Cluster Integrated Updation Strategies for ACO Algorithms

Raghavendra G S
Faculty in Computer Science,
BITS-Pilani K.K. Birla Goa Campus, GOA

Prasanna Kumar N
Faculty in Mathematics,
BITS-Pilani K.K. Birla Goa Campus, GOA

ABSTRACT
Ant Colony Optimization (ACO) algorithm has evolved as the most popular way to attack the combinatorial problems. The ACO algorithm employs multi agents called ants that are capable of finding optimal solution for a given problem instances. These ants at each step of the computation make probabilistic choices to include good solution component in partially constructed solution, so that better solution can be obtained in the search process. The ant algorithms are typically characterized by co-operation among the ants, greedy, heuristics and feedback approaches that helps them to achieve their goals. In this paper, we propose new updation mechanism based on clustering techniques, which is aimed at exploring the nearby solutions region. We also report in detail the impact on performance due to integration of cluster and ACO.

Keywords
Ant, Combinatorial, Optimization, Greedy, Cluster.

1. INTRODUCTION
In recent years, there has been splurge in the algorithms that belongs to a metaheuristic class. One such algorithm is ACO, a nature inspired population based algorithm fascinated by the foraging behavior of the ants. The tiny creature ants, which are almost blind, can establish the shortest path from nest to the food source. The food hunting activity provides the formal framework to solve the combinatorial optimization problems and it is most commonly applied to Travelling Salesman Problem (TSP). Ants deposit a chemical substance called pheromone trial during their journey. The pheromone trial has a important role to play in ACO framework. It acts as an indirect medium of communication through which ants can share their journey experience. It is basically a volatile substance that evaporates over the period of time and this mechanism helps to forget the bad experiences of the journey. The first algorithmic framework that captures the essence of ant activity was given by Dorigo te al [1]. In literature, many variants of ant algorithms have been proposed and each algorithm improvises the earlier versions [2-6]. These improvised algorithms try to balance the intensification and diversification factors. An Ant which intensifies the search near the optimally best solution region may not get globally best solution. Similarly, diversifying the search in search space will get the globally best solution, but need more time to converge. Therefore it is necessary to strike the balance between intensification and diversification for better performance in terms of quality of solution found and the optimal time needed to converge. The Ant algorithms have been successfully applied to various benchmark problems like Job-Shop Scheduling (JSP), the Vehicle Routing Problem (VRP), Graph Coloring Problem (GRP) and Quadratic Assignment Problem (QAP) and have been extended to continuous search domain also. Based on the literature survey, research work related to ACO can be classified into following categories: Devising new strategies for pheromone updation [8,11], Reward -Penalty approaches [13], Dynamic parameter adjustment [9-10], Hybridization of ant algorithms [14], Proofs for convergence [7,12,16,17] and applying the ant algorithms to multidisciplinary fields.

2. CLUSTERING
Data clustering [20] is one of the most important human activity that involves discovering groups and identifying interesting distribution and patterns in the underlying data. Clustering problem is about partitioning the data into groups/classes of objects in such a way that the objects within the group are very similar and the objects across the group are quiet different. Clustering is an unsupervised approach, where no labeled data will be available. The ultimate goal of the clustering is to assign the unlabelled data to labeled classes. The labels to the classes are categorical in nature and are purely data driven; that is, they are obtained from data. It is possible that sometime even class labels may not be defined, but still cluster process should identify the natural closeness among the data and should group them. In general, data belong to only one cluster. However, it is possible for the data to belong more than one clusters and its association with particular cluster is determined by the degree of membership. We will discuss some of the important clustering algorithms available in the literature and these algorithms will be used to come up with cluster integrated ACO algorithm.

