DYNAMIC CLUSTER TRACKING USING DB SCAN ALGORITHM

A Report Submitted

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF TECHNOLOGY

In

COMPUTER SCIENCE & ENGINEERING

(Specialization: Software Engineering)

By

Jagdatta Singh
1170449003

Under the Supervision of
Gaurav Kumar Srivastava
Assistant Professor
BabuBanarasi Das University, Lucknow

To the
SCHOOL OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
BABU BANARASI DAS UNIVERSITY LUCKNOW
May, 2019

CERTIFICATE

It is certified that the work contained in this thesis entitled **—Dynamic Cluster Tracking Using DB Scan Algorithm‖**, by **Jagdatta Singh** (**Roll No. 1144709003**), for the award of Master of Technology from Babu Banarasi Das University has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

(Signature)

Gaurav Kumar Srivastava
Assistant Professor
Babu Banarasi Das University,
Lucknow

TABLE OF CONTENT

ABSTRACT

Multi-agent Systems have seen a tremendous growth since last two decades. They have been successfully used in various fields like telecommunication, distributed systems, decision support systems and robotics. Agents work in collaboration or competition exchanging messages and continuously interacting with the environment. Greater autonomy means larger complexity. Dynamically changing environment poses problems for the pre-programmed agents. Agents need to be reactive and instantaneous in order to solve a problem which pre-programming cannot achieve. Learning an environment lies at the core of agent functions. Then there should be incorporation of application of this learning in inter-agent coordination and competition. Agents collaborate for common organization specific processes and compete for their self goals. This research work is an effort to provide various contemporary learning techniques that agent(s) can employ for better consensus, coordination and understanding. In Markovian transitions the probability of reaching state s‘ from state s is only dependant on s and not on the history of earlier states. In this research work we will also discuss the Markovian nature of Multi-agent learning system. Lastly we will throw light upon the Q-value function approximation in the deep reinforcement learning paradigm. The author explained that there is an agent for handling the camera, another agent has sensor functionality built into it, and lastly the robotic cars also have agent software built into it.

Data mining refers to the process of retrieving data by discovering novel and relative patterns from large database. Clustering is a distinct phase in data mining that work to provide an established, proven structure from a collection of databases. A good clustering approach should be efficient and detect clusters of arbitrary shapes. Density Based Clustering is a well-known density based clustering algorithm which having advantages for finding out the clusters of different shapes and size from a large amount of data, which containing noise and outliers. In this research work , author discussed integrated Density Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm that is multiphase clustering algorithms which improves scalability and efficiency of clusters. Different DBSCAN algorithms perform different task to make cluster more dynamic and effective. Several DBSCAN clustering methods and their corresponding algorithms are described below which helps to further analysis. The authors try to develop a model for dynamic or moving video camera vigilance using Density Based Clustering and Ontologically Defined Environment. The authors are in the way to exploit the rich functionality exposed by the machine learning paradigm in which the stochastic environment to learn is depicted as a two dimensional graph where the position of an object can be given by its coordinates. In the research work the use of Density based clustering is depicted and the whole functionality is governed by the use of agents. In the research work the author can explain that there is an agent for handling the camera, another agent has sensor functionality built into it, and lastly the robotic cars also have agent software built into it.

The author is trying to develop a model for dynamic or moving video camera vigilance using Density Based Clustering and location sensors. The authors try to exploit the rich functionality exposed by the machine learning paradigm in which the stochastic

environment to learn is depicted as a two dimensional graph where the position of an object can be given by its coordinates. The author uses DBSCAN algorithm along with sensor enabled test ground area that keeps the X and Y co-ordinates of the moving objects. The idea here is to capture continuous video of the densest cluster of objects moving together. One practical usage of such system is a wild landscape where groups of animals are moving together to some destination. There will be a somewhat unorganized haphazard movement but we intend to capture only those animals that are greater in number as a group and the camera should move picturing them. This can be achieved by the DBSCAN algorithm.Regarding the use of ontologies, it is kept for future research as stated in the conclusion of the research work.

