Comparison of Swarm Intelligence Techniques for Improved Information Retrieval System

Priyanka J. Howale
Department of Information Technology
Pimpri Chinchwad College of Engineering Pune, India

Sanketa S. Pradhan
Department of Information Technology
Pimpri Chinchwad College of Engineering Pune, India

Shraddha G. Lohar
Department of Information Technology
Pimpri Chinchwad College of Engineering Pune, India

Mehul D.Redekar
Department of Information Technology
Pimpri Chinchwad College of Engineering Pune, India

Anagha N. Chaudhari
Department of Information Technology
Pimpri Chinchwad College of Engineering Pune, India

ABSTRACT
Optimization is an important and critical step in the data mining process and it has a huge impact on the success of a data mining process. Selecting a set of feature which is optimal for a given task is a problem which plays an important role in a wide variety of context including pattern recognition, adaptive control and machine learning

Clusters are formed of the reduced dataset using Swarm Intelligence Technique algorithms i.e. Particle Swarm Optimization(PSO),Ant Colony Optimization(ACO),Cluster Hypothesis is verified which is the intra cluster distance should be minimum and inter cluster distance should be maximum. Most relevant documents are stored i+n the clusters

An Information Retrieval System is used for retrieval of data from the clusters. When user enters a query from a Graphical User Interface, using Information Retrieval algorithm the document is searched and retrieved from the clusters. It is then given as an output to the user

Keywords
Optimization, Swarm Intelligence Technique, Clusters.

1. INTRODUCTION
For the complex data sets there is a problem in retrieval the necessary information from particular records. As the original datasets are multidimensional in nature, so for retrieving the particular information, datasets need to be multidimensionality reduced.
Hence, for these there are different optimization techniques or algorithms , and with the help of those algorithms the datasets are first reduced and then that datasets are provided as an input to the algorithms i.e Particle Swarm Optimization, Ant Colony Optimization and Then clusters are obtain for information retrieval system[1].
Evolutionary algorithms technique:

- *Swam Intelligence(SI)*

Swarm Intelligence algorithms are -

- *Particle Swarm Optimization(PSO)*
- *Ant Colony Optimization(ACO)*

Using all these algorithms and with the help of comparison between these algorithms there is a retrieval of information from the particular data sets, and as well as development of IR system also takes place.

2. MODULES
Module 1(Study and Preparation of Dataset):
Input Dataset is required for optimization of IR using Evolutionary Algorithm. Here Instead of applying already available dataset for System , preparing new dataset in this module.

Module 2 (Comparison of Swarm Intelligence Technique algorithms)
Comparison of Swarm intelligence techniques algorithms based on computation time, reduction of dataset using optimization.

Algorithms:

- Particle Swarm Optimization
- Ant Colony optimization

Module 3 (Formation of Cluster)
Best Algorithm between these four Algorithms are selected and clusters are form for this technique. Best algorithm is identified on the basis of their performance and accuracy, efficiency, time complexity.

Algorithms:

- Particle Swarm Optimization
- Ant Colony Optimization

Module 4 (Information Retrieval)
Clusters obtain for the Best algorithm are used for information retrieval from system .There are many applications of IR so any kind of information may retrieve

from the system .information retrieval is based on the input given to that System(Input Dataset).

3. SYSTEM ARCHITECTURE

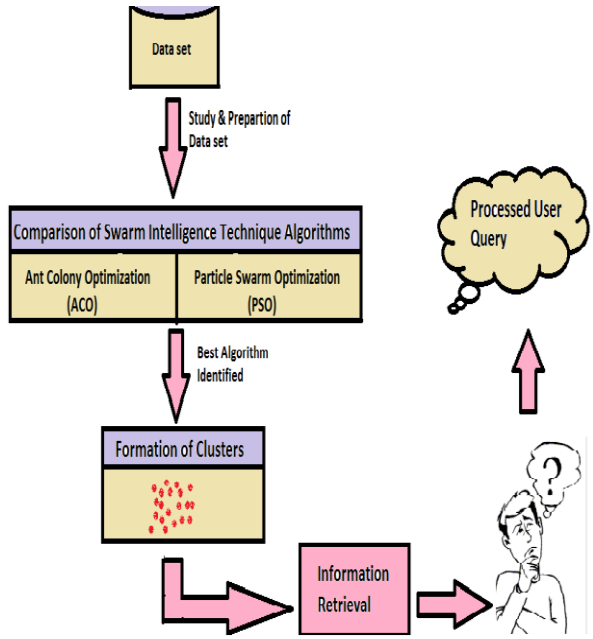

Figure.1: System Architecture

4. ALGORITHMS

4.1 Particle Swarm Optimization(PSO)

Particle Swarm Optimization (PSO) is a global optimization algorithm for dealing with problems in which a best solution can be represented as a point or surface in an n-dimensional space. Hypotheses are plotted in this space and seeded with an initial velocity, as well as a communication channel between the particles. Particles then move through the solution space, and are evaluated according to some fitness criterion after each timestep. Over time, particles are accelerated towards those particles within their communication grouping which have better fitness values. The main advantage of such an approach over other global minimization strategies such as simulated annealing is that the large number of members that make up the particle swarm make the technique impressively resilient to the problem of local minima[1].

