A Modified ACO for Classification on different Data Set

Dharmpal Singh
Department of CSE
Pailan College of Management and Technology
Sector - 1, Phase I, Bengal Pailan Park, Off Diamond Harbour Road, Joka, Kolkata, West Bengal 700104

J. Paul Choudhury
Department of IT
Kalyani Govt. Engineering College, Kalyani, Nadia-741235, West Bengal, INDIA

Mallika De
Department of Engineering & Technological Studies (Retired), Kalyani University Kalyani, Nadia-741235, West Bengal, INDIA

ABSTRACT
Ant colony optimization algorithms have been applied to many combinatorial optimization problems, ranging from quadratic assignment to protein folding or routing vehicles and a lot of derived methods have been adapted to dynamic problems in real variables, stochastic problems, multi-targets and parallel implementations. It has also been used to classification of the data set based on the attribute. It has been observed that construct solution and pheromone update play an important role in the ACO algorithm. The selection of the pheromone update is based on the construct solution which is further base on the probability function and initial selection. So if the selection of the pheromone done properly then ACO algorithm will terminate in less number of the iteration and it will be produce the good result. It has further observed that difference result have been possible for the different selection of the construct and pheromone on the same data set. Therefore, in this paper an effort has been made to suggest the techniques to select the initial construct and pheromone update for data set and the classification has to be done using the concept of clustering.

Keywords
Data mining, soft computing, Ant colony optimization, Particle swarm optimization, fuzzy, neural network, data mining preprocessing.

1. INTRODUCTION
Ants live together in colonies and they use chemical cues called pheromones to provide a sophisticated communication system. An isolated ant moves essentially at random but an ant encountering a previously laid pheromone will detect it and decide to follow it with high probability and thereby reinforce it with a further quantity of pheromone. The repetition of the above mechanism represents the collective behaviour of a real ant colony which is a form of autocatalytic behaviour where the more the ants follow a trail, the more attractive that trail becomes. The above behaviour of real ants has inspired ACO which has proved to be an effective metaheuristic technique for solving many complex problems. This technique uses a colony of artificial ants that behaves as cooperative agents in a mathematical space where they are allowed to search and reinforce pathways (solutions) in order to find the optimal ones. The features of artificial ants are: having some memory, not being completely blind and the process time is discrete. In the ACO technique an initialisation phase takes place during which ants are positioned on different nodes (sessions) with empty tabu lists and initial pheromone distributed equally on paths connecting these sessions. Ants update the level of pheromone while they are constructing their schedules by iteratively adding new sessions to the current partial schedule. At each time step, ants compute a set of feasible moves and select the best one according to some probabilistic rules based on the heuristic information and pheromone level. The higher value of the pheromone and the heuristic information, the more profitable is to select this move and resume the search. The selected node is putted in the tabu list related to the ant to prevent to be chosen again**.** Heuristic information represents the nearer sessions around the current session, while pheromone level "memory" of each path represents the usability of this path in the past to find good schedules. At the end of each iteration, the tabu list for each ant will be full and the obtained cheapest schedule is computed and memorized. For the following iteration, tabu lists will be emptied ready for use and the pheromone level will be updated. This process is repeated till the number of iterations (stopping criteria) has been reached.

1.1 ACO algorithm
A general outline of the ACO algorithm is furnished below.

Algorithm ACO meta heuristic();
while (termination criterion not satisfied)
ant generation and activity();
pheromone evaporation();
daemon actions(); "optional"
end while
end Algorithm

2. RELATED WORK AND LITERATURE SURVEY
Thomas Stutzle and Marco Dorigo[1], have tried to solve the travelling salesman problem with an ant colony optimization and authors have given an overview on the available ACO algorithms for the TSP. Thereafter authors have outlined how ACO algorithms can be applied to TSP problem and presented the available ACO algorithms for the TSP. In section 1.4, they have briefly discusses local search for the TSP, while Section 1.5 presented experimental results which have been obtained with MAX –MIN Ant System. The authors have further opined that they have been applied to several other combinatorial optimization problems because it was the first application of ACO algorithms to the TSP. Furthermore, the authors opined that ACO algorithms have proved to be among the best available algorithms and in section 1.6 they have given a concise overview of these other applications of ACO algorithms.

Wei Zhao et al.[2] have tried to solve combination optimization problems using ant colony optimization. The authors have presented revised pheromones in local and global update mode for solving TSP. The authors have proposed a pheromone increment model called ant constant and a pheromone diffusion model and shown that proposed model given better optimal solutions on different benchmark data sets.

Héctor D. Menéndez, et al. [3] have presented an ACO-based clustering algorithm inspired by the ACO Clustering (ACOC) algorithm and authors have restructures ACOC from a centroid-based technique to a medoid-based technique, where the properties of the search space are not necessarily known. The authors have compared proposed algorithm with ACO Clustering (ACOC) algorithm on both synthetic datasets and real-world datasets extracted from the UCI Machine Learning Repository. The authors have opined that proposed algorithm outperformed the early algorithm based on its accuracy.

Xiaoyong Liu [4] has presented an improved clustering algorithm with Ant Colony optimization (ACO) based on dynamical pheromones. The author has proposed two strategies to improve the performance of the proposed algorithm. The author has adjusted the rate of pheromone evaporation dynamically, and the other has to adjust the strength of pheromone dynamically. The author has compared the proposed method based on the two indices named as Precision and Recall. The author has opined that it has given the best result as compared to the other algorithm.

R. F. Tavares Neto and M. Godinho Filho [13] have presented a literature survey on the uses of ACO approach to solve scheduling problems. The authors have opined that it will not only able to derive certain guidelines for the implementation of ACO algorithms but also to determine possible directions for future research.

Though the ant colony algorithms can resolve several optimization problems successfully, but it cannot establish its convergence. The many limitations have been observed from various survey paper which has been furnished below.

1) ACO has slow convergence speed and low efficiency [7, 8, 10, 11, and 12].

2) It is likely to fall in the local optimal solution [7, 8, 10, 11, and 12]. The main cause of this premature convergence is that the ant colony algorithms update the pheromone corresponding to the current better path, and after certain iterations, the pheromone on the best path becomes very strong, while the pheromone on the worthy path remains very weak. All the ants stick on the best path and it becomes very difficult to jump out from this best path. This raises the likelihood that the obtained optimal solution is only a local optimal solution.

3) Limited domain of inputs can be explored by using ACO [5, 10, and 12]. The studies also reported that in those cases ACO could be applied only to integral input domain.

4) Some studies also stated that redundant paths could not be deleted by using the ACO technique [6, 9]. This implies that various ants can find the test cases that meet the same test adequacy criterion and hence become redundant, whereas ants should be able to explore new paths without taking into account the redundant test cases.

5) Restricted applications of ACO were also spotted by a couple of studies [10, 12]. According to these, their developed systems could not be useful for Object Oriented Programming. The reason behind this could be that the initial approach followed by them did not include the concept of object-oriented programming.

Several authors have used the ACO algorithm for the solving of TSP problem [1], combination optimization problems [2] and Clustering [3, 4]. Moreover authors have also pointed out the problem of ACO in slow convergence speed and low efficiency [7, 8, 10, 11, and 12], fall in the local optimal solution [7, 8, 10, 11, and 12], limited domain of inputs can [5, 10, and 12], redundant paths could not be deleted by using the ACO technique [6, 9] and restricted applications of ACO [10, 12]. It has been observed that construct solution and pheromone update play an important role in the ACO algorithm. The selection of the pheromone update is based on the construct solution which is further base on the probability function and initial selection. So if the selection of the pheromone done properly then ACO algorithm will terminate in less number of the iteration and it will be produce the good result. It has further observed that difference result have been possible for the different selection of the construct and pheromone on the same data set. Therefore, in this paper an effort has been made to suggest the techniques to select the initial construct and pheromone update for data set such that it will produce same result for different data set. Here modified form of the ACO algorithm has also been proposed.

In first section, abstract and introduction have been furnished. In second section, related work and literature survey of the models have been furnished. In third section, implementation has been furnished in detail. In fourth section, result and conclusion have been furnished

3. METHODOLOGY AND IMPLEMENTATION

The concept of data mining [14], association rule [15], factor analysis [15], fuzzy logic (14,15 16) neural network [14], PSO [14] and ABC [17] have been elaborated in the mentioned paper. Therefore, due to size of paper only detail explanation of ACO has been furnished in section 3.1.

3.1 Data cleaning and formation of association rule

3.1.1 Purpose of work

The available data contains the information of Iris flowers with various items. The data related to iris flower containing 150 data values. It is to note that the quality of flower depends on the sepal length of the flower. If the sepal length of a particular flower is known, the quality of flower can be ascertained. Therefore the sepal length of flower (D) has been chosen as the objective item (consequent item) of the flowers. The other parameters i.e. sepal width (A), petal length (B) and petal width (C) have been chosen as the depending items (antecedent items).

The purpose of this work is to correlate the items A, B and C with D, so that based on any value of A, B and C, the value of D can be estimated. From the value of D the quality of that type of flower can be ascertained.

3.1.2 Data mining Preprocessing

Now it is necessary to check whether the data items are proper or not. If the data items are proper, extraction of information is possible otherwise the data items are not suitable for the extraction of knowledge. In that case preprocessing of data is

necessary for getting the proper data. Therefore, the data mining preprocessing techniques like data cleansing, data integration, data transformation and data reduction have to be applied on the available data as follows:

Data Cleansing
The data cleansing techniques include the filling in the missing value, correct or rectify the inconsistent data and identify the outlier in the available data.

It has been observed that all the data sets which have been considered do not contain any missing value. The said data sets do not contain any inconsistent data i.e. any abnormally low or any abnormally high value. All the data values are regularly distributed within the range of that data items. Therefore the data cleansing techniques are not applicable for the available data.

Data Integration
The data integration technique has to be applied if data has been collected from different sources. The available data have been taken from a single source therefore the said technique is not applicable here.