2.1 STING
STING (Statistical Information Grid) [18] is a grid based clustering technique. The grid based approach divides the spatial area into rectangular cells. The rectangular cells are partitioned recursively and hierarchical structure is employed to represent them. The Figure 1 shows the hierarchical representation of the 2-d data space. The hierarchical structure represents several layers of rectangular cells and each layer represents the data at different level of resolution. The hierarchical structure follows the parent - child relationship in which lower level cells are partitioned cells of higher level cells. The size of the leaf level cells can be set to desired density of the data to be present in the area. The cells at higher level store the statistical information like maximum, minimum, mean and type of distribution etc about lower level cells. The STING method is useful for query

answering, where relevant cells are considered for searching data. The statistical information is used in top-down fashion to compute the confidence interval (CI). The CI will reflect the cell's relevancy to the posed query. The cells with highest relevancy will be considered for further processing. The process of finding relevant cells will continue until most relevant leaf cell is found.

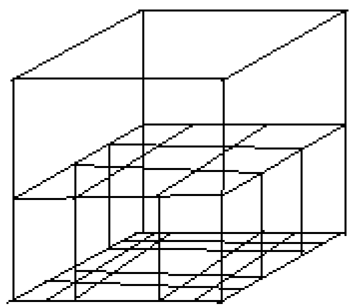


Figure 1 : Hierarchical Structure of 2-D Data space

2.2 k-Means algorithm

The k -means algorithm is one of the simplest and most commonly used clustering algorithms [15]. The algorithm creates k partitions of the given n data points. The algorithm employs the square-error criterion with the intension to minimize the variance within the clusters and to maximize the variance across the clusters. Typically, squared-error criterion computes the distance between the data points and the center (mean) of all the clusters and then assigns the data points to the closest cluster center. The working of k-Means algorithm is as follows:

Input: The number of clusters k and the n data sets.
Output: A set of k clusters that minimizes the squared-error criterion.

Method:
1. Arbitrarily choose k data as the initial cluster centers.
2. (Re)assign each of the data to the nearest cluster C_w , i.e.,
$x_j \in C_w$, if $\| x_j - m_w \| < \|x_j - m_i\|$
for j = 1, 2, ...N , i = w, and i = 1,2,...k.
3. Recalculate the cluster center (mean) for the current partition.
4. Repeat the steps 2 and 3 until there is no change for each cluster.

3. INTEGRATION OF CLUSTERS IN ACO

The experimental simulation reveals that there exist a correlation between the quality of solution found and the distance from good or optimal solutions. In literature, several measures to access the quality of solution can be found and one such measure is fitness-distance correlation (FDC) function [19]. The FDC computes the correlation coefficient which determines the goodness of the obtained solutions with respect to global best solution. The correlation coefficient will have high positive value, if the obtained solution is near to the global best solution. Infact, for the problems like TSP [6], large number of local optimum solutions is concentrated in a small region near the global best solution. Inorder to exploit the regions near the best solution, we propose a cluster based updation strategy which reinforces the toured paths in an unconventional manner. The cluster based updation strategy has the following characteristics:

• It groups the nearby tour performances and each tour performances within the group is reinforced with the same amount of pheromone trial, thereby supports the exploitation of (best) solutions.
• It reinforces all the paths, thereby supports the exploration.

Ideally, best tour in the group will be selected and its performance will be taken as a reference for updating the rest of the paths present in the group. The following subsections will discuss the incorporation of clustering mechanism in ACO algorithm.
The general outline of the cluster integrated ACO is as follows:

Input: A dataset D, number of cities n, number of ants m, number of cluster k and parameter τ , η, ρ.
Output: The best tour length s_{bs}.
Method: Initialize the Pheromone trial, parameters and set sbs to null.
while termination condition not met **do**
$\chi_{iter} \leftarrow$ null
for j = 1, · · · , n **do**
s ← ConstructSolution
s ← LocalSearch Optional
if(f (s) < f (s_{bs})) or (s_{bs} = NULL) then $s_{bs} \leftarrow s$
$\chi_{iter} \leftarrow \chi_{iter} \cup \{s\}$
end for
Identify the clusters in χ_{iter}
Apply the Pheromone Updates for each cluster by choosing the best solution in the corresponding group.
end while

3.1 Grid Structured ACO (GS-ACO)

The grid based strategy quantizes the solution space into finite number of blocks. The tour performance of the ants are sorted in incremental fashion and then placed into the grid structure. Let $G = \{T L_1, T L_2 \cdots T L_n \}$ represents the grid structure, where each T Li represents the n sorted tour length. The grid structure containing the tour performance is divided into blocks. Let $B = \{B_1 , B_2 \cdots B_m \}$ represents the m equisize partitioned block of the grid structure and $m \le n$. The Figure 2 shows the grid structure containing four equisize blocks and eight tour performances. It can be observed that each block contains variable number of tour performances and some blocks may be empty as well. The grid strategy treats each block as a cluster and updates the individual tour paths in the cluster with same amount of pheromone trial as that of best tour path. It should be noted that algorithm degenerates to normal AS algorithm, when the number of blocks is equal to number of ants.

