A Superlative Approach for Clustering Techniques—Brief Study

Ms. M. Divyavani
Assistant Professor
Department of Computer Application
Kovai Kalaimagal College of Arts and Science,
Bharathiar University, Coimbatore, Tamil Nadu, INDIA

Ms. T. Hashni
Assistant Professor
Department of Information Technology
P.S.G.R. Krishnammal College for Women,
Bharathiar University, Coimbatore, Tamil Nadu, INDIA

ABSTRACT

Clustering is the one of the foremost technique in the data mining and its applied in various areas such as artificial intelligence, bio-informatics, biology, computer vision, city planning, data mining, data compression, earth quake studies, image analysis, image segmentation, information retrieval, machine learning, marketing, medicine, object recognition, pattern recognition, spatial database analysis, statistics and web mining. Clustering means the act of partitioning an unlabeled dataset into groups of similar objects. The goal of clustering is to group sets of objects into classes such that similar objects are placed in the same cluster while dissimilar objects are in separate clusters. Over the past few years, several different types of biologically inspired algorithms have been proposed in the various domains. The ant-based clustering algorithms have received special attention from the community over the past few years for two main reasons. First, they are particularly suitable to perform exploratory data analysis and, second, they still require much investigation to improve performance, stability, convergence, and other key features that would make such algorithms mature tools for diverse applications. Ant-based clustering is a biologically inspired data clustering technique. These algorithms have recently been shown to produce good results in a wide variety of real-world applications. During the last five years, research on and with the ant-based clustering algorithms has reached a very promising state. In this paper, a brief study on ant-based clustering algorithms is described. We also present some applications of ant-based clustering algorithms.

Keywords: Clustering, Ant Based Clustering Techniques, Applications.

1. INTRODUCTION

Clustering is the classification of similar objects into different groups, or more precisely, the partitioning of data sets into subsets (clusters), so that the data in each subset share some common trait. Clustering is used as a data processing technique in many different areas of application, such as bioinformatics, data mining, image analysis, etc [3]. Cluster analysis identifies and classifies objects individuals or variables on the basis of the similarity of the characteristics they possess. It seeks to minimize within-group variance and maximize between-group variance. Its aim is to establish a set of clusters such that cases within a cluster are more similar to each other than they are to cases in other clusters. Clustering analysis is used especially as preprocess to another data mining application. A variety of clustering algorithms exists. Mostly used clustering approach is fuzzy c-means and is one of the best for implementing the clustering process. But Fuzzy c-means isn’t suitable for the data where are noise points and the imbalance of examples. This paper study on the new clustering algorithm called Ant-based clustering algorithm [18].

Ant-Based Clustering algorithm is one of those research themes which have been inspired by behaviours and activities of real ant colonies in clustering corpses which has a defined method for clustering a collection of similar and dissimilar objects. The ant behavior in gathering food can be similarly applied to the problem of data clustering. There are several previous works which have used to a real ant behavior to solve these kinds of problems. One of the first works conducted by Deneubourg in 1991 was related to the ant-like agents that are allowed to move objects randomly on a two dimensional (2D) grid as clustering. In 1994, Lumer and Faieta applied the clustering algorithm to data analysis and in 1995; they applied their algorithm to exploratory database analysis as an alternative to the characteristic of the clusters. The application of ant inspired techniques shows to be of great promise to the clustering problem. In fact, ant based clustering techniques are competitive with traditional ones [16].

The paper is organized as follows. In the next section, review the cluster analysis and ant based clustering technique. In section 3, the overview other areas in computer science
where ant-based algorithms have found success. The next section 4 the review the field of ant-based clustering techniques in section 5, followed by our conclusions on the innovative technique.

2. CLUSTER ANALYSIS

Cluster analysis is a tool for exploring the structure of data. Clustering is the process of grouping objects into clusters such that the objects from the same clusters are similar and objects from different clusters are dissimilar. The relationship is often expressed as similarity or dissimilarity measurement and is calculated through distance function. Clustering is useful technique for the discovery of data distribution and patterns in the underlying data. It is an unsupervised learning technique. Unsupervised learning is learning from observations and discovery. In this mode of learning, there is no training set or prior knowledge of the classes. The system analyses the given set of data to observe similarities emerging out of the subsets of the data. The outcome is a set of class descriptions, one for each class, discovered in the environment. The basic objective of clustering is to discover both the dense and sparse regions in a data set. In cluster analysis, one does not know what classes or clusters exist and the problem to be solved is to group the given data into meaningful clusters [12].