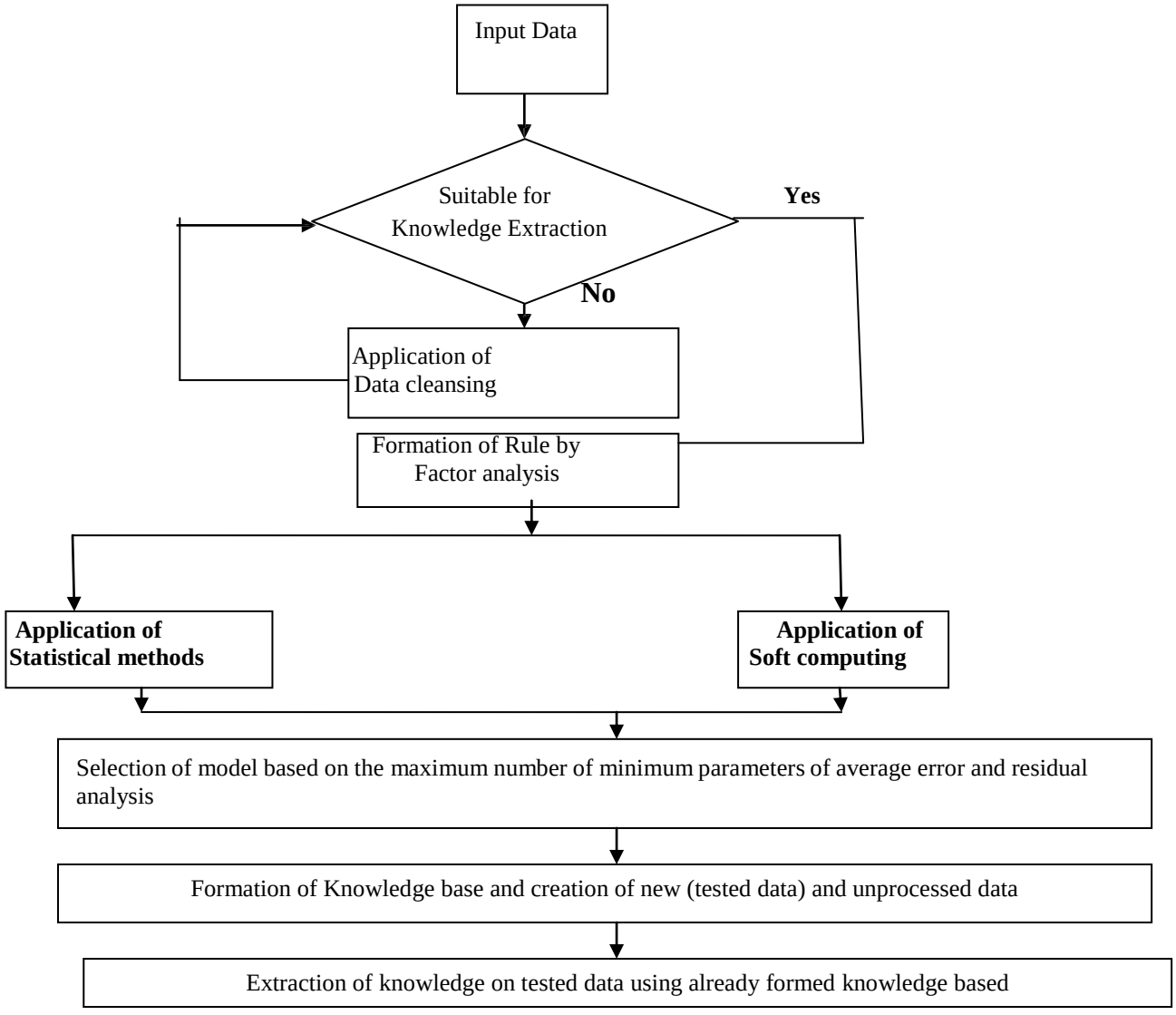
Data Transformation
To make the data within specific range smoothing and normalization techniques have to be applied. Decimal scaling technique has to be applied to move the decimal point to particular position of all data values. Out of these data transformation, smoothing and decimal scaling (multiply by 1000) techniques have been applied on the set.

Data Reduction
The data reduction method has to be applied on huge amount of data (tetra byte) to get the certain portion of data. Since here the amount of data is not large, said technique has not been applied.

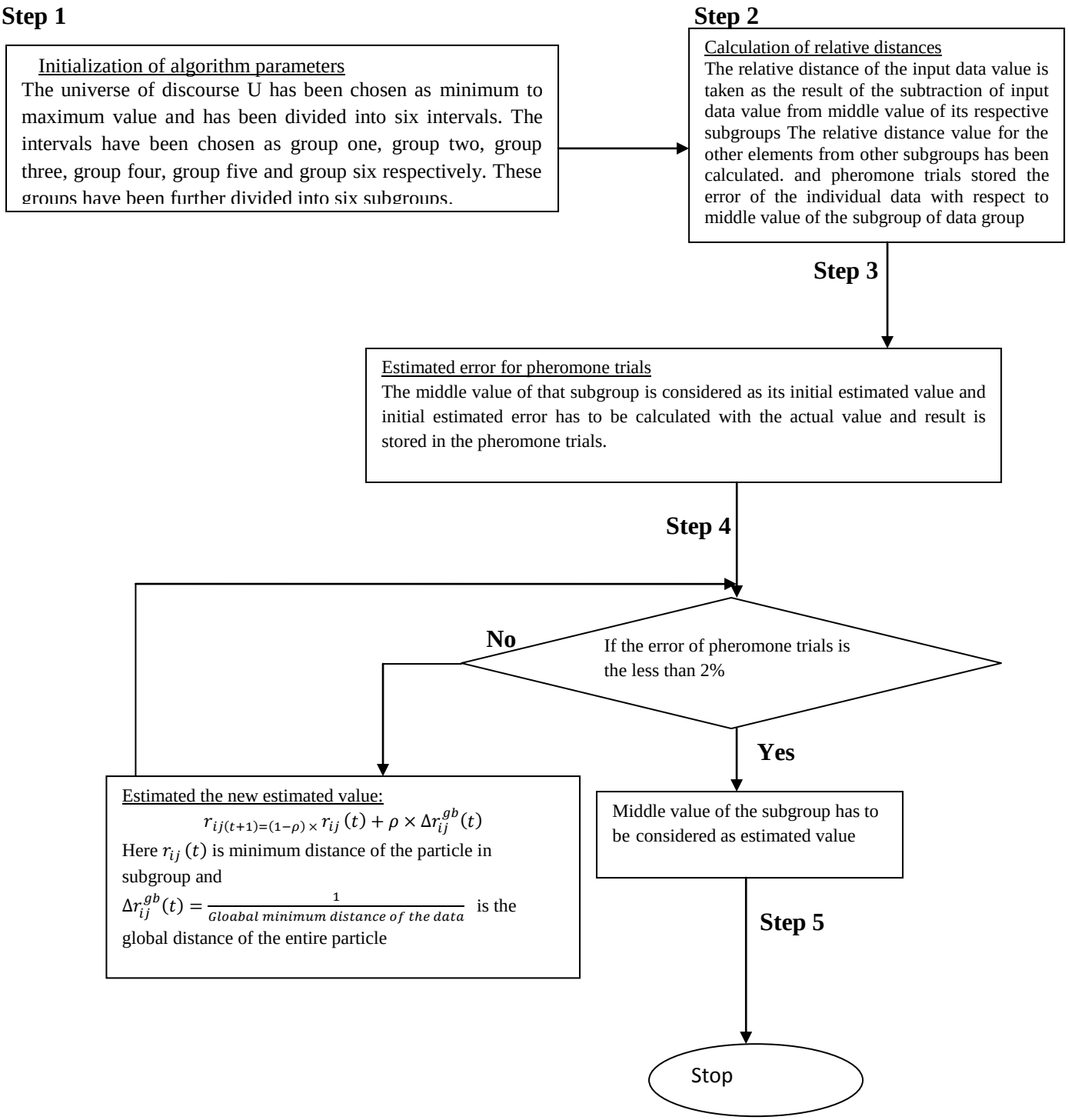
The data set has been taken from well know web site where it have been stored in well from Therefore, in this paper only data scaling has been applied on the said data set

3.2 Implementation

The details of the formation of association rule and application of the factor analysis and principal component analysis has been discussed in the paper [18]. The concept of data mining [14], association rule[15], factor analysis [15], fuzzy logic (14,15 16) neural network [14], PSO [14] and ABC [17] have been elaborated in the mentioned paper. Due to size of the paper, only details procedure of ACO is discussed in this paper.



3.3 Modified Ant Colony algorithm



3.4 Implementation Ant Colony Optimization

The value encoding has to be considered as the data encoding and the (1/ estimated error) of the estimated data with the actual data has to be considered as a fitness function of the estimated data.

The universe of discourse U has been chosen as 3330 to 9816 and has been divided into several intervals. The intervals have been chosen as A_1, A_2, A_3, A_4, A_5 and A_6 which are termed as group one, group two, group three, group four, group five and group six respectively. These groups have been further divided into six subgroups. The ranges of groups and subgroups have been furnished in table 1. Fuzzy sets have been defined on the universe and some linguistic values have also been determined. Let, A_1 = (many), A_2 = (many, many), A_3 = (very many, A_4 = (too many) be the possible values.