CHAPTER 1

INTRODUCTION

As two major communication technologies, the internet and wireless, are maturing rapidly to dominate our civilized life, the authors urgently need to re-establish users' confidence to harvest new potential applications of large-scale distributed systems. Service agents and distributed multi-agent systems (MASs) have shown the potential to help with this move as the lack of trust caused by heavily compromised security issues and concerns coupled with the out-of-date solutions are hindering the progress. The authors therefore seek new remedies to ensure that the continuity in developing new economies is maintained through building new solutions to address today's techno-economical problems. Following a scan of the literature the authors discuss the state-of-the-art progress followed by some observations and remarks for the researchers in the field. Here the authors try to develop a model for dynamic or moving video camera vigilance using Density Based Clustering and Ontologically Defined Environment. The authors are in the way to exploit the rich functionality exposed by the machine learning paradigm in which the stochastic environment to learn is depicted as a two dimensional graph where the position of an object can be given by its coordinates and explained that there is an agent for handling the camera, another agent has sensor functionality built into it, and lastly the robotic cars also have agent software built into it.

1.1Agent

An agent can be anything that perceives its environment through sensors and acts upon the environment through effectors or actuators. The perception is achieved through percepts at any instant derived from the environment. The agents may maintain complete history of the percepts at assorted instants of time. This is called percept sequence.

The agents' behavior is defined by the ‗agent function'. The purpose of the agent function is that it maps any percept sequence to some action. Our research is focused on learning ability of the agents. In order to better comprehend the learning aspect of the agents we must consider the following.

1.1.1 Simple Reflex Agent:

A single agent can be most appropriately defined and understood by considering a simple reflex agent[1]. These agents work upon a problem and act according to the current percept and not on the percept history.

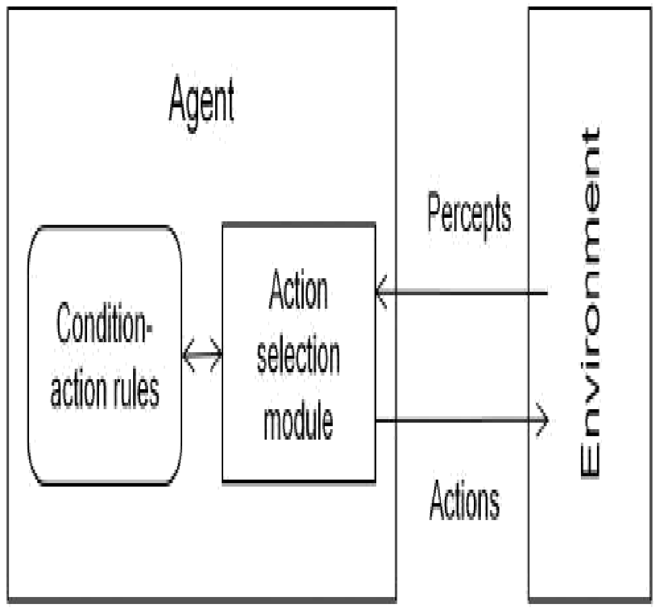

Figure 1.1: Simple Reflex Agent

A programmed function for a simple reflex agent in Python programming language is given below:

```
def RECEIVE_PERCEPT(percept):
state=percept
i = 1
while i< 6:
rule=RULE[i]
i=i+1
action=rule.ACTION
return ACTION
```

1.1.2 Utility-based Agent:

All MAS development methodologies like Prometheus, TROPOS, GAIYA, MASE, etc give importance to the goals that the agent(s) need to accomplish. Binding agents with goals does produce manifestation of their objectives but there is no performance meter as to judge the best accomplishments. That is, there can be many alternate ways to achieve goals and it becomes obvious that the best or at least the better alternative be given privilege. This is because agents learn from themselves and better paths to goal mean even better paths in future up till a near ideal solution is generated if the agent is faced by a similar problem or situation. For this purpose an agent's utility function is essential as it is the self performance measure. A utility based agent is intended to choose action which maximizes the expected utility of the action outcomes.