Cluster analysis is a difficult problem because many factors (such as effective similarity measures, criterion functions, algorithms and initial conditions) come into play in devising a well-tuned clustering technique for a given clustering problem. Moreover, it is well known that no clustering method can adequately handle all sorts of cluster structures (shape, size and density). Sometimes the quality of the clusters that are found can be improved by pre-processing the data. Another common technique is to use post-processing steps to try to fix-up the clusters that have been found. Generally clustering algorithms can be categorized into hierarchical methods, partitioning methods, density-based methods, grid-based methods, and model-based methods[12]. The desirable features of clustering algorithms are scalability, ability to deal with different data types, discovery of clusters with arbitrary shape, able to deal with noise and outliers, insensitive to order of input records, incorporation of user-specified constraints, interpretability and usability, minimal requirements for domain knowledge to determine input parameters.

There exist a large number of clustering algorithms in the literature [14]. No single algorithms are suitable for all types of objects, nor all algorithms appropriate for all problems. Unfortunately, many of the traditional clustering algorithms share a number of drawbacks. Recently, algorithms inspired by nature used for clustering. It is claimed that ant-based clustering algorithms can overcome these drawbacks. These algorithms have advantages in many aspects, such as self-organization, flexibility, robustness, no need of prior information, and decentralization [12].

2.1 Ant-based Clustering Technique

Recently, several biologically-inspired algorithms have been introduced to solve computer problems such as combinatorial optimization, among others. In many advances on the computer sciences have been based on the observation and emulation of processes of the natural world. In bio-inspired artificial intelligence concepts like the swarm intelligence approach, where the behavior of social insects like ants or bees is copied, communication is carried out exclusively through the environment. The ants, bees, termites, and wasps are classified as social insects because they live in colonies [14].

Bio-informatics also focuses into observing how nature solves situations that are similar to engineering problems to face. The study of ant colonies has offered great insight in this aspect. Clustering and sorting behavior of ants has stimulated research in design of new algorithms for data analysis and partitioning. Several species of ants cluster corpses to form a “cemetery”, or sort their larvae into several piles. This behavior is still not fully understood, but a simple model, in which ants move randomly in space and pick up and deposit items on the basis of local information, may account for some of the characteristic features of clustering and sorting in ants (Bonabeau, Dorigo and Theraulaz, 1999). In several species of ants, workers have been reported to form piles of corpses — cemeteries — to clean the nests. Chretien (1996) has performed experiments with the ant Lasius niger to study the organization of cemeteries. Other experiments on the ant Phaidole pallidula are also reported in Deneubourg et al. (1991). Brood sorting was observed by Franks and Sendova-Franks (1992) in the ant Leptothorax unifasciatus. Workers of this species gather the larvae according to their size. Franks and Sendova-Franks (1992) have intensively analyzed the distribution of brood within the brood cluster. Deneubourg et al. (1991) have proposed two closely related models to account for the two above–mentioned phenomena of corpse clustering and larval sorting in ants. General idea is that isolated items should be picked up and dropped at some other location where more items of that type are present [14].

Ant colonies provide a means to formulate some powerful nature-inspired heuristics for solving the clustering problems. Several clustering methods based on ant behavior have been proposed in the literature. This section provides a brief description of these methods. Ant-based clustering algorithms are based upon the brood sorting behavior of ants. Larval sorting and corpse cleaning by ant was first modeled by Deneubourg et al. for accomplishing certain tasks in robotics. Their work was actually focused on clustering objects by using group of real world robots. Their model is known as Basic Model (BM).

This model can be described as follows: The data items are randomly scattered into a two-dimensional grid. Initially, each data object that represents a multi-dimensional pattern is randomly distributed over the 2D space. Each ant moves...
randomly around this grid picking and dropping the data items. The decision to pick-up or drop an item is random but is influenced by the data items in the ant’s immediate neighborhood. The probability of dropping an item is increased if ants are surrounded with similar data in the neighborhood. In contrast, the probability of picking an item is increased if a data item is surrounded by dissimilar data, or when there is no data in its neighborhood. In this way, clustering of the elements on the 2D grid is obtained. The following some of the application described used as ant based clustering methods [14].