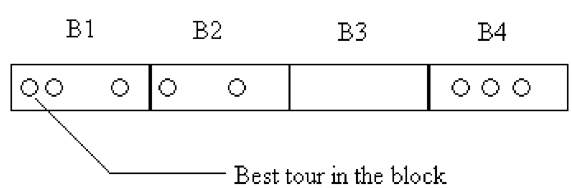


Figure 2 : Grid Structure containing tour length.

3.2 k-Means ACO (kM-ACO)

In previous section, we discussed about grid strategy that treats the partitioned blocks as a cluster. The number of partitions is fixed as a part of parameter setting, but the evolved blocks doesn't reflect as a natural cluster. Consider two blocks B1, B2 and two points in the Figure 2. The first point under consideration is the one located at right end of the block B1 and the second one is located at the left end of the block B2. The two points may be logically belonging to the same group, but the rigid partition scheme puts them in two separate blocks. This necessitates using the clustering scheme that place the logically near data into same groups. The k-Mean algorithms have been employed, that clusters the data depending on the distribution of data points. The updation strategy followed is similar to that of grid strategy, where all the tour paths in the cluster will be reinforced with the same amount of pheromone trial as that of best tour path in the cluster.

4. EXPERIMENTAL STUDY

4.1 Parameter Settings

Since clustering mechanism has been incorporated in the ACO algorithms, additional parameters pertaining to cluster need to be specified as a part of parameter settings. The Grid structured ACO and k-Means ACO need k number of clusters as a parameter for clustering process. We made extensive simulations by varying the parameters α, β from 1 to 5, ρ from 0.7 to 1.0 with the increment of 0.03 and number of ants m were varied in range from {10,n/2,n}, where n is the number of ants in the system. The parameter k was varied in the range of 20-80% of the number of ants and the total number of iterations was set to 1,00,000.

4.2 Result Analysis for Primary Updation

The Table 1 shows the comparative results of cluster integrated ACO for primary updation. The general observation is that grid strategy exhibits larger deviation from optimal solution compared to k-Mean strategy. In grid strategy, most of the obtained solutions deviate by 1%. Inorder to access the behavior of algorithm, graphs are plotted for varying number of ants, clusters and pheromone at different intensity level. In subsequent section, we will present the graphical analysis of the behavior of all the cluster integrated ACO algorithms.

Table 1. Comparative Results for Cluster integrated ACO algorithms for primary updation.

Datasets	Algorithm	Best (Std Dev)	Average (Std Dev)
bays29	GS-ACO	2046.4 (1.30%)	2061.2 (2.03%)
	kM-ACO	2061.2 (2.03%)	2038.5 (0.91%)
att48	GS-ACO	10745.8 (1.10%)	10807.5 (1.68%)
	kM-ACO	10685.6 (0.54%)	10704.4 (0.71%)
eil51	GS-ACO	432.7 (1.57%)	435.1 (2.13%)
	kM-ACO	428.3 (0.53%)	432.9 (1.61%)
st70	GS-ACO	684.3 (1.37%)	688.2 (1.95%)
	kM-ACO	681.6 (0.97%)	684.4 (1.39%)
eil76	GS-ACO	547.4 (1.74%)	552.3 (2.65%)
	kM-ACO	543.7 (1.05%)	546.7 (1.61%)
kroa100	GS-ACO	21397.4 (0.54%)	21439.9 (0.74%)
	kM-ACO	21348.5 (0.31%)	21376.2 (0.44%)
kroa200	GS-ACO	29740.5 (1.26%)	29793.7 (1.44%)
	kM-ACO	29545.6 (0.60%)	29580.4 (0.72%)
lin318	GS-ACO	42778.7 (1.78%)	42838.6 (1.92%)
	kM-ACO	42434.2 (0.96%)	42516.9 (1.16%)