1.1.3 Learning Agent:

The last decade has seen much attention paid on the attributes of MAS like:

- If the agents are cooperating or competing.
- How does the adaptation process done in agents.
- The algorithms for adaptation.

The conceptual structure of a learning agent has been defined by Russel. This structure defines four core elements in learning agents. These are a learning element, a performance element, a critic and a problem generator. Figure 1.2 shows the learning agent architecture.

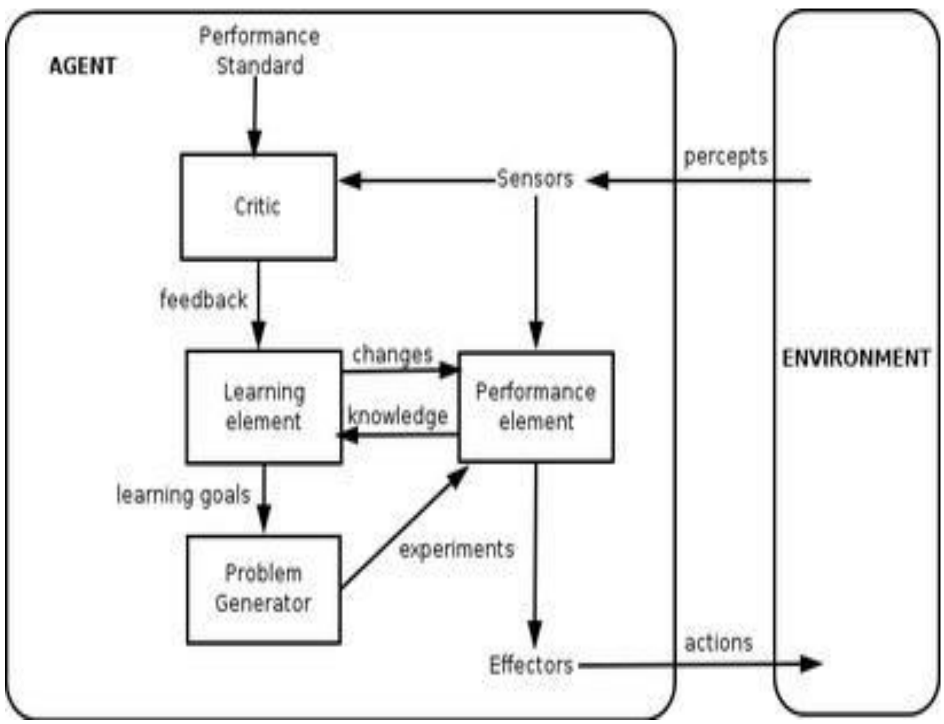

Figure1.2: Learning Agent

The performance element has percepts as input at the moment from the environment and decides the action to take. The critic element analyses the quality of the learning process and how good the agent is doing. The critic also provides the feedback to the learning agent. The critic modifies the performance element for best results in future.

Lastly, the performance generator suggests actions for new informative experiences.

Weiss proposes six aspects that concern the learning process of the agents:

- **Degree of Decentralization:**

The learning can be done by either a single agent called central agent alone or it can be done by few or all the agents collectively. This employs understanding, collaboration and unified distribution in perceiving the environment. This will also raise a question regarding whether the agents have a unified view and perception of the environment.

- **Interaction specific agents:**

Interactions can be based on observation, indirect effects by environment or explicit relationships. Interactions change with time and are modeled using some message exchange between the agents.

- **Aspect related to involvement:**

This depicts the level of involvement of agents with the environment in the learning process.

- **Type of goals**

Goals of the agents can be local (self goals or selfish goals) or global (organizational). MAS may have agents that are collaborating, competitive or sometimes both. Competitive agents have selfish goals like a Pong game with 2 agents, player1 and player2. Agents can be cooperative like a fire extinguishing robotic system. The competitive agents work to maximize their individual reward rather than collective. The cooperative agents work for common goals and share the reward. These two types of the agents have different learning phenomenon. Their perception, learning and actions follow different processes.

- **The Learning Algorithm:**

A learning algorithm describes the procedure the agents follow in order to learn from the environment.