3. ANT-BASED CLUSTERING TECHNIQUES USED TO IMPROVING PERFORMANCE

Here described some of the application used to ant-based clustering method and the given promising results.

3.1 Towards Improving Clustering Ants: An Adaptive Ant Clustering Algorithm

The Adaptive Ant Clustering Algorithm (A2CA) was developed by taking further inspiration from biological systems. In particular, A2CA was inspired by the fact that termites, while building their nests, deposit pheromone on soil pellets and this serves as a reinforcement signal to other termites placing more pellets on the same region of the space (Camazine et al., 2001). Another biological observation taken into account while developing A2CA was the fact that ants can sense not only its immediate neighborhood environment, but a broader range that may vary from ant to ant and with time. Therefore, A2CA has two main modifications in relation to SACA: (i) a progressive vision scheme, (ii) the inclusion of pheromone on the grid cells. In addition, we adopt a cooling schedule for the parameter that drives the picking probability ($k_e$) [2].

3.2 Ant-MST: An Ant-Based Minimum Spanning Tree for Gene Expression Data Clustering

The proposed a new ant-based clustering algorithm for document clustering based on the travelling salesperson scenario. In this paper, the presented an approach called Ant-MST for gene expression data clustering based on both ant-based clustering and minimum spanning trees (MST). The ant-based clustering algorithm is firstly used to construct a fully connected network of nodes. Each node represents one gene, and every edge is associated with a certain level of pheromone intensity describing the co-expression level between two genes. Then MST is used to break the linkages in order to generate clusters [7].

The new Minimum Spanning Tree (MST), a concept from the graph theory, is used for representing multidimensional gene expression data. Based on the representation, gene expression data clustering problem is converted to a tree partitioning problem. Advantages of using this method have been described and demonstrated as follows: (1) the simple structure of a tree facilitates efficient implementations of rigorous clustering algorithm; (2) clustering based on MST does not depend on detailed geometric shape of a cluster; (3) inter-data relationship is greatly simplified in MST representation and no essential information for clustering is lost [7].

3.3 A New Clustering Algorithm Based on Hybrid Global Optimization Based on a Dynamical Systems Approach Algorithm

Many methods for local optimization are based on the notion of a direction of a local descent at a given point. A local improvement of a point in hand can be made using this direction. As a rule, modern methods for global optimization do not use directions of global descent for global improvement of the point in hand. From this point of view, global optimization algorithm based on a dynamical systems approach (GOP) is an unusual method. Its structure is similar to that used in local optimization: a new iteration can be obtained as an improvement of the previous one along a certain direction. In contrast with local methods, is a direction of a global descent and for more diversification combined with Tabu search [1].

This algorithm is called hybrid GOP (HGOP). Cluster analysis is one of the attractive data mining techniques that are used in many fields. One popular class of data clustering algorithms is the center based clustering algorithm. K-means is used as a popular clustering method due to its simplicity and high speed in clustering large datasets. However, K-means has two shortcomings: dependency on the initial state and convergence to local optima and global solutions of large problems cannot found with reasonable amount of computation effort. In order to overcome local optima problem lots of studies have been done in clustering [1].

3.4 Antcolony Inspired Clustering in Biomedical Data Processing

This method is based on the MAX-MIN Ant System algorithm [13]. The proposed method works as follows: first a population of random solutions is created. In each step, the population is populated with new solutions and further improved. Then the population is evaluated and only the specified number of solutions is conserved. Then, a certain amount of pheromone is deposited in proportion to the quality of best individuals and pheromone evaporation is performed [9].