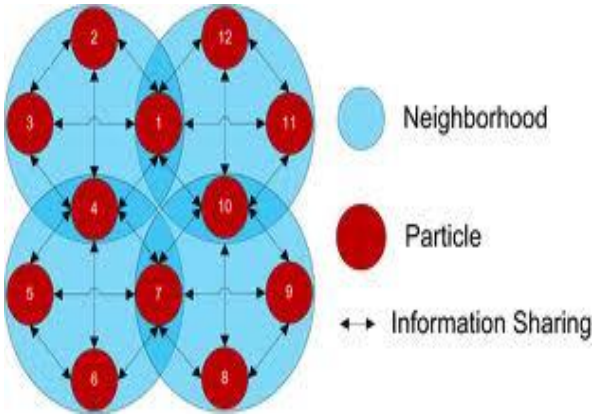

Figure.2: Particle Swarm Optimization

Algorithm: Pseudo code for PSO:

1. Initialize all particles.
2. Calculate fitness value of each particle
3. If the fitness value is better than the best fitness value
   (pBest),then Set pBest = current fitness value
4. If pBest is better than gBest,
   Set gBest = pBest
5. Calculate particle Velocity
   Use gBest and Velocity to update particle Data

4.2 Ant Colony Optimization(ACO):

Ant colony optimization (ACO), introduced by Dorigo in his doctoral dissertation, is a class of optimization algorithms modeled on the actions of an ant colony. ACO is a probabilistic technique useful in problems that deal with finding better paths through graphs. Artificial 'ants' simulation agents locate optimal solutions by moving through a parameter space representing all possible solutions. Natural ants lay down pheromones directing each other to resources while exploring their environment. The simulated 'ants' similarly record their positions and the quality of their solutions, so that in later simulation iterations more ants locate better solutions[1].

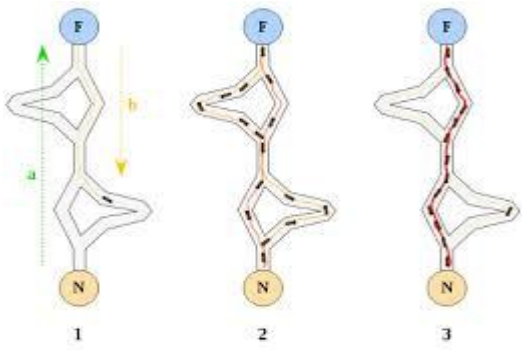

Figure.3: Ant Colony System

Scheme:

1. Construct ant solutions

2. Define attractiveness τ, based on experience from solution

3.Define specific visibility function, η, for a give problem (e.g. distance)

Ant walk

1.Initialize ants and nodes (states)

2.Choose next edge probabilistically according to attractiveness and visibility

3. Each ant maintains a taboo list of infeasible transitions for that iteration

4. Update attractiveness of an edge according to the number of ants that pass through

5.Pheromone update Parameter is called evaporation rate

6.Pheromones = long-term memory of an ant colony

ρ small low → evaporation → slow adaptation

ρ large → high evaporation → fast adaptation

"new pheromone" or $\Delta\tau$ usually contains the base attractiveness constant Q and a factor that you want to optimize

---

Algorithm: General Ant Colony Pseudo Code

Initialize the base attractiveness, τ, and visibility, η, for each edge;
for i < IterationMax do:
for each ant do:
choose probabilistically (based on previous equation)the next state to move into;
add that move to the tabu list for each ant;
repeat until each ant completed a solution;
end
for each ant that completed a solution do:
update attractiveness τ for each edge that the ant traversed;
end;
if (local best solution better than global solution)
save local best solution as global solution;
end;
end;

---

5. EXPERIMENTAL RESULTS

5.1 Study and Preparation of Dataset

Textual dataset is necessary for the comparison of Algorithms following dataset is sound track dataset and it contains list of 200 songtracks.

```
# "#JustDating" (2014) {Pilot (#1.1)}
- "Wait Up (Boots of Danger)"
  Performed by 'Tokyo Police Club' (qv)

# "#LawstinWoods" (2013) {The Incident (#
1.6)}
- "This Is the End (I Don't Want You Back)"
  By Bradley Voorhees
```