Table 1. Group and Subgroup Range

Group/Subgroup	Subgroup					
Group One (A_1)	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{16}
(3300-4386)	(3300-3481)	(3481-3662)	(3662-3843)	(3843-4024)	(4024-4205)	(4205-4386)
Group Two (A_2)	A_{21}	A_{22}	A_{23}	A_{24}	A_{25}	A_{26}
(4386-5475))	(4386-4567)	(4567-4748)	(4748-4929)	(4929-5110)	(5110-5291)	(5291-5472)
Group Three(A_3)	A_{31}	A_{32}	A_{33}	A_{34}	A_{35}	A_{36}
(5472-6588)	(5472-5653)	(5653-5834)	(5834-6015)	(6015-6196)	(6196-6377)	(6377-6558)
Group Four (A_4)	A_{41}	A_{42}	A_{43}	A_{44}	A_{45}	A_{46}
(6588-7644)	(6558-6739)	(6739-6920)	(6920-7101)	(7101-7282)	(7282-7463)	(7463-7644)
Group Five(A_5)	A_{51}	A_{52}	A_{53}	A_{54}	A_{55}	A_{56}
(7644-8730)	(7463-7825)	(7825-8006)	(8006-8187)	(8187-8386)	(8386-8549)	(8549-8730)
Group Six (A_6)	A_{61}	A_{62}	A_{63}	A_{64}	A_{65}	A_{66}
(8730-9816)	(8730-8911)	(8911-9092)	(9092-9273)	(9273-9454)	(9454-9635)	(9635-9816)

Step 1

The relative distances of the data value from each subgroup of that group have been calculated. The relative distance of the input data value is taken as the result of the subtraction of input data value from middle value of its respective subgroups as shown in step 2 of table 2 of particle swarm optimization. The relative distance value for the other elements from other subgroups has been calculated and pheromone trials stored the error of the individual data with respect to middle value of the subgroup of data group.

As for example, the minimum distance of data element (3331.597) of serial number 1 of the of table2 is 58.903 (minimum distance of serial number one of group one is denoted by $r_{ij}(t)$).The middle value of that subgroup is considered as its initial estimated value and initial estimated error has to be calculated with the actual value and result is stored in the pheromone trials.

Table 2

Distance of each Element from Respective Group

Serial No.	Available Data	Range of Data					
	Group One(A_1)	A_{11}(3300-3481)	A_{12} (3481-3662)	A_{13} (3662-3843)	A_{14} (3843-4024)	A_{15} (4024-4205)	A_{16} (4205-4386)
1	3331.597	58.903	239.903	420.903	601.903	782.903	963.903
2	3734.387	343.887	162.887	18.113	199.113	380.113	561.113
3	3912.366	521.866	340.866	159.866	21.134	202.134	383.134
4	3939.355	548.855	367.855	186.855	5.855	175.145	356.145
5	3941.134	550.634	369.634	188.634	7.634	173.366	354.366
6	4009.457	618.957	437.957	256.957	75.957	105.043	286.043
7	4066.994	676.494	495.494	314.494	133.494	47.506	228.506
8	4079.558	689.058	508.058	327.058	146.058	34.942	215.942
9	4104.769	714.269	533.269	352.269	171.269	9.731	190.731
10	4104.769	714.269	533.269	352.269	171.269	9.731	190.731
11	4104.769	714.269	533.269	352.269	171.269	9.731	190.731
12	4135.317	744.817	563.817	382.817	201.817	20.817	160.183
13	4135.317	744.817	563.817	382.817	201.817	20.817	160.183
14	4146.102	755.602	574.602	393.602	212.602	31.602	149.398
15	4174.871	784.371	603.371	422.371	241.371	60.371	120.629
16	4203.639	813.139	632.139	451.139	270.139	89.139	91.861
17	4243.193	852.693	671.693	490.693	309.693	128.693	52.307
18	4300.73	910.23	729.23	548.23	367.23	186.23	5.23
19	4318.713	928.213	747.213	566.213	385.213	204.213	23.213
20	4340.285	949.785	768.785	587.785	406.785	225.785	44.785
Group Two (A_2)		A_{21} (4386-4567)	A_{22} (4567-4748)	A_{23} (4748-4929)	A_{24} (4929-5110)	A_{25} (5110-5291)	A_{26} (5291-5472)
21	4397.822	78.678	259.678	440.678	621.678	802.678	983.678
22	4426.59	49.91	230.91	411.91	592.91	773.91	954.91
23	4466.144	10.356	191.356	372.356	553.356	734.356	915.356
24	4466.144	10.356	191.356	372.356	553.356	734.356	915.356
25	4467.923	8.577	189.577	370.577	551.577	732.577	913.577
26	4494.913	18.413	162.587	343.587	524.587	705.587	886.587
27	4496.692	20.192	160.808	341.808	522.808	703.808	884.808
28	4534.467	57.967	123.033	304.033	485.033	666.033	847.033
29	4563.236	86.736	94.264	275.264	456.264	637.264	818.264
30	4565.014	88.514	92.486	273.486	454.486	635.486	816.486
31	4592.004	115.504	65.496	246.496	427.496	608.496	789.496
32	4602.79	126.29	54.71	235.71	416.71	597.71	778.71
33	4606.347	129.847	51.153	232.153	413.153	594.153	775.153
34	4674.67	198.17	17.17	163.83	344.83	525.83	706.83
35	4716.003	239.503	58.503	122.497	303.497	484.497	665.497
36	4739.435	262.935	81.935	99.065	280.065	461.065	642.065
37	4757.418	280.918	99.918	81.082	262.082	443.082	624.082
38	4757.418	280.918	99.918	81.082	262.082	443.082	624.082
39	4843.723	367.223	186.223	5.223	175.777	356.777	537.777
40	4897.621	421.121	240.121	59.121	121.879	302.879	483.879
41	4911.964	435.464	254.464	73.464	107.536	288.536	469.536
42	4922.832	446.332	265.332	84.332	96.668	277.668	458.668
43	4924.611	448.111	267.111	86.111	94.889	275.889	456.889
44	4955.158	478.658	297.658	116.658	64.342	245.342	426.342
45	5034.134	557.634	376.634	195.634	14.634	166.366	347.366
46	5061.256	584.756	403.756	222.756	41.756	139.244	320.244
47	5075.681	599.181	418.181	237.181	56.181	124.819	305.819
48	5174.551	698.051	517.051	336.051	155.051	25.949	206.949
49	5188.762	712.262	531.262	350.262	169.262	11.738	192.738
50	5228.449	751.949	570.949	389.949	208.949	27.949	153.051
51	5248.078	771.578	590.578	409.578	228.578	47.578	133.422
52	5268.003	791.503	610.503	429.503	248.503	67.503	113.497
53	5285.853	809.353	628.353	447.353	266.353	85.353	95.647
Group Four (A_3)		A_{31} (5472-5653)	A_{32} (5653-5834)	A_{33} (5834-6015)	A_{34} (6015-6196)	A_{35} (6196-6377)	A_{36} (6377-6558)
54	5559.144	3.356	184.356	365.356	546.356	727.356	908.356
55	5569.929	7.429	173.571	354.571	535.571	716.571	897.571
56	5577.259	14.759	166.241	347.241	528.241	709.241	890.241
57	5616.681	54.181	126.819	307.819	488.819	669.819	850.819
58	5697.568	135.068	45.932	226.932	407.932	588.932	769.932
59	5862.982	300.482	119.482	61.518	242.518	423.518	604.518
60	5877.325	314.825	133.825	47.175	228.175	409.175	590.175
61	6071.508	509.008	328.008	147.008	33.992	214.992	395.992
62	6098.497	535.997	354.997	173.997	7.003	188.003	369.003
63	6123.708	561.208	380.208	199.208	18.208	162.792	343.792
64	6127.266	564.766	383.766	202.766	21.766	159.234	340.234
65	6150.616	588.116	407.116	226.116	45.116	135.884	316.884
66	6186.582	624.082	443.082	262.082	81.082	99.918	280.918
67	6262.05	699.55	518.55	337.55	156.55	24.45	205.45
68	6267.469	704.969	523.969	342.969	161.969	19.031	200.031
69	6362.781	800.281	619.281	438.281	257.281	76.281	104.719
70	6371.788	809.288	628.288	447.288	266.288	85.288	95.712
71	6402.335	839.835	658.835	477.835	296.835	115.835	65.165
72	6431.104	868.604	687.604	506.604	325.604	144.604	36.396
Group Five (A_4)		A_{41} (6558-6739)	A_{42} (6739-6920)	A_{43} (6920-7101)	A_{44} (7101-7282)	A_{45} (7282-7463)	A_{46} (7463-7644)
73	6596.518	51.982	232.982	413.982	594.982	775.982	956.982
74	6603.664	44.836	225.836	406.836	587.836	768.836	949.836
75	6623.507	24.993	205.993	386.993	567.993	748.993	929.993
76	6625.286	23.214	204.214	385.214	566.214	747.214	928.214
77	6664.84	16.34	164.66	345.66	526.66	707.66	888.66
78	6664.84	16.34	164.66	345.66	526.66	707.66	888.66
79	6693.527	45.027	135.973	316.973	497.973	678.973	859.973
80	6704.395	55.895	125.105	306.105	487.105	668.105	849.105
81	6770.939	122.439	58.561	239.561	420.561	601.561	782.561
82	6826.615	178.115	2.885	183.885	364.885	545.885	726.885
83	6833.812	185.312	4.312	176.688	357.688	538.688	719.688
84	6869.809	221.309	40.309	140.691	321.691	502.691	683.691
85	6900.356	251.856	70.856	110.144	291.144	472.144	653.144
86	6912.92	264.42	83.42	97.58	278.58	459.58	640.58
87	6941.689	293.189	112.189	68.811	249.811	430.811	611.811
88	6979.464	330.964	149.964	31.036	212.036	393.036	574.036
89	6997.447	348.947	167.947	13.053	194.053	375.053	556.053
90	7008.233	359.733	178.733	2.267	183.267	364.267	545.267
91	7037.001	388.501	207.501	26.501	154.499	335.499	516.499
92	7038.78	390.28	209.28	28.28	152.72	333.72	514.72
93	7038.78	390.28	209.28	28.28	152.72	333.72	514.72
94	7175.344	526.844	345.844	164.844	16.156	197.156	378.156
95	7227.544	579.044	398.044	217.044	36.044	144.956	325.956
96	7231.102	582.602	401.602	220.602	39.602	141.398	322.398
97	7232.962	584.462	403.462	222.462	41.462	139.538	320.538
98	7245.445	596.945	415.945	234.945	53.945	127.055	308.055
99	7254.534	606.034	425.034	244.034	63.034	117.966	298.966
100	7259.87	611.37	430.37	249.37	68.37	112.63	293.63
101	7272.517	624.017	443.017	262.017	81.017	99.983	280.983
102	7299.506	651.006	470.006	289.006	108.006	72.994	253.994
103	7331.863	683.363	502.363	321.363	140.363	40.637	221.637
104	7409.162	760.662	579.662	398.662	217.662	36.662	144.338
105	7437.849	789.349	608.349	427.349	246.349	65.349	115.651
106	7437.849	789.349	608.349	427.349	246.349	65.349	115.651
107	7454.053	805.553	624.553	443.553	262.553	81.553	99.447
108	7468.396	819.896	638.896	457.896	276.896	95.896	85.104
109	7497.246	848.746	667.746	486.746	305.746	124.746	56.254
110	7520.597	872.097	691.097	510.097	329.097	148.097	32.903
111	7522.375	873.875	692.875	511.875	330.875	149.875	31.125