3.5 An Improved Probabilistic Ant-based Clustering for Distributed Databases

The use of ant-based clustering for distributed databases was explored in [Chandrasekar et al., 2006]. An algorithm called probabilistic ant based clustering for distributed databases (PACE) was proposed based on user-interaction or queries from the distributed database. The main advantage is that highly probable or most likely keywords from the query can
be further analyzed instead of concentrating on the entire set of data available. It utilizes a commonly heard concept of hit ratio, which it calculates for the user query. Depending on this, a number of zones are formed throughout the database with priorities assigned to them. The sizes of these zones and their logical placement in the database are discussed in that paper. A colony building algorithm of ants is utilized for formation of the clusters with an extensive odor analysis model which determines the number and type of agents or ants surrounding a data item. The results obtained from PACE showed highly efficient retrieval from the final clusters formed [6].

In this method present an improved version of the Probabilistic Ant-based Clustering Algorithm for Distributed Databases (PACE). The most important feature of this algorithm is the formation of numerous zones in different sites based on corresponding user queries to the distributed database. Keywords, extracted out of the queries, are used to assign a range of values according to their corresponding probability of occurrence or hit ratio at each site. The propose the introduction of weights for individual or groups of data items in each zone according to their relevance to the queries alongwith the concept of familial pheromone trails as part of an Ant Odor Identification Model to bias the movements of different types of ants towards the members of their own family. Its performance is compared against PACE and other known clustering algorithms for different evaluation measures and an improvement is shown in terms of convergence speed and quality of solution obtained [6].

3.6 Ant Colony Metaphor in a New Clustering Algorithm

In this method focuses on the behavior of clustering procedures in this new approach called ACA. The proposed algorithm is evaluated on a number of well-known benchmark data sets. Empirical results clearly show that the ant clustering algorithm (ACA) performs well when compared to other techniques. The ant-clustering algorithms are mainly based on versions proposed by Deneubourg, Lumer and Faieta. A number of slight modifications have been introduced that improve the quality of the clustering and, in particular, the spatial separation between clusters on the grid. Recently, Handl and Meyer (2002) extended Lumer and Faieta’s algorithm and proposed an application to classification of Web documents. The model proposed by Handl and Meyer has inspired us to use this idea to classical cluster analysis. The basic idea is to pick-up or drop a data item on the grid. The employed a modified version of the “short–term memory” introduced by Lumer and Faieta (1994). Each ant has a permission to exploit its memory according to the following rules: if an ant is situated at grid cell \( p \) and carries a data item \( i \), it uses its memory to proceed to all remembered positions, one after the other [15].

3.7 Entropy-Based Metrics in Swarm Clustering

In an ant-clustering system, the cluster generator consists of a group of agents. Each agent computes a measure based on its local surroundings, and clusters objects by picking them up, by dropping them, and by moving (with or without objects), with probabilities dependent on the measure. An incremental ant clustering algorithm incorporates the following modules: cluster initialization, dynamic cluster modification, and cluster model maintenance (in a changing grid) [5].

Entropy used in a number of classical approaches to clustering, as a means to drive the clustering process. ENCLUS22, COOLCAT23 and ACODF24 all use entropy as a criterion to drive their algorithms. ENCLUS aims to discover clusters embedded in the subspaces of a high-dimensional data set. It estimates the entropy and determines if it satisfies a goodness criterion: its entropy has to be lower than a pre-defined threshold value. COOLCAT is an incremental algorithm, which places the next object in the cluster where it minimizes the overall expected entropy. Since it has no way to autonomously determine the number of clusters, this must be specified a priori. ACODF uses a metric based on Renyi’s quadratic entropy [25] to estimate the quality of clusters. The authors compare a number of metrics, arguing for their choice primarily based on its low computational complexity. While this is an important issue for global cluster quality metrics, it is less important for the local metrics used in ant algorithms, so we have chosen not to follow their lead in this, but to stay with Shannon entropy [26]. Their algorithm searches for natural boundaries in the data, using the entropy measure to determine regions where the data is diverse, on the assumption that a diverse region is likely to represent a boundary [5].

3.8 Density-Based and Ant-Colony Clustering Algorithm

In this new method presented the algorithm comparison of two alternatives for data grouping, i.e., Ant colony clustering algorithm and Density-based Clustering (DENCELUE). In Ant colony clustering Algorithm, ants are assumed to distribute over two-dimensional board. Whenever they move, they will randomize their pathways in either picking or dropping objects for grouping. They re-randomize their pathways in every movement. In addition, each ant helps to accelerate the speed parameter. But in DENCELUE, the task is divided into two steps. Firstly, grids construct to fine Density function. Secondly, divide real data with density attractors together objects into density areas, the so-called density cluster. After trial with Iris plant data and Synthesis data, it was shown that Ant colony clustering Algorithm gave the test result slightly better than DENCELUE data grouping, hence, depending on the types of trial data [11].