4.3 Result Analysis for Secondary Updation

We extended the proposed approach by incorporating additional reinforcement mechanism. The additional reinforcement is done after primary updation. The additional/ secondary updation is adopted to provide the diversification for the search process. The primary updation mechanism updates the pheromone trial proportional to the quality of solution found. The secondary updation mechanism uses cluster based updation strategy to reinforce the traveled paths. The Table 2 shows the comparative results for the cluster integrated ACO for secondary updation. On comparing Table 2 with Table 1, it can be observed that secondary updation strategy improvises most of the best solutions and average solutions.

Table 2. Comparative Results for Cluster integrated ACO algorithms for secondary updation.

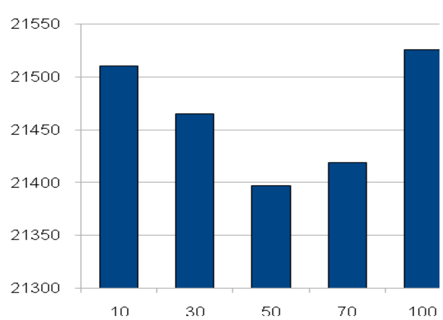
Datasets	Algorithm	Best (Std Dev)	Average (Std Dev)
bays29	GS-ACO	2040.4 (1.00%)	2055.1 (1.73%)
	kM-ACO	2028.4 (0.41%)	2034.4 (0.71%)
att48	GS-ACO	10711.7 (0.78%)	10756.3 (1.20%)
	kM-ACO	10672.6 (0.42%)	10680.1 (0.49%)
eil51	GS-ACO	430.3 (1.00%)	433.4 (1.73%)
	kM-ACO	429.1 (0.72%)	431.6 (1.31%)
st70	GS-ACO	680.1 (0.75%)	685.5 (1.55%)
	kM-ACO	679.4 (0.65%)	682.9 (1.17%)
eil76	GS-ACO	544.9 (1.28%)	548.8 (2.00%)
	kM-ACO	541.4 (0.57%)	545.3 (1.35%)
kroa100	GS-ACO	21355.7 (0.34%)	21381.3 (0.46%)
	kM-ACO	21338.4 (0.26%)	21355.8 (0.34%)
kroa200	GS-ACO	29518.8 (0.51%)	29572.4 (0.69%)
	kM-ACO	29478.1 (0.37%)	29508.8 (0.47%)
lin318	GS-ACO	42481.2 (1.07%)	42570.6 (1.28%)
	kM-ACO	42360.6 (0.75%)	42394.9 (0.87%)

4.4 Performance Analysis of Algorithms

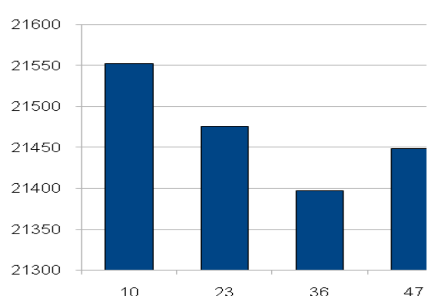
In this section, we will discuss the impact of various parameters affecting the performance of the algorithm. A comparative graph are drawn by varying number of ants, number of clusters and pheromone trials of different strength for each variant of the algorithm in order to access the performance. The execution profile graph is plotted to access the nature of convergence of algorithm.