112	7536.801	888.301	707.301	526.301	345.301	164.301	16.699
113	7574.494	925.994	744.994	563.994	382.994	201.994	20.994
Group Five (A_5)		A_{51} (7463-7825)	A_{52} (7825-8006)	A_{53} (8006-8187)	A_{54} (8187-8368)	A_{55} (8366-8549)	A_{56} (8549-8730)
114	7648.235	86.265	267.265	448.265	629.265	810.265	991.265
115	7651.824	82.676	263.676	444.676	625.676	806.676	987.676
116	7659.021	75.479	256.479	437.479	618.479	799.479	980.479
117	7867.546	133.046	47.954	228.954	409.954	590.954	771.954
118	7885.447	150.947	30.053	211.053	392.053	573.053	754.053
119	7903.543	169.043	11.957	192.957	373.957	554.957	735.957
120	7932.312	197.812	16.812	164.188	345.188	526.188	707.188
121	7993.406	258.906	77.906	103.094	284.094	465.094	646.094
122	7997.077	262.577	81.577	99.423	280.423	461.423	642.423
123	8016.756	282.256	101.256	79.744	260.744	441.744	622.744
124	8029.403	294.903	113.903	67.097	248.097	429.097	610.097
125	8077.851	343.351	162.351	18.649	199.649	380.649	561.649
126	8086.858	352.358	171.358	9.642	190.642	371.642	552.642
127	8106.619	372.119	191.119	10.119	170.881	351.881	532.881
128	8142.616	408.116	227.116	46.116	134.884	315.884	496.884
129	8171.385	436.885	255.885	74.885	106.115	287.115	468.115
130	8218.167	483.667	302.667	121.667	59.333	240.333	421.333
131	8340.356	605.856	424.856	243.856	62.856	118.144	299.144
132	8381.802	647.302	466.302	285.302	104.302	76.698	257.698
133	8415.907	681.407	500.407	319.407	138.407	42.593	223.593
134	8417.686	683.186	502.186	321.186	140.186	40.814	221.814
135	8468.107	733.607	552.607	371.607	190.607	9.607	171.393
136	8518.334	783.834	602.834	421.834	240.834	59.834	121.166
137	8570.535	836.035	655.035	474.035	293.035	112.035	68.965
138	8602.861	868.361	687.361	506.361	325.361	144.361	36.639
139	8613.647	879.147	698.147	517.147	336.147	155.147	25.853
140	8692.755	958.255	777.255	596.255	415.255	234.255	53.255
141	8698.205	963.705	782.705	601.705	420.705	239.705	58.705
Group Six(A_6)		A_{61} (8730-8911)	(8911-9092)	A_{63} (9092-9273)	A_{64} (9273-9454)	A_{65} (9454-9635)	A_{66} (9635-9816)
142	8750.292	70.208	251.208	432.208	613.208	794.208	975.208
143	8809.608	10.892	191.892	372.892	553.892	734.892	915.892
144	8850.941	30.441	150.559	331.559	512.559	693.559	874.559
145	8850.972	30.472	150.528	331.528	512.528	693.528	874.528
146	8894.166	73.666	107.334	288.334	469.334	650.334	831.334
147	9055.909	235.409	54.409	126.591	307.591	488.591	669.591
148	9415.505	595.005	414.005	233.005	52.005	128.995	309.995
149	9464.148	643.648	462.648	281.648	100.648	80.352	261.352
150	9809.32	988.82	807.82	626.82	445.82	264.82	83.82

This method has been applied to every data set for calculating distance value of the all groups. Based on the minimum distance value, the group has been selected and the error with respect to that group has been calculated and stored in memory of pheromone trials and have been furnished in table 3 (column 4 and 10 respectively)

Step 2

If the error of pheromone trials is the less than 2% then middle value of the subgroup has to be considered as estimated value of data otherwise if the estimated error is not less than 2% and positive then error as calculated has to be subtracted from the earlier estimated value. If the estimated error is not less than 2% and negative then error as calculated has to be added with estimated value. The Global parameter is furnished below.

$$r_{ij(t+1)=(1-\rho)\times} r_{ij}\,(t) + \rho \times \Delta r_{ij}^{gb}(t)$$

Here $r_{ij}\,(t)$ is minimum distance of the particle in subgroup and $\Delta r_{ij}^{gb}(t) = \frac{1}{Gloabal\ minimum\ distance\ of\ the\ data}$ is the global distance of the entire particle in that group.

As for an example, as the estimated error of data element (3331.597) of serial number 1 of the group one of table 3 (third column) is 1.77% (positive), the value of $r_{ij^{(t+1)}}$ will be

$$r_{ij^{(t+1)}} = (1\text{-}0.5) \times 58.903 + 0.5(1/5.23) = 29.5471$$

Here 3331.597 belongs to serial number one and lies in group one which contains the minimum distance value as 5.23 in the group one (according to table 2 serial no. 18 of column A_{16}). Therefore the value of $\Delta r_{ij}^{gb}(t)$ is taken as 1/5.23. The estimated value has been calculated as (available data – the value of $r_{ij^{(t+1)}}$). Therefore the estimated value = 3331.597-29.5471= 3302.05 for serial no 1. The value of ρ is taken as 0.5. The estimated error is -0.89%. This step has to be repeated until the estimated error is less than 2% or number of iteration has been reached as 50 whichever is less.

The instructions as narrated in step 2 has been repeated for other data elements of other groups to calculate the estimated value based on ant colony optimization algorithm. The error analysis has been made to produce estimated error as furnished in table 3 (6 and 12 column). The average error has been found as 0.73%.

Table 3. Estimated Output and Estimated Error Based on ACO

Sl. No	Available Data	Estimated Data(initial)	Estimated Error (initial) (%)	Estimated data(Final)	Estimated Error (%)	Sl. No.	Available Data	Estimated Data (initial)	Estimated Error (initial) (%)	Estimated data(Final)	Estimated Error (%)
1	3331.597	3390.5	1.77	3302.05	-0.89	76	6625.286	6648.5	0.35	6648.5	0.35
2	3734.387	3752.5	0.49	3752.5	0.49	77	6664.84	6648.5	-0.25	6648.5	-0.25
3	3912.366	3933.5	0.54	3933.5	0.54	78	6664.84	6648.5	-0.25	6648.5	-0.25
4	3939.355	3933.5	-0.15	3933.5	-0.15	79	6693.527	6648.5	-0.67	6648.5	-0.67
5	3941.134	3933.5	-0.19	3933.5	-0.19	80	6704.395	6648.5	-0.83	6648.5	-0.83
6	4009.457	3933.5	-1.89	3933.5	-1.89	81	6770.939	6829.5	0.86	6829.5	0.86
7	4066.994	4114.5	1.17	4114.5	1.17	82	6826.615	6829.5	0.04	6829.5	0.04
8	4079.558	4114.5	0.86	4114.5	0.86	83	6833.812	6829.5	-0.06	6829.5	-0.06
9	4104.769	4114.5	0.24	4114.5	0.24	84	6869.809	6829.5	-0.59	6829.5	-0.59
10	4104.769	4114.5	0.24	4114.5	0.24	85	6900.356	6829.5	-1.03	6829.5	-1.03
11	4104.769	4114.5	0.24	4114.5	0.24	86	6912.92	6829.5	-1.21	6829.5	-1.21
12	4135.317	4114.5	-0.50	4114.5	-0.50	87	6941.689	7010.5	0.99	7010.5	0.99
13	4135.317	4114.5	-0.50	4114.5	-0.50	88	6979.464	7010.5	0.44	7010.5	0.44
14	4146.102	4114.5	-0.76	4114.5	-0.76	89	6997.447	7010.5	0.19	7010.5	0.19
15	4174.871	4114.5	-1.45	4114.5	-1.45	90	7008.233	7010.5	0.03	7010.5	0.03
16	4203.639	4114.5	-2.12	4143.55	-1.43	91	7037.001	7010.5	-0.38	7010.5	-0.38
17	4243.193	4295.5	1.23	4295.5	1.23	92	7038.78	7010.5	-0.40	7010.5	-0.40
18	4300.73	4295.5	-0.12	4295.5	-0.12	93	7038.78	7010.5	-0.40	7010.5	-0.40
19	4318.713	4295.5	-0.54	4295.5	-0.54	94	7175.344	7191.5	0.23	7191.5	0.23
20	4340.285	4295.5	-1.03	4295.5	-1.03	95	7227.544	7191.5	-0.50	7191.5	-0.50
21	4397.822	4476.5	1.79	4476.5	1.79	96	7231.102	7191.5	-0.55	7191.5	-0.55
22	4426.59	4476.5	1.13	4476.5	1.13	97	7232.962	7191.5	-0.57	7191.5	-0.57
23	4466.144	4476.5	0.23	4476.5	0.23	98	7245.445	7191.5	-0.74	7191.5	-0.74
24	4466.144	4476.5	0.23	4476.5	0.23	99	7254.534	7191.5	-0.87	7191.5	-0.87
25	4467.923	4476.5	0.19	4476.5	0.19	100	7259.87	7191.5	-0.94	7191.5	-0.94