Alexander Hinneburg and Daniel A Keim [3D] suggested DENCELUE from the basis of good development in statistics and pattern recognition, well-known as “Kernel density estimation” which contained two important parameters, i.e., \( s \) and \( x \) While parameter \( s \) would determine impact of near-by spot, and parameter \( x \) would explain how density attractors decrease the number of density-attractors and increase the effectiveness of grouping [11].
3.9 A Queen Lead Ant-based Algorithm for Database Clustering

The main aim of the Queen Lead ant-based (QLA) algorithm is to partition the given data set into clusters of similar objects. The idea of the algorithm is to simulate the way real ants form the colonies, which are queen dominant. The ants that possess similar cuticle odor to that of the queen of a colony will join that colony. Each object of the database is treated as an ant and it follows a common set of rules, which leads to the required partition. A local optimization technique is applied to assist the ants in achieving good partition. The input for the algorithm is just the database to be clustered. It does not need any other additional parameters such as the number of desired clusters or any information about the structure of the database. Each object in the database has a certain number of attributes, and attributes can be of either numeric or symbolic types [8].

The QLA algorithm consists of four main stages. The first stage is the initialization stage in which all the ants are initialized to the attributes of the object they represent. The second stage is the training stage in which the ants learn about their possible colony mates by meeting and learning about other ants. The third stage consists of two phases: cluster formation and the local optimization. In the cluster formation phase, we stimulate meetings between ants so that the ants can decide which colony they belong to, thus giving us the clusters of similar ants. In the second phase, the local optimization step is designed to eliminate any ants that have been prematurely added to a colony even though they should not belong to it. It also checks to see if the queen represents a colony better than any other ant in that colony. The last stage is the final clustering stage in which the free ants, i.e., the ants which do not belong to any colony, are added to the closest colony [8].

3.10 Incremental Clustering Based on Swarm Intelligence

In the dynamic system, our cluster generator consists of a group of agents. Each agent computes the information entropy or pheromone concentration of the area surrounding it, and clusters objects by picking-up, dropping and moving. The incremental clustering algorithm incorporates the following modules: initialize clusters, modify clusters dynamically, maintain cluster model with a changing grid [4].

It presents a clustering model, including initializing clusters, modifying the previously discovered knowledge using the new data without retraining the old data, and maintaining clusters with a changing grid. The algorithm applies information entropy to model behaviors of agents, such as picking-up and dropping objects, and guides agent movement by pheromone in incremental stages. Our method has main advantages over related work in two aspects: fewer parameters are needed to set, and clustering speed is fast. The experiments show that the incremental method resulted in almost the same good quality clusters as the static method, but much faster than the static method; the entropy function led to slightly better quality clusters than non-entropy functions, and the number of parameters was greatly reduced. In a word, in our new clustering model, each agent acts on environmental cues and the behaviors among them are correlated, so dynamic data can adapt the placement of the clusters. We propose a simple approach, and it is applicable to data mining systems which modify databases either periodically or in batches [4].
novel Ants Sleeping Model (ASM), including its complete formal definition and explanation, to solve clustering problems in data mining. In the ASM model, each datum is represented by an agent (artificial ant) with a small amount of basic information about the datum per se and its current position. In the agents’ environment, which is a two-dimensional grid, each agent interacts with its neighbors and exerts influence on others. Consequently, those with similar features form into groups, and those with different features repel each other. Based on the ASM model, we propose an effective formula for computing the fitness and activating probability of the agents [17].

A analyze the effect of each parameter, propose several efficient agents moving strategies, and design an adaptive ant clustering algorithm A*C. In A*C, the ants can form into high-quality clusters by making simple movements according to little local information from its neighborhood. Compared with the BM and LF algorithms, the ASM and A*C are direct, adaptive and easy to implement. A*C has less parameter to tune, produces higher quality clusters, and is computationally more efficient and also proposed several ants moving strategies that have salient effect in accelerating the clustering process. Experimental results on standard clustering benchmarks demonstrate significant superiority in both computational time and clustering quality of ASM and A*C over previous methods [17].

