNN), 2014 International Joint Conference on , vol., no., pp.1792,1796, 6-11 July 2014

[10] Goel, Sharma, Bedi, "Cuckoo Search Clustering Algorithm: A novel strategy of biomimicry," Information and Communication Technologies (WICT), 2011 World Congress on , vol., no., pp.916,921, 11-14 Dec. 2011

[11] Joachims, "Transductive Inference for Text Classification Using Support Vector Machines," Proc. 16th Int'l Conf. Machine Learning, pp. 200-209, 1999.

[12] Karthikeyan, Venkatalakshmi, "Energy conscious clustering of Wireless Sensor Network using PSO incorporated cuckoo search," Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on , vol., no., pp.1,7, 26-28 July 2012

[13] Long, Wu, Zhang, Yu, "Spectral Clustering for Multi-Type Relational Data," Proc. 23rd Int'l Conf. Machine Learning, pp. 585-592, 2006.

[14] Mohapatra, Chakravarty, Dash, An improved cuckoo search based extreme learning machine for medical data classification, Swarm and Evolutionary Computation, Volume 24, October 2015, Pages 25-49

[15] Molisch, "Ultrawideband propagation channels—Theory, measurement, and modeling," IEEE Trans. Veh. Technol., vol. 54, no. 5, pp. 1528–1545, Sep. 2005.

[16] Rui Tang; Fong, S.; Xin-She Yang; Deb, S., "Integrating nature-inspired optimization algorithms to K-means clustering," Digital Information Management (ICDIM), 2012 Seventh International Conference on , vol., no., pp.116,123, 22-24 Aug. 2012.

[17] Shutin and Kubin, "Cluster analysis of wireless channel impulse responses with hidden Markov models," in Proc. IEEE ICASSP, May 2004, pp. 949–952.

[18] Xu, Liu, Gong, "Document Clustering Based on Non-Negative Matrix Factorization," Proc. 26th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 267-273, 2003.

[19] Yanhua Chen, Lijun Wang, Ming Dong, Member,"Non-Negative Matrix Factorization for Semisupervised Heterogeneous Data Coclustering", IEEE Transactions On Knowledge And Data Engineering, Vol. 22, No. 10, October 2010.