A Review of Clustering Algorithm Based On Swarm Intelligence

RAJNI MISHRA
M.Tech, Department of CSE
TIT BHOPAL, India
Tomar.ranu1989@gmail.com

BHUPESH GAUR
Head of Department of CSE
TIT BHOPAL, India

ABSTRACT

Clustering algorithm is powerful tool in data analysis. The process of clustering done by the selection of similar features and create group. The selection of features and creation of group based on iteration process. Due to iteration sometimes loss the meaning full data and not generate accurate cluster group. For the generation of accurate and reeducation of data loss various authors used swarm intelligence algorithm along with clustering algorithm. the swarm intelligence gives various algorithm such as ant colony optimization, particle of swarm optimization, glow swarm algorithm beep algorithm and many more algorithm. in this paper presents the review of clustering technique along with swarm intelligence. The swarm intelligence derived one algorithm is called KANTS. The KANTS algorithm used for medical data analysis, but it faced a problem of proper feature selection and seed selection.

Keywords: - Clustering Algorithm, KANTS, and Swarm Intelligence.

INTRODUCTION

Swarm intelligence is a revolutionary technique for solving optimization problems that formerly took its inspiration from the biological examples that can be observed in nature, such as ant colonies, flocks of birds, fish schools and bee hives, where a number of individuals with limited capabilities are able to come to intelligent solutions for complex problems [9]. Swarm art is a type of generative art, a term used to classify artistic creations that, with a varying degree of human intervention, are generated by artificial intelligence systems or other computational models [6]. Swarm Intelligence (SI) is a computational Paradigm that arose as a result of the developments in complexity studies. Typically, SI systems rely on a population of simple entities that interact with each other and/or with the environment by means of simple rules. The rules lead the system to a state in which global patterns emerge.[2] This process is usually known as self-organization. swarm intelligence based algorithms provide new ways to solve complex and difficult optimization problems, new ways to manage and control traffic and communications networks, accomplished and effective ways to generate realistic simulations of swarms, and last, but not least, strategies and ideas for organizing goal-oriented groups of robots (so-called “swarm-bots”) [6]. The idea of using KANTS as an art tool is contextualized within the swarm art research and artistic field. KANTS is an SI clustering algorithm based on ant colonies and stigmergy. The term stigmergy describes a species ability to communicate via environment, together with procedures that modify that same environment. In KANTS, the ants are the input data vectors and communicate via a steroidal square grid of vectors with the same dimension of the input vectors. The ants move on the grid and update it the grid vectors are adjusted toward the input vectors (ants) that visit them. In addition, ants tend to move toward regions of the grid with similar vectors. When properly tuned, KANTS guides the swarm to a self-organized state in which complex patterns of global behavior emerge [7].

Figure 1: Main Principles of Collective Behavior of Swarms

Cluster analysis, or clustering, refers to a set of mathematical techniques for sorting observed data into
groups so as to maximize the similarity of observations within the same group and minimize the similarity of observations across different groups. These techniques can be used to discover associations and structures within a data set that may not have been known. Cluster analysis has been widely used in the biological and social sciences to help define classification schemes or taxonomies. It has also been used to suggest new ways of describing a population in business and marketing applications. Clustering is a division of data into groups of similar objects. Each group, called cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups. In other words, the goal of a good document clustering scheme is to minimize intra-cluster distances between documents, while maximizing inter-cluster distances (using an appropriate distance measure between documents). A distance measure (or, dually, similarity measure) thus lies at the heart of document clustering [23]. Clustering is the important step for many errands in machine learning. Section-II gives the information swarm intelligence. In section III discuss about related work in the area of swarm intelligence and clustering algorithm. In section IV discuss the problem of current scenario of clustering and finally discuss in section V conclusion and future scope.

II. Swarm Intelligence

In this section discuss the swarm intelligence technique used for the process of algorithm optimization such as clustering, classification and data regression. Basically swarm optimization algorithm derived from the biological animals and kits. The behavior of kits and animal are very inspiring behavior during the selection of path and foods. It also gives the information about survival factor of process. Here discuss two swarm algorithm one is ant colony system and other is particle of swarm optimization.

ANT COLONY SYSTEM

ACO is an optimization algorithm inspired in the collective foraging behavior of ants to find and exploit the food source that is nearest to the nest. The design of an ACO algorithm implies the specification of the following aspects:

- A proper definition of the problem that needs to be solved by the ants so that some strategies can be made to incrementally build solution to the problem based on some transition rule, amount of pheromone on the path and some local information.
- A method to construct valid solutions that can be considered legally permissible in the real world situation.[3]
- A heuristic function that should be designed based on the problem to measure the significance of the item/terms that can be added to the current solution so as to give right direction to the process of finding optimal solution.