4.4.1 Grid Structured ACO

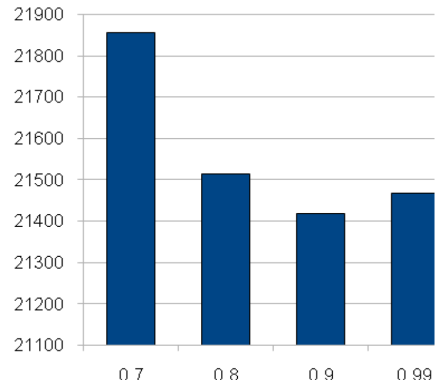
The Kroa100 dataset is used to access the behavior of algorithm, since it provides better result compared to other datasets. The Figure 3 shows the comparative results of primary updation for grid strategy. It can be observed from the Fig 3a that better results are obtained, when number of ants is around 50-70% of the total number of cities. The Fig 3b shows the results for varying number of clusters. The number of ants was set to 50 and the number of the clusters was set to {10, 23, 36, 47}. It can be observed that as the number of clusters increases, quality of solution improves and better solution was obtained, when number of cluster is around 70% of the number of ants. The pheromone trial is responsible for remembering the past experiences of the ants. In GS-ACO, lower persistent factor leads to poor quality of solution and search stagnation occurs, when pheromone trial strength is 0.7%. The Figure 4 shows the comparative results of secondary updation for GS-ACO. One can observe from Fig 3a and Fig 4a, quality of solution has improvised across the variation in the number of ants. Similar observation can be done for varying number of clusters. The algorithm yields poorer solution for lower pheromone strength, but search stagnation doesn't occur.



(a) Variation in number of ants

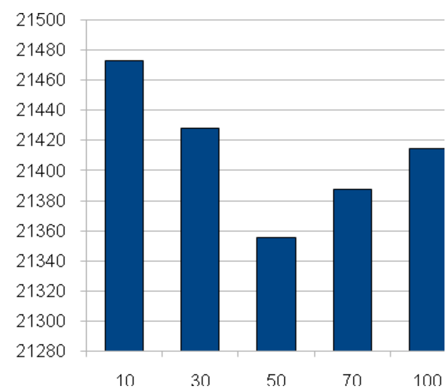


(b) Variation in number of clusters

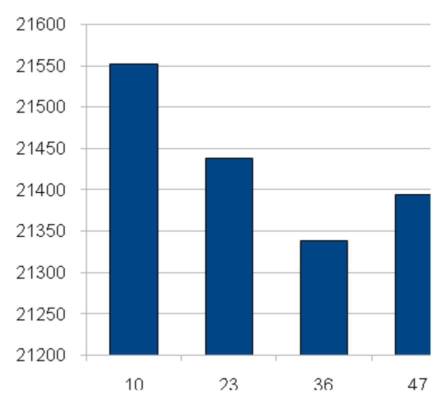


(c) Variation in pheromone trial

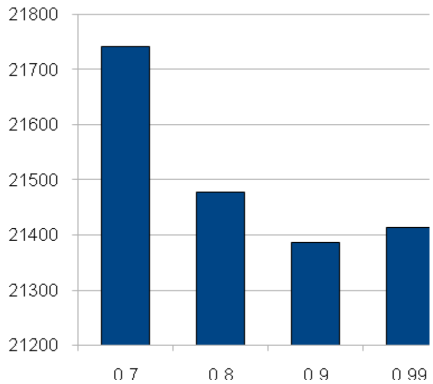
Figure 3: Comparative graph of GS-ACO for primary updation.



(a) Variation in number of ants



(b) Variation in number of clusters

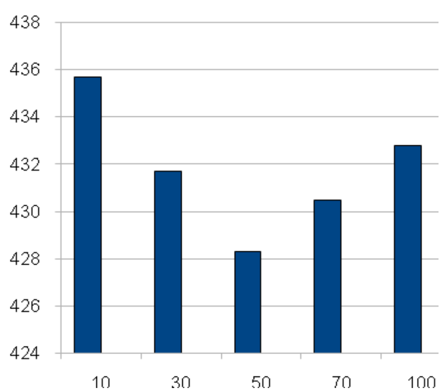


(c) Variation in pheromone trial

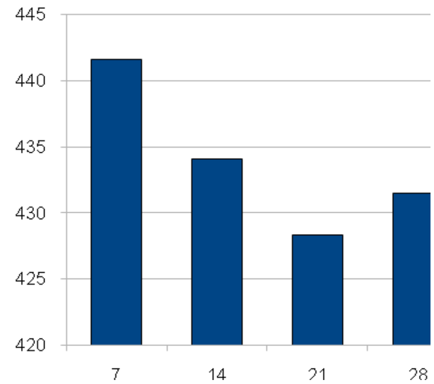
Figure 4: Comparative graph of GS-ACO for secondary updation.