26	4494.913	4476.5	-0.41	4476.5	-0.41	101	7272.517	7191.5	-1.11	7191.5	-1.11
27	4496.692	4476.5	-0.45	4476.5	-0.45	102	7299.506	7372.5	1.00	7372.5	1.00
28	4534.467	4476.5	-1.28	4476.5	-1.28	103	7331.863	7372.5	0.55	7372.5	0.55
29	4563.236	4476.5	-1.90	4476.5	-1.90	104	7409.162	7372.5	-0.49	7372.5	-0.49
30	4565.014	4476.5	-1.94	4476.5	-1.94	105	7437.849	7372.5	-0.88	7372.5	-0.88
31	4592.004	4657.5	1.43	4657.5	1.43	106	7437.849	7372.5	-0.88	7372.5	-0.88
32	4602.79	4657.5	1.19	4657.5	1.19	107	7454.053	7372.5	-1.09	7372.5	-1.09
33	4606.347	4657.5	1.11	4657.5	1.11	108	7468.396	7553.5	1.14	7553.5	1.14
34	4674.67	4657.5	-0.37	4657.5	-0.37	109	7497.246	7553.5	0.75	7553.5	0.75
35	4716.003	4657.5	-1.24	4657.5	-1.24	110	7520.597	7553.5	0.44	7553.5	0.44
36	4739.435	4657.5	-1.73	4657.5	-1.73	111	7522.375	7553.5	0.41	7553.5	0.41
37	4757.418	4838.5	1.70	4838.5	1.70	112	7536.801	7553.5	0.22	7553.5	0.22
38	4757.418	4838.5	1.70	4838.5	1.70	113	7574.494	7553.5	-0.28	7553.5	-0.28
39	4843.723	4838.5	-0.11	4838.5	-0.11	114	7648.235	7734.5	1.13	7734.5	1.13
40	4897.621	4838.5	-1.21	4838.5	-1.21	115	7651.824	7734.5	1.08	7734.5	1.08
41	4911.964	4838.5	-1.50	4838.5	-1.50	116	7659.021	7734.5	0.99	7734.5	0.99
42	4922.832	4838.5	-1.71	4838.5	-1.71	117	7867.546	7915.5	0.61	7915.5	0.61
43	4924.611	4838.5	-1.75	4838.5	-1.75	118	7885.447	7915.5	0.38	7915.5	0.38
44	4955.158	5019.5	1.30	5019.5	1.30	119	7903.543	7915.5	0.15	7915.5	0.15
45	5034.134	5019.5	-0.29	5019.5	-0.29	120	7932.312	7915.5	-0.21	7915.5	-0.21
46	5061.256	5019.5	-0.83	5019.5	-0.83	121	7993.406	7915.5	-0.97	7915.5	-0.97
47	5075.681	5019.5	-1.11	5019.5	-1.11	122	7997.077	7915.5	-1.02	7915.5	-1.02
48	5174.551	5200.5	0.50	5200.5	0.50	123	8016.756	8096.5	0.99	8096.5	0.99
49	5188.762	5200.5	0.23	5200.5	0.23	124	8029.403	8096.5	0.84	8096.5	0.84
50	5228.449	5200.5	-0.53	5200.5	-0.53	125	8077.851	8096.5	0.23	8096.5	0.23
51	5248.078	5200.5	-0.91	5200.5	-0.91	126	8086.858	8096.5	0.12	8096.5	0.12
52	5268.003	5200.5	-1.28	5200.5	-1.28	127	8106.619	8096.5	-0.12	8096.5	-0.12
53	5285.853	5200.5	-1.61	5200.5	-1.61	128	8142.616	8096.5	-0.57	8096.5	-0.57
54	5559.144	5562.5	0.06	5562.5	0.06	129	8171.385	8096.5	-0.92	8096.5	-0.92
55	5569.929	5562.5	-0.13	5562.5	-0.13	130	8218.167	8277.5	0.72	8277.5	0.72
56	5577.259	5562.5	-0.26	5562.5	-0.26	131	8340.356	8277.5	-0.75	8277.5	-0.75
57	5616.681	5562.5	-0.96	5562.5	-0.96	132	8381.802	8458.5	0.92	8458.5	0.92
58	5697.568	5743.5	0.81	5743.5	0.81	133	8415.907	8458.5	0.51	8458.5	0.51
59	5862.982	5924.5	1.05	5924.5	1.05	134	8417.686	8458.5	0.48	8458.5	0.48
60	5877.325	5924.5	0.80	5924.5	0.80	135	8468.107	8458.5	-0.11	8458.5	-0.11
61	6071.508	6105.5	0.56	6105.5	0.56	136	8518.334	8639.5	1.42	8639.5	1.42
62	6098.497	6105.5	0.11	6105.5	0.11	137	8570.535	8639.5	0.80	8639.5	0.80
63	6123.708	6105.5	-0.30	6105.5	-0.30	138	8602.861	8639.5	0.43	8639.5	0.43
64	6127.266	6105.5	-0.36	6105.5	-0.36	139	8613.647	8639.5	0.30	8639.5	0.30
65	6150.616	6105.5	-0.73	6105.5	-0.73	140	8692.755	8639.5	-0.61	8639.5	-0.61
66	6186.582	6105.5	-1.31	6105.5	-1.31	141	8698.205	8639.5	-0.67	8639.5	-0.67
67	6262.05	6286.5	0.39	6286.5	0.39	142	8750.292	8820.5	0.80	8820.5	0.80
68	6267.469	6286.5	0.30	6286.5	0.30	143	8809.608	8820.5	0.12	8820.5	0.12
69	6362.781	6286.5	-1.20	6286.5	-1.20	144	8850.941	8820.5	-0.34	8820.5	-0.34
70	6371.788	6286.5	-1.34	6286.5	-1.34	145	8850.972	8820.5	-0.34	8820.5	-0.34
71	6402.335	6467.5	1.02	6467.5	1.02	146	8894.166	8820.5	-0.83	8820.5	-0.83
72	6431.104	6467.5	0.57	6467.5	0.57	147	9055.909	9001.5	-0.60	9001.5	-0.60
73	6596.518	6648.5	0.79	6648.5	0.79	148	9415.505	9363.5	-0.55	9363.5	-0.55
74	6603.664	6648.5	0.68	6648.5	0.68	149	9464.148	9363.5	-1.06	9363.5	-1.06
75	6623.507	6648.5	0.38	6648.5	0.38	150	9809.32	9725.5	-0.85	9725.5	-0.85

3.5 Statistical Methods and Soft computing models

The statistical method using least square technique (LSQ) based on linear, exponential, asymptotic, curvilinear and logarithmic equations have been applied on the cumulative antecedent item to produce estimated cumulative antecedent item. The error analysis has been made to compute estimated error and average error. The estimated data based on statistical methods and the estimated errors have been furnished in table 7. It has been observed that the least square technique based on logarithmic equation has given minimum error as compared to other statistical models. The problem of using this model is that if original data (actual data) is not stabilized, some noise may enter with the original data, and as a result that may generate some erroneous results. This sort of unstabilized data can be avoided by using soft computing models.

Under soft computing domain, the harmony search, fuzzy logic, neural network, particle swarm optimization, artificial bee colony algorithm have been used to produce estimated data based on that models.

3.6 Review of models

The statistical methods, fuzzy logic, neural network, particle swarm optimization, artificial bee colony algorithm have been used to produce estimated data based on those models. The error analysis and the parameters of residual analysis have been applied on the estimated data with respect to the available data. The error analysis includes the computation of average error and residual analysis includes the computation of sum of absolute residual, mean of absolute residual, mean of mean of absolute residual, median of absolute residual, maximum of absolute residual and standard deviation of absolute residual. The average error and the parameters of residual analysis have been furnished in table 4. The strength of each model can be ascertained by the error and residual

analysis. The less value of the parameter for the error and residual analysis indicates more strength of the model (method) than others. From table 4 it has been observed that out of seven cases, ACO has been preferred by seven cases. Therefore the ACO has been selected as the preferable optimizing model for the estimation of data. Therefore, the estimated data based on ACO can be used for the extraction of knowledge. The estimated data based on the ACO has been named as intermediate data and that name will be used in further processing.

Table 4
Average Error and Residual Analysis of Applied Methods on Iris Flowers Data

Serial Number	Method Name	Sum of Absolute Residual	Maximum Absolute Residual	Mean Absolute Residual	Mean of Mean Absolute Residual	Median of Absolute Residual	Standard Deviation of Absolute residual	Average Error (%)
1	Linear Equation	20090	710.349	3.375	0.023	286.376	94.90468	2.25
2	Exponential equation	31627.33	618.573	4.907	0.033	495.577	144.1412	3.27
3	Asymptotic Equation	80269.16	2191.526	15.404	0.103	309.903	459.805	10.27
4	Curvilinear Equation	20066.95	716.8084	3.382	0.023	282.7374	95.0663	2.25
5	Logarithm Equation	27677.26	998.64	4.762	0.032	71.827	143.9862	3.17
6	Fuzzy Logic	14493.82	420.903	2.429	0.016	65.507	64.14068	1.62
7	Neural Network	6811.231	191.861	1.104	0.007	24.993	28.12644	0.78
8	PSO	8917.629	324.073	1.529	0.01	63.043	53.0546	1.02
9	ACO	6728.986	121.166	1.098	0.007	24.993	26.82716	0.73

3.7 Creation of Knowledge

Certain unknown data have been taken for the extraction of knowledge for that data item and that have been furnished in table 5. The preprocessing technique namely decimal scaling techniques have been applied on unknown data sets to convert it into proper format for the extraction of knowledge. The unknown data have been termed as tested data. The purpose of this work is to assess the quality of flowers (unknown data) based on the information A, B, C. The proposed work is to estimate the value D based on these primary values i.e. A, B, C using the proposed model. Thereafter an effort has to be made the quality of information which has been obtained (estimated value of D) by calculating the error of the estimated value of D with respect to actual value of D. The items sepal width, petal length petal width and sepal width have been termed as A, B, C and D respectively.

