ABSTRACT
Ant Colony Optimization is a population based meta-heuristic that can be used to find approximate solutions to optimization problems. The clustering analysis divides data into groups (clusters) such that similar data objects belong to the same cluster and dissimilar data objects to different clusters. The clustering based on ant’s behavior is used for process optimization. This paper focuses on water treatment process in which the output chemical needs to be at optimum level to get the desired turbidity range in the outlet water. Initially k-means clustering is used to cluster the input chemical values and then refining of these clusters is done by ant based clustering. A single ant can move the data among clusters depending on picking up and dropping probabilities. The performance measures such as Performance Index (PI) and Standard Deviation (%SD) are used to validate the results from k-means clustering and ant based clustering.

1. INTRODUCTION
The ant colony algorithm is a type of swarm intelligence algorithm with the ability to successfully achieve better solution results of complicated combinatorial optimization problems [1]. It takes inspiration from the observations of ant colonies foraging behavior with which ants can find the shortest paths from food sources to their nest. Research on ant system reveals its potential to solve classic combinatorial optimization problems like Travelling Salesman Problem (TSP) [2]. It is a probabilistic technique for solving computational problems in which artificial ants lay pheromone trials while traversing. The ants choose their paths based on probabilities depending on the strengths of pheromone trials that have been previously laid and a distance function. The cluster analysis is a method of clustering a data set into groups of similar individuals [3] [4]. It identifies and classifies objects variables on the basis of the similarity of the characteristics they possess. There are many clustering methods available, and each of them may give a different grouping of a dataset. One such method is an ant colony clustering which gives better clustering of the input data.

2. WATER TREATMENT PROCESS
Water treatment process is used to make water more acceptable for a desired end-use. There are various stages involved for making them more suitable. The inputs are from sources such as domestic, industrial or storm [5].

A. Block Diagram of Water Treatment Process
The three basics steps of water treatment process are mixing, sedimentation and filtration. The various stages of operation are given in the block diagram which is shown in the Figure 1.

The river water gets mixed with Poly Aluminium Chloride (PAC) coagulant which produces less sludge and sludge formed from Poly Aluminium coagulants tends to settle equally well at low and at normal water temperatures [6]. Then the mixed water enters the sedimentation tank where the water is allowed to sit, heavy suspended particles like sand will settle to the bottom over time since they are denser than water. The water can be collected from the top without disturbing the layer of sediment at the bottom. Sometimes the insoluble particles are too small to settle out quickly in sedimentation alone. So, the gravity filtration is carried out. In this process, water containing solid impurities is passed through a porous medium, typically layers of sand and gravel. The force of gravity is used to push the water through the medium. The solids get stuck in the holes, and are thus retained in the porous medium. The water that passes through the bottom of the filter no longer contains those solid impurities.

B. Control Part of Water Treatment Process
The proposed block diagram discusses the control part of water treatment plant. It is shown in the Figure 2. There are five inputs to the controller and the output is PAC [7].

Figure 2. Control Part in Water Treatment Process
The purpose of the operator is to determine the amount of PAC to be added so that the turbidity of the treated water is kept below a certain level. The optimal amount of PAC depends on the properties of turbid water like
- Turbidity of inlet water (TUB1)
• Temperature of water (TE)
• Alkalinity of water (ALK)
• pH of water (pH)
• Turbidity of inlet water (TUB1)
• Turbidity of outlet water (TUB2)

These parameters are the inputs to the controller and are clustered using the clustering techniques to give the optimum PAC values. This optimized PAC gives the desired turbidity range in the outlet water.

3. CLUSTERING TECHNIQUES

The cluster analysis divides data into groups (clusters) such that similar data objects belong to the same cluster and dissimilar data objects to different clusters.

A. K-means Clustering

The k-means clustering partitions the objects into clusters by minimizing sum of the squared distances between the objects and the centroid of the clusters. The algorithm repeatedly computes the current cluster centers and reassigns each data item to the cluster whose center is closest to it. It terminates when no more reassignment take place [8].

B. Algorithm for K-means Clustering

Step1: Choose a number of clusters k
Step2: Initialize cluster centers $C_1, ..., C_k$
  i. Could pick k data points and set cluster centers to these points
  ii. Or could randomly assign points to clusters and take means of clusters
Step3: For each data point, compute the cluster center it is closest to and assign the data point to this cluster.
Step4: Re-compute cluster centers (mean of data points in cluster)
Step5: Stop when there are no new re-assignments.
   The initial cluster centers are normally chosen either sequentially or randomly. The quality of the final clusters is based on these initial seeds. The cluster centers are calculated by the euclidean distances. The euclidean distance is given by the formula,

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)}$$

Here, x and y are the points between which the distances are measured, n is the total number of values.

The k-means clustering does the initial clustering of five sets of data values such as Turbidity of inlet water (TUB1), Temperature of water (TE), Alkalinity of water(ALK), pH of water (pH) and Turbidity of Outlet water (TUB2). The testing datasets are taken and compared with each cluster centroid. The centroid having the nearest distance is taken and output PAC values are calculated accordingly. Then given datasets are assigned to the respective cluster.