4.4.2 k-Means ACO

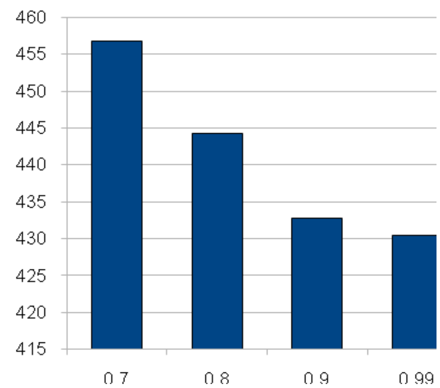
The eil51 dataset was selected to assess the behavior of the algorithm. The Figure 5 shows the comparative results of primary updation for kM-ACO strategy. The Fig 5a shows that kM-ACO provides better results, when there are around 30 number of ants. The number of ants was set to 30 and number of clusters was set to {7, 14, 21, 28} to assess the impact of varying number of clusters on performance as shown in Fig 5b. The better results were obtained, when the number of clusters is 70% of the number of ants. The Fig 5c shows that pheromone persistence has a greater effect on the obtained quality of solution and quality of solution sharply increases with the increase in pheromone trial strength. The Figure 6 shows the comparative results of secondary updation for GS-ACO. However, secondary updation provides comparatively inferior result than the primary updation.



(a) Variation in number of ants

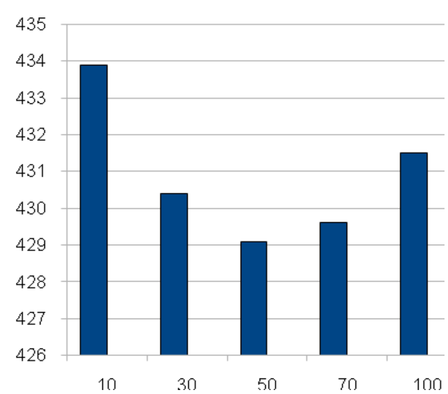


(b) Variation in number of clusters

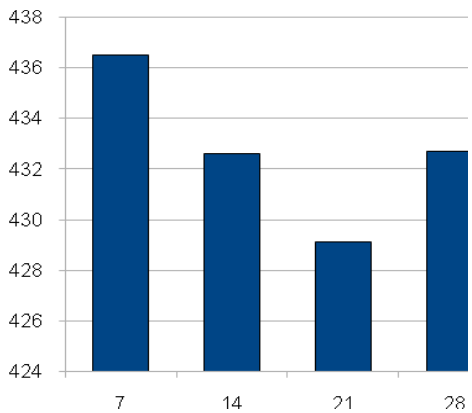


(c) Variation in pheromone trial

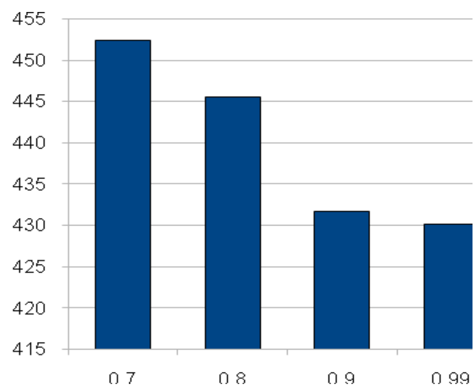
Figure 5: Comparative graph of kM-ACO for primary updation.



(a) Variation in number of ants



(b) Variation in number of clusters



c) Variation in pheromone trial

Figure 6: Comparative graph of kM-ACO for secondary updation.

5. CONCLUSION

In this paper, we have presented some novel approaches to update the pheromone trial. The updation strategy have been incorporated with cluster mechanism that reinforces the logically nearer paths with same amount of pheromone trial. The paper discussed about the incorporation of grid and k-mean strategy and the impact of incorporation on the performance of ACO algorithm. The performance analysis is done with respect to primary updation and secondary updation. For the first time, such a technique has been developed by us that can be used for any randomized search heuristics.