4.3 Ant-MST: An Ant-Based Minimum Spanning Tree for Gene Expression Data Clustering

The proposed an ant-based clustering algorithm for document clustering based on the travelling salesperson scenario. In this paper, the presented an approach called Ant-MST for gene expression data clustering based on both ant-based clustering and Minimum Spanning Trees (MST). The ant-based clustering algorithm is firstly used to construct a fully connected network of nodes. Each node represents one gene, and every edge is associated with a certain level of pheromone intensity describing the co-expression level between two genes. Then MST is used to break the linkages in order to generate clusters. In this paper, we have presented a clustering algorithm Ant-MST for gene expression data clustering. It consists of two stages. First construct a fully connected network of nodes using the ant-based clustering algorithm and then build an MST from the fully connected graph and partition it into K clusters. Experimental results on three different datasets have been presented to illustrate its best feasibility and efficiency. In future work we will continue on the enhancement of the gene expression data clustering component and conduct a large scale of experiments to evaluate the system performance [7].

4.4 A New Clustering Algorithm Based on Hybrid Global Optimization Based on a Dynamical Systems Approach Algorithm

In this paper the hybrid HGOP algorithm is used to solve clustering problems. The HGOP algorithms use the notion of relationship between variables that describes influences of the changes of the variables to each other. The HGOP algorithm takes into account some relatively worse points for further consideration. This is what other methods do, such as Simulated Annealing, Genetic Algorithms and Taboo Search. The HGOP algorithm attempts to jump over local minimum points and tries to find deeper points. In this paper the global optimum is combined with the Tabu search to solve the problem of revisiting the visited region [1].

This hybridization makes the algorithm to be faster. The HGOP algorithm for data clustering can be applied when the numbers of clusters are known a priori and are crisp in nature. To evaluate the performance of the HGOP algorithm, it is compared with other stochastic algorithms viz. ant-colony, genetic algorithm, simulated annealing and Tabu search. The algorithm is implemented and tested on several simulation and real datasets; preliminary computational experience is very encouraging in terms of the quality of solution found and the processing time required [1].

4.5 A New Clustering Algorithm Based on the Chemical Recognition System of Ants

In this paper, describe a new model of the ant recognition system and its first application to the unsupervised clustering problem. Results are good when compared to those of the 10-MEANS algorithm. Our approach does not make any assumption about the nature of the data to be clustered and does not require an initial partition or an initial number of classes. This allows us to test our method in numerous application fields. The first one will be the web mining problem and more precisely the study of the behaviour of Internet users, because of the growing necessity for such tools for webmasters and because it provides a huge source of data. We are currently working on a new version of ANTCLUST that allows the user to see the generation of the nests in a 2D-space in real time. This version will rely on another modeling of the Label and its evolution which will be more accurate. In fact, there are numerous ways left to adapt the mathematical model of the ant recognition system to the unsupervised clustering problem [10].

4.6 An Improved Probabilistic Ant Based Clustering for Distributed Databases

In this paper, present an improved version of the Probabilistic Ant-based Clustering Algorithm for Distributed Databases (PACE). The most important feature of this algorithm is the formation of numerous zones in different sites based on corresponding user queries to the distributed database. Keywords, extracted out of the queries, are used to assign a range of values according to their corresponding probability of occurrence or hit ratio at each site. We propose the introduction of weights for individual or groups of data items in each zone according to their relevance to the queries along with the concept of familial pheromone trails as part of an Ant Odor Identification Model to bias the movements of different
types of ants towards the members of their own family. Its performance is compared against PACE and other known clustering algorithms for different evaluation measures and an improvement is shown in terms of convergence speed and quality of solution obtained [6].

The new proposed algorithm applied an improved version of the PACE algorithm for ant-based clustering applied in distributed databases. The main features of this algorithm are the introduction of weights and familial pheromone trails as part of an Ant Odor Identification Model. The aims are manifold-to reduce the convergence time and thereby improve the runtime of the solution, to improve the quality of clustering by forming compact and clearly defined clusters, to separate out outliers from the final solution, to provide a metric which easily determines the dissimilarity between any two points in the final cluster, to parallelize the algorithm for again reducing the processing time. These aims are reasonably achieved by our algorithm as ably proven by our experimental results. Further investigation could be carried out with a large number of data-sets as an extended version of our algorithm [6].