Figure.4: Dataset

5.2 Ant Colony Optimization (ACO)

Ant Colony Optimization algorithm is executed successfully.Soundtrack dataset is given as input to this algorithm.output of this algorithm is as follows (Figure.5):
Output shows time required to travel dataset and path of the dataset, which follow by each ant. initialize 5 ants then for each ant, algorithm calculate best length of path and required time. out of that 5 paths ,best trail is identified.
Then weight matrix for dataset is shown in (Figure.5).after that length of best trail is found and time required for best trail.

```
Initializing pheromones on trails
Entering UpdateAnts - UpdatePheromones loop
New best length of 1104.0 found at time 0
New best length of 1086.0 found at time 1
New best length of 1077.0 found at time 13
New best length of 1063.0 found at time 15
Time complete
Best trail found:136 105 165 198 101 191 116 183 20 9 86 193 235 226 148 238 67 35
164 190 119
253 199 239 185 120 167 66 175 200 217 33 215 133 114 3 27 187 24 7 94
49 4 81 54 56 234 93 103 144 210 134 12 156 90 73 17 169 11 39 188
22 168 76 150 184 163 255 72 243 96 104 182 46 249 57 45 154 241 177 179
102 52 18 153 171 28 232 126 257 123 0 176 233 59 1 135 201 197 122 142
244 29 118 58 106 218 224 170 43 31 115 194 88 110 82 228 145 174 246 152
202 117 74 107 87 196 212 252 248 166 98 221 97 26 206 71 180 225 50 213
60 62 80 227 6 70 79 121 230 30 251 15 208 223 211 109 205 13 209 159
10 236 160 113 149 137 65 55 155 5 44 173 112 34 91 128 40 95 99 23
92 83 25 219 189 89 16 141 75 84 132 178 21 138 216 85 143 245 61 125
258 38 220 240 250 64 147 186 130 77 69 111 172 204 63 203 214 247 192 51
129 161 256 53 124 108 140 254 41 2 231 36 42 207 14 78 158 131
Length of best trail found: 1063.0
End Ant Colony Optimization demo
```

Figure.5: Ant Colony Optimization Output

5.3 Particle Swarm Optimization(PSO)

Particle Swarm Optimization is executed Successfully.
Input to the algorithm is Dataset and output is various functions of particle swarm optimization. out of this five functions ,function which give best result that only used for further implementation of algorithm.

1.Hole Function

2.Peak function

3.Sphere Function

4.Step Function

5.3.1 Hole Function

In hole function, PSO considered or arrange all the particle in X and Y coordinates axis according to their target (gbest) values. each particle which arrange in x and y coordinate will moved towards the target value. it covered in 31 iteration according to their no of particle. In above yellow circle denotes the no of particle which moved towards the red target value according to their x and y coordinates.

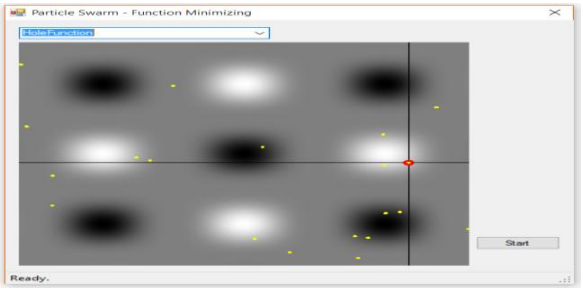

Figure.6.1 : Hole Function

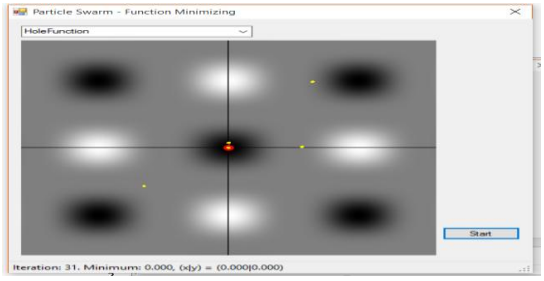

Figure.6.2 : Hole Function

5.3.2 Peak function

Peak Function Centre will stand out as a leader in function venue design.

Peak function of particle will collect the other particles behaviour. Target value set center of x and y coordinates. All the particles moved toward their according to their behaviour and velocity of particle.