- A rule for updating the pheromones on certain paths that leads to good solution to the problem and fine-tune the pheromone-trail.
- A transition rule that is based on the heuristic function and amount of pheromone on a path to decide about the movement of ants and construction of solution.

ARTIFICIAL ANTS

Artificial ants are characterized as agents that imitate the behavior of real ants. However it should be noted that an artificial ant system has some differences in comparison with real ants which are as follows:

- Artificial ants have memory.
- They are not completely blind.
- They follow a discrete time system.

The Ant System works in two major steps:

- Construction of the solution to the problem under consideration.
- Updating the pheromone trails which may increase or decrease the amount of pheromone on certain paths.

Particle of swarm optimization

Particle of swarm optimization is dynamic population based optimization technique. The dynamic population based selection process of feature attribute of network traffic data. The network traffic data categories into different section of number of particle. The number of particle process as the number of attribute are distributed along with range of path. Some steps are

Step 1: the process of feature attribute in range and distribute and define velocity of particle.

\[ v_i = v_i + c_1 R_1 (p_i, \text{best} - p_i) + c_2 R_2 (g_i, \text{best} - p_i) \]

Step 2: the process of velocity update of particle according to their iteration of each particle agent set.

\[ p_i = p_i + v_i \]

where \( p_i \) and \( v_i \) are the position and velocity of particle \( i \), respectively; \( p_i, \text{best} \) and \( g_i, \text{best} \) is the position with the \( \text{best} \) objective value found so far by particle \( i \) and the entire population respectively; \( w \) is a parameter controlling the dynamics of flying; \( R_1 \) and \( R_2 \) are random variables in the range \([0,1]\); \( c_1 \) and \( c_2 \) are factors controlling the related weighting of corresponding terms. The random variables support the PSO with the ability of stochastic searching.

Step 3: Position updating \( \square \) The positions of all particles are updated according to,

\[ p_i = p_i + v_i \]

After updating, pi should be tested and limited to the allowed range.
III RELATED WORK

This section gives an extensive literature survey on the existing swarm intelligence and ant colony optimization methods in the field of data mining specially for clustering technique. They study various research and journal paper related to swarm intelligence and ant colony optimization along with data mining method.

Carlos M. Fernandes, Antonio M. Mora, Juan J. Merelo, Agostinho C. Rosa Et al. [1] In this paper authors discuss a simplified version of KANTS and describes recent experiments with the algorithm in the context of a contemporary artistic and scientific trend called swarm art, a type of generative art in which swarm intelligence systems are used to create artwork or ornamental objects. KANTS is used here for generating color drawings from the input data that represent real-world phenomena, such as electroencephalogram sleep data.

Manju ,Chander Kant Et al. [2] Describes in this paper the idea of Ant Colonies is presented with brief introduction to its applications in different areas of problem solving in computer science. Artificial Swarm/Ant foraging utilizes various forms of indirect communication, involving the implicit transfer of information from agent to agent through modification of the environment. Using this approach, one can design efficient searching methods that can find solution to complex optimization problems.

O.A. Mohamed Jafar , R. Sivakumar Et al. [3] Discuss on In this paper, a brief survey on ant-based clustering algorithms is described. They also present some applications of ant-based clustering algorithms. Most promising among them are swarm intelligence algorithms. Clustering with swarm-based algorithms is emerging as an alternative to more conventional clustering techniques.

O.A. Mohamed Jafar , R. Sivakumar Et al. [3] Discuss on In this paper, a brief survey on ant-based clustering algorithms is described. They also present some applications of ant-based clustering algorithms. Most promising among them are swarm intelligence algorithms. Clustering with swarm-based algorithms is emerging as an alternative to more conventional clustering techniques.

David A. Van Veldhuizen, Gary B. Lamont Et. al. [5] Discuss here a multi objective optimization problems and certain related concepts, present an MOEA classification scheme, and evaluate the variety of contemporary MOEAs. Current MOEA theoretical developments are evaluated; specific topics addressed include fitness functions, Pareto ranking, niching, fitness sharing, mating restriction, and secondary populations.