Table 5
Unknown Data

Serial Number	A	B	C	D	Serial Number	A	B	C	D
1	2.65	1.2	0.2	4.4	31	3	4.55	1.45	5.85
2	2.95	1.25	0.15	4.35	32	2.6	5.05	1.75	6.15
3	3.45	1.55	0.2	5	33	2.7	5	1.7	6.15
4	3.4	1.6	0.3	5.4	34	2.95	4.75	1.65	6.45
5	3.35	1.65	0.45	5.05	35	3.15	4.7	1.45	6.85
6	3.35	1.8	0.35	4.95	36	2.9	5.15	1.4	6.55
7	1.9	3.35	1.1	3.85	37	2.85	5.25	1.45	6.5
8	3.85	1.4	0.35	5.25	38	3.1	4.7	1.8	5.8
9	2.95	2.4	0.7	5.2	39	3.2	4.75	1.65	6.35
10	4.15	1.45	0.15	5.35	40	3	5	1.75	6.35
11	3.25	2.35	0.6	5.25	41	2.95	5.05	1.85	6.15
12	3.15	2.45	0.75	5.1	42	2.95	5.55	1.8	6.4
13	3.1	2.6	0.7	5	43	3.1	5.3	1.9	6.5
14	2.4	3.5	1	5.2	44	3.1	5.45	1.8	6.85
15	2.3	3.85	1	5.75	45	2.9	5.7	1.85	6.8
16	3.3	2.75	0.7	5.85	46	2.95	5.55	1.95	6.4
17	3.5	2.5	0.7	5.7	47	3.05	5.35	2.05	6.55
18	2.5	3.65	1.05	5.6	48	2.9	5.4	2.25	6.55
19	2.7	4.2	1.25	5.8	49	2.95	5.35	2.25	6.65
20	2.65	4.3	1.25	5.55	50	3	5.7	2.1	6.9
21	2.55	4.6	1.4	5.85	51	3.05	5.8	2.05	6.85
22	2.6	4.6	1.35	5.85	52	2.95	6.1	1.95	7.2
23	3	4.15	1.25	5.65	53	3	5.85	2.15	6.8
24	2.95	4.2	1.3	6	54	3.35	5.55	2.2	6.45
25	2.9	4.3	1.3	6.3	55	3.3	5.55	2.3	6.55
26	2.7	4.4	1.5	5.55	56	3.1	5.9	2.3	7.3
27	2.65	4.5	1.5	5.3	57	3.15	6.3	2.3	6.95
28	2.75	4.55	1.5	6.1	58	3.45	6.05	2.5	6.75
29	2.95	4.4	1.4	6.25	59	3.7	6.25	2.25	7.55
30	2.95	4.65	1.4	6.1	60	3.8	6.55	2.1	7.8

Step 2

Now, it is necessary to calculate cumulative antecedent item using the relation as formed in section of factor analysis. The relation has been furnished as

Total effect value = (0.970912) × A + (0.683227) × B + (0.701015) × C.

Step 3

Now using the relation, the total effect value has been computed and furnished in table 6.

Table 6
Total Effect Value

Serial No.	A	B	C	Total Effect Value	Serial No.	A	B	C	Total Effect Value
1	2650	1200	200	3532.992	31	3000	4550	1450	7037.891
2	2950	1250	150	3823.377	32	2600	5050	1750	7201.444
3	3450	1550	200	4548.852	33	2700	5000	1700	7229.323
4	3400	1600	300	4604.569	34	2950	4750	1650	7266.194
5	3350	1650	450	4695.337	35	3150	4700	1450	7286.012
6	3350	1800	350	4727.719	36	2900	5150	1400	7315.685
7	1900	3350	1100	4904.66	37	2850	5250	1450	7370.513
8	3850	1400	350	4939.885	38	3100	4700	1800	7482.821
9	2950	2400	700	4994.646	39	3200	4750	1650	7508.922
10	4150	1450	150	5125.116	40	3000	5000	1750	7555.648
11	3250	2350	600	5181.657	41	2950	5050	1850	7611.365
12	3150	2450	750	5258.041	42	2950	5550	1800	7917.928
13	3100	2600	700	5276.928	43	3100	5300	1900	7962.859
14	2400	3500	1000	5422.499	44	3100	5450	1800	7995.242
15	2300	3850	1000	5564.537	45	2900	5700	1850	8006.917
16	3300	2750	700	5573.594	46	2950	5550	1950	8023.08
17	3500	2500	700	5596.97	47	3050	5350	2050	8053.627
18	2500	3650	1050	5657.125	48	2900	5400	2250	8082.355
19	2700	4200	1250	6367.285	49	2950	5350	2250	8096.739
20	2650	4300	1250	6387.062	50	3000	5700	2100	8279.262
21	2550	4600	1400	6600.091	51	3050	5800	2050	8361.079
22	2600	4600	1350	6613.586	52	2950	6100	1950	8398.855
23	3000	4150	1250	6624.397	53	3000	5850	2150	8416.797
24	2950	4200	1300	6645.063	54	3350	5550	2200	8586.698
25	2900	4300	1300	6664.84	55	3300	5550	2300	8608.254
26	2700	4400	1500	6679.184	56	3100	5900	2300	8653.201
27	2650	4500	1500	6698.961	57	3150	6300	2300	8975.038
28	2750	4550	1500	6830.214	58	3450	6050	2500	9235.707
29	2950	4400	1400	6851.811	59	3700	6250	2250	9439.827
30	2950	4650	1400	7022.617	60	3800	6550	2100	9636.734

Step 4

The cumulative antecedent item for tested data has been furnished in table 6 which has to be processed in subsequent sections. The cumulative antecedent item has been termed as tested antecedent data. The tested antecedent data has been furnished in table 7.

Table 7
Tested Data

Serial Number	Tested Antecedent Data	Serial Number	Tested Antecedent Data	Serial Number	Tested Antecedent Data	Serial Number	Tested Antecedent Data
1	3532.992	16	5573.594	31	7037.891	46	8023.08
2	3823.377	17	5596.97	32	7201.444	47	8053.627
3	4548.852	18	5657.125	33	7229.323	48	8082.355
4	4604.569	19	6367.285	34	7266.194	49	8096.739
5	4695.337	20	6387.062	35	7286.012	50	8279.262
6	4727.719	21	6600.091	36	7315.685	51	8361.079
7	4904.66	22	6613.586	37	7370.513	52	8398.855
8	4939.885	23	6624.397	38	7482.821	53	8416.797
9	4994.646	24	6645.063	39	7508.922	54	8586.698
10	5125.116	25	6664.84	40	7555.648	55	8608.254
11	5181.657	26	6679.184	41	7611.365	56	8653.201
12	5258.041	27	6698.961	42	7917.928	57	8975.038
13	5276.928	28	6830.214	43	7962.859	58	9235.707
14	5422.499	29	6851.811	44	7995.242	59	9439.827
15	5564.537	30	7022.617	45	8006.917	60	9636.734

Step 5

The statistical methods and soft computing models have been applied on the tested antecedent data. The statistical methods include least square technique (LSQ) based on linear, exponential, asymptotic, curvilinear and logarithmic

equations. The soft computing models include fuzzy logic (Fuzzy), neural network (NN), genetic algorithm(GA), particle swarm optimization (PSO), ant colony optimization(ACO. The average errors based on statistical and soft computing models have been furnished in tables 8.

Table 8
Average Error and Residual Analysis using Tested Data (Iris Flowers Data)

Serial Number	Method Name	Sum of Absolute Residual	Maximum Absolute Residual	Mean Absolute Residual	Mean of Mean Absolute Residual	Median of Absolute Residual	Standard Deviation of Absolute residual	Average Error (%)
1	Linear Equation	8978.615	555.933	1.41	0.02	85.876	100.0384	2.34
2	Exponential equation	13170.25	526.431	1.97	0.03	264.337	140.3142	3.29
3	Asymptotic Equation	26623.52	2161.037	4.46	0.07	517.588	346.7657	7.43
4	Curvilinear Equation	8970.36	562.254	1.41	0.02	82.555	100.4892	2.34
5	Logarithm Equation	11435.6	835.499	1.74	0.03	162.808	155.382	2.9
6	Fuzzy Logic	411009.6	9636.734	60	1	7022.617	1470.556	1.43
7	Neural Network	2928.87	180.785	0.63	0.01	12.317	28.8091	0.59
8	PSO	3380.681	253.008	0.54	0.01	4.237	50.62961	0.91
9	ACO	2081.524	92.234	0.32	0.01	12.117	26.11567	0.53

3.8 Extraction of Knowledge

The estimated ACO output of the tested data has been furnished in table 9. The data (serial number 1 of the table 9) is 3571.5 which lie between (serial number 1 and 2 of table 6). Therefore its estimated output of the consequent item will be the average of actual output of the consequent item of table 6(serial number 1 and 2). Similarly output (estimated consequent item) of the all the tested data has been calculated and furnished in table 12.