C. ACO Based Clustering

The quality of the clusters from k-means can be improved using the ant clustering technique. This ant based algorithm is to refine the clusters obtained and avoids misclustering.

In this method only one ant is used to refine the clusters. The ant is allowed to go for a random walk on the clusters. Whenever it crosses a cluster, it will pick an item from the cluster and drop it into another cluster while moving. The pick and drop probabilities are calculated as given:

$$P_{picking} = \left(\frac{k_i + f}{k_i}\right)^2$$

$$P_{dropping} = \left(\frac{f}{k_i + f}\right)^2$$

Here, f is the entropy value of the clusters calculated before the item was picked and dropped, $k_i$ and $k_f$ are threshold constants (picking up threshold and dropping threshold, respectively). If the dropping probability is lower than the picking then the item is dropped into another cluster and the entropy value is calculated again [9]. This random walk is repeated for N number of times.

D. Algorithm for Ant Based Clustering

Step1: Choose a number of clusters k
Step2: Initialize cluster centers $C_{1...k}$
Step3: For each data point, compute the cluster center it is closest to and assign the data point to this cluster.
Step4: Re-compute the cluster centers
Step5: Stop when there are no new re-assignments.
Step6: Ant based refinement
  i. Input the clusters from improved k-means.
  ii. For i=1 to N do
    a. Let the ant go for a random walk to pick an item
    b. Calculate the pick and drop probability
    c. Decide to drop the item.
    d. Re-calculate the entropy value of check whether the quality is improving or not.
  iii. Repeat

4. RESULTS AND DISCUSSION

The input and output parameters of the water treatment process for the experiment are taken from [10]. The quality of the output can be determined by comparing the clustering outputs with statistical output which are obtained by linear regression. A linear regression model fits a linear function to a set of data points. The form of the function is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$$

Where, y is the target variable, $x_1, x_2 ... x_n$ are the predictor variables, and $\beta_1, \beta_2 ... \beta_n$ are coefficients that multiply the predictor variables. $\beta_0$ is a constant.

The form of linear equation obtained for water treatment process is

$$y = -3.6644 + 0.0092x_1 - 0.1228x_2 + 0.6362x_3 + 0.0078x_4 - 0.0042x_5$$

where, $x_1, x_2, x_3, x_4, x_5$ are the input values of water treatment process. The clustering is done with 19 sets of input values.
The plots between the observed PAC values, statistical, K-means and ant based PAC values are shown in the Figure 3.

![Figure 3. Plots of Observed, Statistical, K-means and Ant based Clustering Output](image)

**Table 1. Input/Output Parameters of Water Treatment Process**

<table>
<thead>
<tr>
<th>Input Parameters (ppm)</th>
<th>Output Parameters (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TUR1</td>
<td>TUR2</td>
</tr>
<tr>
<td>27.0</td>
<td>1.0</td>
</tr>
<tr>
<td>59.0</td>
<td>1.0</td>
</tr>
<tr>
<td>11.0</td>
<td>1.0</td>
</tr>
<tr>
<td>12.0</td>
<td>3.0</td>
</tr>
<tr>
<td>14.0</td>
<td>4.0</td>
</tr>
<tr>
<td>25.0</td>
<td>1.0</td>
</tr>
<tr>
<td>29.0</td>
<td>1.0</td>
</tr>
<tr>
<td>12.0</td>
<td>5.0</td>
</tr>
<tr>
<td>0.0</td>
<td>1.5</td>
</tr>
<tr>
<td>11.0</td>
<td>2.0</td>
</tr>
<tr>
<td>20.0</td>
<td>1.5</td>
</tr>
<tr>
<td>25.8</td>
<td>1.5</td>
</tr>
<tr>
<td>50.8</td>
<td>1.5</td>
</tr>
<tr>
<td>65.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>

**Table 2. Validation Measures**

<table>
<thead>
<tr>
<th>Validation Measures</th>
<th>Statistical Model</th>
<th>K-means Clustering</th>
<th>K-means with ACO Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Index</td>
<td>211.78</td>
<td>129.7</td>
<td>107.11</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>20.61</td>
<td>12.39</td>
<td>9.9</td>
</tr>
</tbody>
</table>

From the validation measures, the standard deviation for ant based clustering is 2.49% lesser than k-means deviation and 10.71% lesser than statistical deviation. Thus the PAC values obtained from k-means with ACO is an optimal one than the values obtained from statistical calculation and k means clustering.

**V. CONCLUSION**

In the water treatment process, the amount of chemicals to be added to obtain the desired quality of water is considered as an important aspect. In this paper, the addition of Poly Aluminium Chloride coagulant at an optimal amount is considered since it produces less sludge and quick settlement of particles. Thus the experimental input datasets are clustered using k means clustering initially. The output PAC values are obtained with respect to the five sets of input values and it is compared with observed PAC values. The output quality depends on the initial clustering of data and its misclustering will give wrong result. So the ant based clustering algorithm refined the clusters from k-means by picking, moving and dropping the values among clusters. The standard deviation is calculated and it is less for ant based refined clusters when compared to k-means and statistical output giving the optimized values of PAC. Thus the clustering quality of k-means is improved. The same algorithm can be used for implementing multiple inputs and multiple outputs processes for future work.

**REFERENCES**