Eric Bonabeau David Corne , Riccardo Poli Et.al. [6] Authors in this paper discuss , swarm intelligence based algorithms provide new ways to solve complex and difficult optimization problems, new ways to manage and control traffic and communications networks, accomplished and effective ways to generate realistic simulations of swarms, and last, but not least, strategies and ideas for organising goal-oriented groups of robots (so-called 'swarm -bots') for potential future applications ranging through agriculture, flexible manufacturing and space exploration.

Mehdi Neshat ,Ghodrat Sepidnam ,Mehdi Sargolzaei , Adel Najaran Toosi Et. al. [7] Present, a review of AFSA algorithm and describes the evolution of this algorithm along with all improvements, its combination with various methods as well as its applications. This algorithm has many advantages including high convergence speed, flexibility, fault tolerance and high accuracy. AFSA (artificial fish-swarm algorithm) is one of the best methods of optimization among the swarm intelligence algorithms. This algorithm is inspired by the collective movement of the fish and their various social behaviors.

Pradeep Jha, Krishan Kant Lavania, Deepak Dembla Et. al. [8] discuss in this paper how to reduce the execution time of SVC procedure as well as to improve the ability of proposed SVC scheme in dealing with classification problems. The procedure contains a Ant colony optimization (ACO) technique .they have used Ant Colony Optimization (ACO) based data preprocessing step to remove noise points, outliers, and insignificant points from the dataset. Experiments showed reduction in the execution time of SVC procedure without altering the final cluster configurations. Using their proposed method, the classification accuracy of all dataset is better than the SNN-SVC method.
A.M. Mora, L.J. Herrera, P.A. Castillo, J.J. Merelo et al. [9] This paper reports the investigations and experimental procedures conducted for designing an automatic sleep classification tool based solely in the features extracted with wavelets from EEG, EMG and EOG signals, without any visual aid or context-based evaluation. Real data collected from patients was processed and classified by several traditional and bio-inspired heuristics.

Stefan Bornhofen Vincent Gardeux, Andréa Machizaud et al. [10] This paper presents the swarm paradigm in the context of artistic creation, and more particularly explores the interest of enhancing swarm models with dynamics inspired from natural ecosystems. they introduce an energy budget to the agents of a swarm system, and show how mapping the energy level to visual information such as line width or color, combined with mechanisms such as resource chasing and consumption.

S. Dehuri, S.B. Cho et al. [11] Authors In this paper, proposed a multi objective Pareto based particle swarm optimization (MOPPSO) to minimize the architectural complexity and maximize the classification accuracy of a polynomial neural network (PNN). To support this, they provide an extensive review of the literature on multi objective particle swarm optimization and PNN.

IV PROBLEM STATEMENT

For the purpose of medical data clustering various machines learning algorithm are applied, such as clustering, weighted clustering, and regression. Two of the most critical and well generalized problems of medical data are its new evolved feature and concept-drift. Since a medical data is a fast and continuous event, it is assumed to have infinite length. Therefore, it is difficult to store and use all the historical data for training. The most discover alternative is an incremental learning technique. Several incremental learners have been proposed to address this problem [8], [7].

In addition, concept-drift occurs in the medical when the underlying concepts of the medical change over time. A variety of techniques have also been proposed in the literature for addressing concept-drift [2], [6], [7] in data medical clustering. However, there are two other significant characteristics of data multi-categories, such as concept evolution and feature evolution that are ignored by most of the existing techniques. Concept-evolution occurs when new classes evolve in the data. On the category process we found some important problem in cluster oriented medical data clustering. These problems are given below.

1. Medical data clustering suffered from multiple feature evaluation,
2. Selection of number of cluster for multi-level [1].
3. Diversity of feature selection process [12].
4. Boundary value of cluster.
5. Outlier data treat as noise.

V CONCLUSION AND FUTURE WORK

In this paper resents the review of clustering technique along with swarm intelligence. The process of swarm intelligence reduces the number of iteration and data loss. The swarm intelligence give verity of algorithm such as ant colony optimization, particle of swarm optimization and many more kits based algorithm. on the basis of review estimated that all authors used swarm intelligence algorithm for the optimization of cluster features and reduces the number of iteration. But the analysis of medical data faced a problem of uncertain noise. The processing of noise along with these data creates a problem for the generation of cluster. For the better generation of cluster used particle of swarm optimization technique.

REFERENCES

[8] Pradeep Jha, Krishan Kant Lavania, Deepak Dembla
"Ant Colony Optimization Technique: Robust Data Pre-processing Procedure Support Vector Clustering" IJITKM 2013 PP. 128-133.

"Sleeping with Ants, SVMs, ANNs and SOMs".

"From Swarm Art toward Ecosystem Art"