4.7 Towards Improving Clustering Ants: An Adaptive Ant Clustering Algorithm

This paper provided a number of contributions to the field in two main front lines. First, several modifications were introduced in the standard ant-clustering algorithm so as to enhance its performance and convergence properties. In particular, we proposed a cooling schedule for the parameter that controls the rate of picking-up objects from the grid. This guarantees that the algorithm always stabilizes after a number of iteration steps. Furthermore, we developed the ideas of progressive vision (Sherafat et al., 2004) and proposed a new form of implementing the pheromone heuristics on the grid in such a way that groups of data reinforce the attraction to those regions of the grid that contain data [2].

The proposed adaptive algorithm, named AcCA, was applied to a number of benchmark data sets and to a real world bio-informatics data set. The obtained results were compared to the standard ant-clustering algorithm with cooling schedule and modified dropping probability, and stress the benefits of the modifications introduced in the proposed algorithm. Most importantly, AcCA demonstrated a good robustness in terms of finding the correct number of clusters in the data set, low variations of the results in terms of number of clusters found, and always stabilized after a fixed number of iterations automatically defined by the algorithm. Despite the encouraging results presented here, there are still several avenues for investigation that deserve to be pursued [2].

5. CONCLUSION

The new study and most promising technology is called ant-based clustering techniques as superlative approach for clustering techniques and they are applied for image processing, data mining and so on. These algorithms are an appropriate alternative to traditional clustering algorithms. Ant-based clustering is a biologically inspired data clustering technique. Clustering task aims at the unsupervised classification of patterns in different groups. Clustering problem has been approached from different disciplines during last year’s. In recent years, many algorithms have been developed for solving numerical and combinatorial optimization problems and solving clustering problems. Ant-based clustering algorithms have been used in a large variety of applications. The algorithm has a number of features that make it an interesting study of cluster analysis. It has the ability of automatically discovering the number of clusters. The nature of the algorithm makes it fairly robust to the effects of outliers within the data. Research on ant-based clustering algorithms is still an on-going field of research. In this paper, we address a brief study of ant-based clustering algorithms and an overview of some of its applications and their techniques. There are a number of directions in which research on ant-based clustering can be continued. We conclude the study with listing some important algorithms and applications or fields for ant-based clustering algorithms: study the effect based on reasonably good validity index function to judge the fitness of several possible partitioning of the data of ant-based clustering schemes and validating mathematically; study the possibility of dynamic clustering using ant clustering with applications; study of transformation of ant clustering algorithms into supervised algorithms; developing new theoretical results of behavior of ant clustering algorithms and study of hierarchical ant-based clustering algorithms; to analyze the working principles that ant-based clustering shares with other clustering methods and their applications; finally the brief study, analyze about various ant based algorithms which are used to improving the clustering performance and to be supported they clustering quality and finding the good results to be applied the different applications.

REFERENCES

4. Bo Liu, Jihui Pan, Bob McKay, “Incremental Clustering Based on Swarm Intelligence”, Department of Computer Science, Dept. of Computer Science and Engineering, JINAN University, Seoul National University, Guangzhou, China, 510632 Seoul 151744 Korea, lbxldd@sohu.com, jhpan@jnu.edu.cn, rim@cse.snu.ac.kr, 2003.


6. Chandrasekar, R., Srinivasan, T., “An Improved Probabilistic Ant-based Clustering for Distributed Databases”, Department of Information Technology, 2 Assistant Professor, Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Sriperumbudur, India, 1chandra85@gmail.com, 2tsrini1969@gmail.com, IJCAI-07, 2701, 2007.

7. Deyu Zhou, Yulan He, Chee Keong Kwoh and Hao Wang, “Ant-MST: an Ant-Based Minimum Spanning Tree for Gene Expression Data Clustering”, School of Computer Engineering, Nanyang Technological University, Singapore 639798, fzhou0063, asylhe, asckkwoh, wang0046@ntu.edu.sg, 2006.