Table 9
Tested with ACO output on tested data

Tested Antecedent Data	ACO	Tested Antecedent Data	ACO	Tested Antecedent Data	ACO	Tested Antecedent Data	ACO
3532.992	3571.5	5573.594	5562.5	7037.891	7191.5	8023.08	8096.5
3823.377	3843	5596.97	5653	7201.444	7191.5	8053.627	8096.5
4548.852	4476.5	5657.125	6286.5	7229.323	7191.5	8082.355	8096.5
4604.569	4657.5	6367.285	6377	7266.194	7282	8096.739	8277.5
4695.337	4657.5	6387.062	6648.5	7286.012	7372.5	8279.262	8368
4727.719	4862.75	6600.091	6648.5	7315.685	7372.5	8361.079	8458.5
4904.66	5019.5	6613.586	6648.5	7370.513	7553.5	8398.855	8458.5
4939.885	5110	6624.397	6648.5	7482.821	7553.5	8416.797	8639.5
4994.646	5200.5	6645.063	6648.5	7508.922	7553.5	8586.698	8639.5
5125.116	5200.5	6664.84	6648.5	7555.648	7644	8608.254	8639.5
5181.657	5200.5	6679.184	6648.5	7611.365	7915.5	8653.201	8911
5258.041	5381.5	6698.961	6829.5	7917.928	7915.5	8975.038	9182.5
5276.928	5562.5	6830.214	6829.5	7962.859	7915.5	9235.707	9363.5
5422.499	5562.5	6851.811	7010.5	7995.242	8006	9439.827	9544.5
5564.537	5569.5	7022.617	7010.5	8006.917	8096.5	9636.734	9594.5

Table 10
Estimated Error on Tested Data by ACO

Serial Number	Tested Data	Tested data (ACO)	Actual Consequent Item	Estimated Consequent Item	Estimated Error (%)
1	3532.992	3571.5	5573.594	4500	2.27
2	3823.377	3843	5596.97	4300	-1.15
3	4548.852	4476.5	5657.125	5150	3.00
4	4604.569	4657.5	6367.285	5150	-4.63
5	4695.337	4657.5	6387.062	4966.67	-1.65
6	4727.719	4862.75	6600.091	4966.67	0.34
7	4904.66	5019.5	6613.586	5150	33.77
8	4939.885	5110	6624.397	5150	-1.90
9	4994.646	5200.5	6645.063	5000	-3.85
10	5125.116	5200.5	6664.84	5000	-6.54
11	5181.657	5200.5	6679.184	5166.67	-1.59
12	5258.041	5381.5	6698.961	5166.67	1.31
13	5276.928	5562.5	6830.214	5166.67	3.33
14	5422.499	5562.5	6851.811	5166.67	-0.64
15	5564.537	5569.5	7022.617	5680	-1.22
16	5573.594	5562.5	7037.891	5680	-2.91
17	5596.97	5653	7201.444	5680	-0.35
18	5657.125	6286.5	7229.323	5680	1.43
19	6367.285	6377	7266.194	5716.67	-1.44
20	6387.062	6648.5	7286.012	5716.67	3.00
21	6600.091	6648.5	7315.685	5775	-1.28
22	6613.586	6648.5	7370.513	5775	-1.28
23	6624.397	6648.5	7482.821	5775	2.21
24	6645.063	6648.5	7508.922	5775	-3.75
25	6664.84	6648.5	7555.648	5775	-8.33
26	6679.184	6648.5	7611.365	5775	4.05
27	6698.961	6829.5	7917.928	5775	8.96
28	6830.214	6829.5	7962.859	6333.33	3.83
29	6851.811	7010.5	7995.242	6333.33	1.33
30	7022.617	7010.5	8006.917	6100	0.00
31	7037.891	7037.891	7191.5	6100	4.27
32	7201.444	7201.444	7191.5	6300	2.44
33	7229.323	7229.323	7191.5	6300	2.44
34	7266.194	7266.194	7282	6300	-2.33
35	7286.012	7286.012	7372.5	6300	-8.03
36	7315.685	7315.685	7372.5	6300	-3.82
37	7370.513	7370.513	7553.5	6300	-3.08
38	7482.821	7482.821	7553.5	6160	6.21
39	7508.922	7508.922	7553.5	6160	-2.99
40	7555.648	7555.648	7644	6160	-2.99
41	7611.365	7611.365	7915.5	6160	0.16
42	7917.928	7917.928	7915.5	6275	-1.95
43	7962.859	7962.859	7915.5	6275	-3.46
44	7995.242	7995.242	8006	6275	-8.39
45	8006.917	8006.917	8096.5	6625	-2.57
46	8023.08	8023.08	8096.5	6625	3.52
47	8053.627	8053.627	8096.5	6666.67	1.78
48	8082.355	8082.355	8096.5	6666.67	1.78
49	8096.739	8096.739	8277.5	6666.67	0.25
50	8279.262	8279.262	8368	7033.73	1.94
51	8361.079	8361.079	8458.5	6866.67	0.24
52	8398.855	8398.855	8458.5	6866.67	-4.63
53	8416.797	8416.797	8639.5	6866.67	0.98
54	8586.698	8586.698	8639.5	7040	9.15
55	8608.254	8608.254	8639.5	7040	7.48
56	8653.201	8653.201	8911	7040	-3.56
57	8975.038	8975.038	9182.5	6300	-9.35
58	9235.707	9235.707	9363.5	6300	-6.67
59	9439.827	9439.827	9544.5	7550	0.00
60	9636.734	9636.734	9594.5	7700	-1.28

The particular model has to be selected based on maximum number of minimum value of the parameters of residual analysis and average error. The estimated data based on the selected model has to be used to form a knowledge base. It has been observed that ACO has been preferred over all models (table 4) for initial data sets (training data set) using Iris data set. The said concept has also been tallied (table 8) using new data for Iris data set. The modified ACO and other algorithms have been applied on the other data sets like wine data set and Boston city data to produce the estimated data and estimated error. Using Wine data set, it has been observed that out of seven cases modified ACO has been preferred in seven cases (training data table 11) and seven cases in testing data (Table 12). It has been further observed that average error of modified ACO 0.97% (training data) and 0.55% (tested data) as compared to conventional ACO 2.33% (training data) and 2.67% (tested data). Therefore, modified ACO has been considered as preferable optimizer for tested data with an objective of knowledge extraction of tested antecedent data for Wine data. Using Boston city data set, it has been observed that out of seven cases, modified ACO has been preferred by seven cases in both training (Table 13) and testing data (Table 14) respectively. It has been further observed that average error of modified ACO 0.84% (training data) and 1.04% (tested data) as compared to conventional harmony search algorithm 1.97% (training data) and 1.56% (tested data).Therefore the modified ACO has been selected

as the preferable optimizing model for extraction of knowledge for Boston city data. Thereafter for any other new data set, the application of rules as decided by factor analysis and the algorithm of ACO can be applied to from the total effect value (antecedent item) for that.

Table 11
Average Error and Residual Analysis Using Wine Data Set (Training)

Serial Number	Method Name	Sum of Absolute Residual	Maximum Absolute Residual	Mean Absolute Residual	Mean of Mean Absolute Residual	Median of Absolute Residual	Standard Deviation of Absolute residual	Average Error (%)
1	Linear Equation	563239.8	20864.46	11.83	0.07	3715.2	2706.837	6.65
2	Exponential equation	263976.8	16307.86	4.99	0.03	1216.55	1906.527	2.8
3	Asymptotic Equation	916816.2	22028.48	23.31	0.13	6067.63	3349.786	13.1
4	Curvilinear Equation	272310	12907.52	5.82	0.03	363.57	1348.284	3.27
5	Logarithm Equation	500072.4	20067.69	9.84	0.06	3227.99	2545.412	5.53
6	Fuzzy Logic	191872.5	6591.45	3.89	0.02	1043.65	841.5369	2.19
7	Neural Network	90976.54	1019.61	1.97	0.01	17.65	298.4702	1.11
8	GA	77766.62	1019.61	1.6	0.01	17.65	288.5529	0.9
9	PSO	82189.65	1584.68	1.85	0.01	6.18	365.586	1.03
10	ACO	82429.31	1037.39	1.73	0.01	17.65	269.9827	0.97

Table 12
Average Error and Residual Analysis Using Wine Data Set(Tested)

Serial Number	Method Name	Sum of Absolute Residual	Maximum Absolute Residual	Mean Absolute Residual	Mean of Mean Absolute Residual	Median of Absolute Residual	Standard Deviation of Absolute residual	Average Error (%)
1	Linear Equation	242461.8	17217.92	4.764317	0.071109	4009.03	3294.108	7.11092
2	Exponential equation	121704.4	12757.49	2.082021	0.031075	1494.3	2341.779	3.107494
3	Asymptotic Equation	359025	18334.91	9.129402	0.13626	6385.15	3794.556	13.62597
4	Curvilinear Equation	100678.5	9395.47	2.079314	0.031035	69.74	1615.817	3.103454
5	Logarithm Equation	212114.9	16416.51	3.91513	0.058435	3535.42	3061.499	5.843478
6	Fuzzy Logic	79867.98	3048.42	1.541123	0.023002	1029.17	798.3215	2.300183
7	Neural Network	19454.74	996.42	0.386516	0.005769	3.17	233.6207	0.576889
8	GA	19068.94	996.42	0.372862	0.005565	3.17	234.3948	0.55651
9	PSO	35301.4	2398.57	0.7353	0.0110	8.3	526.5675	1.094
10	ACO	19056.6	996.42	0.372102	0.005554	3.17	234.1222	0.555375

Table 13
Average Error and Residual Analysis Using Boston City Data (Training)