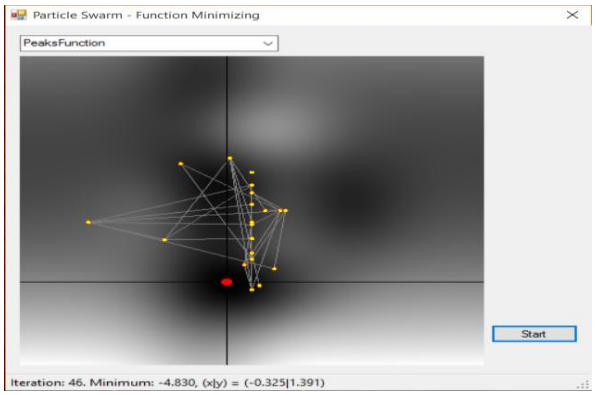

Figure.7 : Peak Function

5.3.3 Sphere Function

Sphere function of particle located all the function are in centered to target value.

Sphere function covered the global optimum. It take to covered global optimum in 11 iteration. This function is simple and strongly convex function.Sphere function is well known benchmark for testing local search ability of optimization algorithm

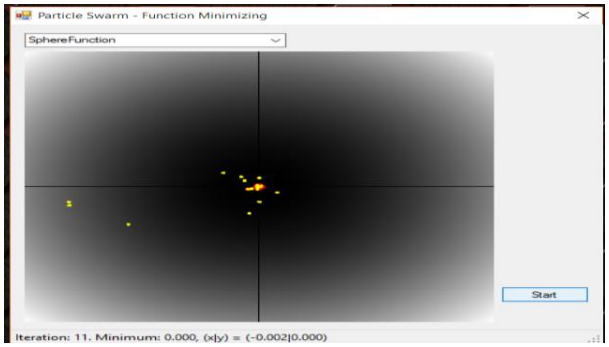

Figure.8 : Sphere Function

5.3.4 Step Function

In step function particle, particle are in gbest value depends on experience of each other.

It covered to complete in iteration 2 according to no of particle.

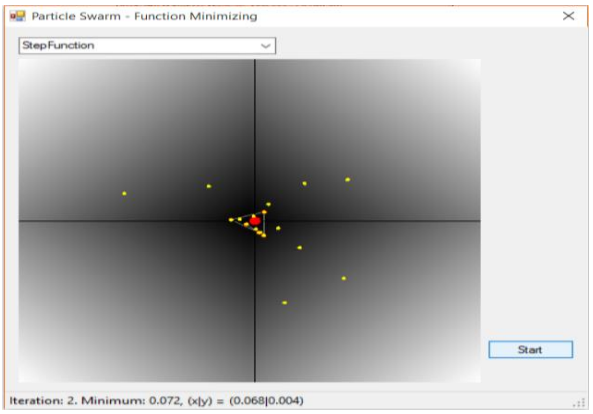

Figure.9: Step Function

6. CONCLUSION

This paper helps to retrieve relevant data with the help of optimization using evolutionary algorithm techniques. i.e Swarm Intelligence Technique. Algorithms applied on input datasets to obtain output in the form of clusters. Clusters are formed from multidimensional datasets where complexities of datasets are reduced using algorithms present under Swarm Intelligence. Comparison between output of these two algorithms(PSO,ACO) in the form of clusters gives best and efficient algorithm for optimization of Information Retrieval as per user query. Paper includes implementation of PSO and ACO algorithms. The future scope for this research would be development of information retrieval system using swarm intelligence technique, hence forming an improved IR system on high dimensional textual datasets.

7. REFERENCES

[1] Sandeep U. Mane Assistance Professor, Dept. of CSE, Pankaj G. Gaikwad.M. Tech Student, Dept. of CSE, "Nature Inspired Techniques for Data Clustering"2014

[2] C. Ahn, J. An, J. Yoo,"Estimation of particle swarm distribution algorithms: combining the benefits of PSO and EDAs",2012

[3] D. Karaboga, B. Akay," A modified artificial bee colony (ABC) algorithm for constrained optimization problems",2011.

[4] Ant Colony Optimization by Marco Dorigo and Thomas Stützle, MIT Press, 2004. ISBN 0-262-04219-3 Particle Swarm Optimization by Maurice Clerc, ISTE, ISBN 1-905209-04-5, 2006

[5] .Karaboga,Dervis"Artificial bee colony algorithm". *Scholarpedia* ,2010.

[6] A.E. Rizzoli, R. Montemanni, E. Lucibello, L.M. Gambardella,"Ant colony optimization for real-world vehicle routing problems",2007

[7] X. Zhang, L. Tang,"A new hybrid ant colony optimization algorithm for the vehicle routing problem",2009

[8] S. Mazzeo, I. Loiseau,"An ant colony algorithm for the capacitated vehicle routing problem",2004