6. REFERENCES

[1] Dorigo,M., Maniezzo, V., and Colorni, A.(1996) Ant System: Optimization by a colony of cooperating agents, IEEE Transaction on Systems, Man and Cybernetics, Vol 26,No.1,pp.29-41.

[2] Bullnheimer, Hartl, R. F., and Strauss, C.(1999) A new rank based version of the Ant System: A computational study, Central European Journal for Operation Research and Economics, Vol 7,No.1,pp.25-38.

[3] Blum, C., Roli, A., and Dorigo, M.(2004) HC-ACO: The Hypercube framework for Ant Colony Optimization, IEEE Transaction on Systems, Man and Cybernetics, Vol 34,No.2,pp.1161-1172.

[4] Gambardella, L.M., and Dorigo, M.(1995) Ant-Q: A reinforcement learning approach to the traveling salesman problem. In Proceedings of 12th International Conference on Machine Learning.

[5] Gambardella, L.M., and Dorigo. M.(1997) Solving symmetric and asymmetric TSP by ant colonies. In Proceedings of IEEE International Conference on Evolutionary Computation. Porto,Spain.

[6] Stutzle, T., and Hoos, H.H.(2000) MAX - MIN Ant System, Future Generation Computer System, Vol 16,No 8,pp.889-914.

[7] Hong-hao, Z., and Fan-lun, X.(2006) A New Approach of Ant Colony Optimization and its Proof of Convergence. In Proceedings of 6th World Congress on Intelligent Control and Automation.

[8] Jun, S., Sheng-Wu and Fu-Ming, G.(2004) A New Pheromone Updating Strategy in Ant Colony Optimization. In Proceedings of 3rd International Conference on Machine Learning and Cybernetics.

[9] Lijie, L., Shangyou, J., and Ying, Z.(2008) Improved Ant Colony Optimization for the Traveling Salesman Problem. In Proceedings of International Conference on Intelligent Computation Technology and Automation.

[10] Masaya, Y., Masahiro, F., and Hidekazu, T.(2008) A New Pheromone Control Algorithm of Ant Colony Optimization. In Proceedings of International Conference on Smart Manufacturing Application.

[11] Naimi, H.M., and Taherinejad, N.(2009) New Robust and efficient ant colony algorithms: Using new interpretation of local updating process, Expert System with Applications, Vol 36,No.1,pp.481-488.

[12] Stutzle, T., and Dorigo, M.(2002) A Short Convergence Proof for a Class of Ant Colony Optimization, IEEE Transaction on Evolutionary Computation, Vol 20,No.3,pp.1-9.

[13] Raghavendra, G. S., and Prasanna K.N.(2008) Relative Reward Pheromone Update Strategy for ACO Algorithms. In Proceedings of MSAST2008, IMBIC,Kolkata,India.

[14] Blum, C.(2007). Ant Colony Optimization: Introduction and Hybridization. In Proceedings of 7th International Conference on Hybrid Intelligent Systems.

[15] MacQueen J (1967) Some methods for classification and analysis of multi-variate observations. In Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, pp.281-297.

[16] Prasanna, K., and Raghavendra, G.S.(2011) On the Evaporation Mechanism in the Ant Colony Optimization Algorithms, Annales of Computer Science Series, Vol 9,No.1,pp.51-56.

[17] Prasanna, K., and Raghavendra, G.S.(2011) A note on the Parameter of Evaporation Mechanism in the Ant Colony Optimization Algorithms, International Mathematical Forum, Vol 6,No.34,pp.1655-1659.

[18] Wang, W., Yang, J., Muntz, R (1997) STING: A Statistical information grid approach to spatial data mining. In Proceedings of International Conference on Very Large Databases, pp.186-195.

[19] Jones T, Forrest S (1995) Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In Proceedings of the 6th International Conference on Genetic Algorithms, pp.184-192.

[20] Rui X, Donald W (2005) Survey of Clustering Algorithms. IEEE Transaction On Neural Networks, Vol.16,No 3,pp.645-678.