Serial Number	Method Name	Sum of Absolute Residual	Maximum Absolute Residual	Mean Absolute Residual	Mean of Mean Absolute Residual	Median of Absolute Residual	Standard Deviation of Absolute residual	Average Error (%)
1	Linear Equation	2608225	30884.83	57.81	0.11	6071.3	4104.223	11.43
2	Exponential equation	4386720	30783.82	78.66	0.16	10190.87	5003.399	15.55
3	Asymptotic Equation	1034558	17429.51	24.25	0.05	674.99	2149.16	4.79
4	Curvilinear Equation	1250565	18522.46	27.79	0.05	118.06	2495.727	5.49
5	Logarithm Equation	1970980	26466.41	45.18	0.09	3671.98	3411.724	8.93
6	Fuzzy Logic	635987.1	2821.05	10.06	0.02	341.66	813.3729	1.99
7	Neural Network	365430.3	1408.05	6	0.01	1071.34	414.4377	1.18
8	GA	330512.2	1590.05	4.84	0.01	1071.34	411.074	0.96
9	PSO	526228.4	3402.18	6.8	0.01	101.41	976.207	1.35
10	ACO	316549.6	5652.02	4.3	0.01	1071.34	493.7084	0.85

Table 14
Average Error and Residual Analysis Using Tested Data on Boston City Data(Tested)

Serial Number	Method Name	Sum of Absolute Residual	Maximum Absolute Residual	Mean Absolute Residual	Mean of Mean Absolute Residual	Median of Absolute Residual	Standard Deviation of Absolute residual	Average Error (%)
1	Linear Equation	1116324	29669.76	36.8	0.24	5493.49	6083.05	24.37
2	Exponential equation	1591262	29546.42	42	0.28	9629.36	5755.885	27.81
3	Asymptotic Equation	404876.6	16309.43	15.18	0.1	3180.23	3246.44	10.06
4	Curvilinear Equation	458333.7	17380.78	16.63	0.11	3770.79	3551.805	11.02
5	Logarithm Equation	841487.6	25258.46	28.89	0.19	6333.96	5108.946	19.13
6	Fuzzy Logic	198607.2	2804.41	4.09	0.03	304.97	840.9338	2.71
7	Neural Network	114148.3	1391.41	2.26	0.01	1108.03	419.2308	1.5
8	GA	100868.2	1391.41	1.77	0.01	1108.03	414.0762	1.17
9	PSO	147805.6	1791.11	1.52	0.04	803.74	932.314	1.51
10	ACO	99462.27	2845.43	1.57	0.01	1108.03	527.8877	1.04

From previous knowledge, the necessary consequent item (output) may be decided, which is the predicted value or inference knowledge gathered from the applied early data set. In case of Iris flower data the quality of flower can be inference from the input of its attributes. The quality of wine and property tax rate to be pays by the customer can also be predicted based on the attribute of Wine and Boston city data sets. Thereafter, if the said knowledge is available in advance, necessary planning work can be decided by the Governments and various other agencies in the country.

4. CONCLUSION

The applications of rules as decided by factor analysis and principal component analysis have been applied to from the total effect value (antecedent) on the Iris flower data. The methods of statistical analysis and soft computing have been applied on the total effect value to select the preferable method. The decision of the preferable model has been decided by maximum number of minimum parameters of average error and the residual analysis. For checking the validity of model, an effort has been made to get relevant information (knowledge) using new data items (tested data). It has been mentioned that modified harmony search algorithm has been preferred over all models for initial data sets (training data set) using Iris data set. The said concept has been tallied using new data (tested data) for Iris data set.

For any new data set, the said models can be applied to from the cumulative antecedent item for that data set and accordingly a relation can be formed between antecedent item and consequent item for the formation of knowledge base. From previous knowledge, the necessary consequent item may be decided, which is the predicted or inference knowledge gathered from the applied new data set.

Knowledge discovery (output extraction) can also be made for different new unknown data set.

Thereafter, if the said information is available in advance, necessary planning work can be decided by the Governments and various other agencies in the country

5. REFERENCES

[1] Thomas STUTZLE and Marco DORIGO "ACO Algorithms for the Traveling Salesman Problem", *John Wiley & Sons*, 1999

[2] Wei Zhao ; Coll. of Inf. Technol., JiLin Agric. Univ., Changchun, China ; Xingsheng Cai ; Ying Lan, "A New Ant Colony Algorithm for Solving Traveling Salesman Problem", Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference, Vol.3, pp. 530-533 23-25 March 2012

[3] Héctor D. Menéndez, Fernando E. B. Otero, David Camacho, "MACOC: A Medoid-Based ACO Clustering Algorithm" 9th International Conference, ANTS 2014, Brussels, Belgium, September 10-12, 2014. Proceedings, pp. 122-133, 2014.

[4] Xiaoyong Liu, "Ant Colony Optimization Algorithm Based on Dynamical Pheromones for Clustering Analysis", International Journal of Hybrid Information Technology, Vol.7, No.2 (2014), pp.29-38

[5] K. Ayari, S. Bouktif, G. Antoniol," Automatic Mutation Test Input Data Generation via Ant Colony," In the Proceedings of the *9th annual conference on Genetic and evolutionary computation [GECCO]*, London, England, July 2007.

[6] D.J. Mala and V. Mohan, "IntelligenTester – Software Test Sequence Optimization Using Graph Based Intelligent Search Agent," In the Proceedings of *International Conference on Computational Intelligence and Multimedia Applications [ICCIMA]*, 2007, pp. 2227.

[7] K. Li, Z. Yang, "Generating Method of Pair-wise Covering Test Data Based on ACO," In the Proceedings of *International Workshop on Educational Technology & International Workshop on Geoscience and Remote Sensing*, 2008, pp. 776-779.

[8] D.J. Mala, M. Kamalapriya, R. Shobhana, V.Mohan, "A NonPheromone based Intelligent Swarm Optimization Technique in Software Test Suite Optimization," In the *Proceedings of IAMA*,2009.

[9] P. R. Srivastava, K. Baby, and G Raghurama, "An Approach of Optimal Path Generation using Ant Colony Optimization," In the Proceedings of *TENCON 2009, IEEE Press*, 2009, pp. 1-6.

[10] K. Li, Z. Zhang, and W. Liu, "Automatic Test Data Generation Based On Ant Colony Optimization," In the Proceedings of *Fifth International Conference on Natural Computation (ICNC)*, IEEE Press, 2009, pp. 216-219.

[11] W. Ding, J. Kou, K. Li, Z. Yang, "An Optimization Method of Test Suite in Regression Test Model," In the Proceedings of the *2009 WRI World Congress on Software Engineering (WCSE)*, IEEE Computer Society Washington, DC, USA, Vol. 4, 2009, pp. 180-183.

[12] M. Chis, "A Survey of the Evolutionary Computation Techniques for Software Engineering," 1st Chapter In Chis, M. (Ed.), *Evolutionary Computation and Optimization Algorithms in Software Engineering: Applications and Techniques*, 2010, pp. 1-12.

[13] R. F. Tavares Neto and M. Godinho Filho, "Literature review regarding Ant Colony Optimization applied to scheduling problems: Guidelines for implementation and directions for future research", Engineering Applications of Artificial Intelligence, Volume 26 Issue 1, pp 150-161. January, 2013

[14] D. P. Singh, J. P. Choudhury and M. De, " A Comparative Study on the performance of Soft Computing models in the domain of Data Mining," *International Journal of Advancements in Computer Science and Information Technology*, Vol. 1, No. 1, pp. 35-49, September, 2011

[15] D. P. Singh, J. P. Choudhury and M. De, "A Comparative Study to Select a Soft Computing Model for Developing the Knowledge Base of Data mining with Association Rule Formation by Factor Analysis", *International Journal of Artificial Intelligence and Knowledge Discovery*, Vol. 3, No. 3, October, 2013

[16] D. P. Singh, J. P. Choudhury and M. De, "A comparative study on the performance of Fuzzy Logic, Bayesian Logic and neural network towards Decision Making"*International Journal of Data Analysis Techniques and Strategies (IJDATS)*, Vol. 4, No. 2, pp. 205-216, April, 2012

[17] D. P. Singh, J. P. Choudhury and M. De, "A Comparative Study to Select a Soft Computing Model for Knowledge Discovery in Data Mining", *International Journal of Artificial Intelligence and Knowledge Discovery*, Vol. 2, No. 2, pp. 6-19, April, 2012.

[18] D. P. Singh, J. P. Choudhury and M. De, "A comparative study on principal component analysis and factor analysis for the formation of association rule in datamining domain", *Proceedings of the 2nd International Conference on Mathematical, Computational and Statistical Sciences (MCSS '14), Gdansk*, Poland, ISBN: 978-960-474-380-3, pp.442-452, May 15-17, 2014, ISI Index

6. AUTHOR PROFILE

Dr. Dharmpal Singh received his Bachelor of Computer Science and Engineering and Master of Computer Science and Engineering from West Bengal University of Technology. He has about eight years of experience in teaching and research. At present, he is with Pailan College of Management and Technology, Kolkata, and West Bengal, India as an Associate Professor. Currently, he had done his Ph. D in University of Kalyani. He has about 23 publications in national and international journals and conference proceedings.

Prof. (Dr.) Jagannibas Paul Choudhury received his Bachelor of Electronics and Tele-Communication Engineering with Honours from Jadavpur University, Kolkata and Masters of Technology from Indian Institute of Technology Kharagpur. He received his PhD (Engineering) from Jadavpur University. He has about 24 years experience in teaching, research and administration. Now, he is with Department of Information Technology, Kalyani Government Engineering College, West Bengal, India as an Associate Professor and the Head in Information Technology. He has about 100 publications in national and international journals and conference proceedings. His research fields are soft computing, data mining, clustering and classification, routing a computer network, etc.

Prof. (Dr.) Mallika De received her BSc in Physics from Calcutta University in 1973 and MSc from Jadavpur University in 1976. She received MTech in Computer Science from Indian Statistical Institute, Calcutta in 1980 and 1985, respectively, and PhD in Engineering from Jadavpur University in 1997. She is currently JIS Group, India; she has 30 years as a faculty. Her research interests are parallel algorithms and architectures, fault-tolerant computing, image processing and soft computing. She has authored of ten refereed journal articles and ten conference papers.