- **Feedback from learning**

Proper assessment of the learning agent needs to be done in order for the agent to know if it is progressive ie., advantageous or deleterious.

1.2 Clustering:

Clustering is a popular data analysis technique. Clustering algorithms can be widely applied in many fields including: pattern recognition, machine learning, image processing, and information retrieval and so on. It also plays an important role in data mining. All the existing clustering algorithms have their own characteristics, but also have their own flaws. As a kind of other clustering, density based algorithm is simple and high efficiency algorithm. Density-based clustering algorithms, which are designed to discover clusters of arbitrary shape in databases with noise, a cluster is defined as a high-density region partitioned by lowdensity regions in data space. Density Based Spatial Clustering of Applications with Noise (DBSCAN) is a typical density-based clustering algorithm. DBSCAN can discover clusters of arbitrary shape. But it is sensitive to the input parameters, especially when the density of data is nonuniform. The DBSCAN clustering algorithms usually can be classified into the following different categories:

- Partitioning based DBSCAN clustering
- Grid-based DBSCAN clustering
- Hierarchical DBSCAN clustering
- Detection Based DBSCAN clustering
- Incremental DBSCAN clustering
- Spatial–temporal DBSCAN clustering

1.2.1 Partition based DBSCAN clustering:

This method generally results in a set of M clusters, each object belonging to one cluster. Each cluster may be represented by a centroid or a cluster representative; this is some sort of summary description of all the objects contained in a cluster. The precise form of this description will depend on the type of the object which is being clustered. In case where real-valued data is available, the arithmetic mean of the attribute vectors for all objects within a cluster provides an appropriate representative; alternative types of centroid may be required in other cases, e.g., a cluster of documents can be represented by a list of those keywords that occur in some minimum number of documents within a cluster. If the number of the clusters is large, the centroids can be further clustered to produces hierarchy within a dataset.

1.2.2 Grid-based DBSCAN clustering:

This technique quantizes the data set in to a no of cells and then work with objects belonging to these cells. They do not relocate points but ratter builds several hierarchical levels of groups of objects. The merging of grids and consequently clusters, does not depend on a distance measure .It is determined by a predefined parameter.

1.2.3 Hierarchical DBSCAN clustering:

It is a method of cluster analysis which seeks to build a hierarchy of clusters. Thebasics of hierarchical clustering include Lance-Williams formula, using idea of conceptual clustering. The hierarchical algorithms build clusters gradually (as crystals are grown) Strategies for hierarchical clustering generally fall into two types: In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object. Hierarchical Clustering is subdivided into agglomerative methods, which proceed by series of fusions of the n objects into groups, and divisive methods, which separate n objects successively into finer groupings. Agglomerative techniques are more commonly used. Hierarchical clustering may be represented by a two dimensional diagram known as dendrogram which illustrates the fusions or divisions made at each successive stage of analysis.

1.2.4 Detection Based DBSCAN clustering:

In this technique simplified detection problem can be solved efficiently as an upper bound on a discredited likelihood function. This is an efficient algorithm for recovering the maximum likelihood number of sides and orientation only at the locations of the most likely polygons. The first stage of the detection is posed as a discrete Hough-based algorithm. The second stage takes an approximation to the full likelihood function to recover orientation and number of sides.

1.2.5 Incremental DBSCAN clustering:

This algorithm is used to handle dynamic databases. It has the ability to changing the radius threshold value dynamically. The algorithm restricts the number of the final clusters and reads the original dataset only once. At the same time the frequency information of the attribute values is introduced by this algorithm. It can be used for the categorical data. The algorithm can not only overcome the impact of the inadequate of the memory when clustering the large scale data set, but also accurately reflect the characteristics of the data set.

1.2.6 Spatial Temporal DBSCAN clustering:

Spatial Temporal DBSCAN clustering is new clustering algorithm designed for storing and clustering a wide range of spatial–temporal data. Environmental data, from a variety of sources, were integrated as coverages, grids, shapefiles, and tables. Special functions were developed for data integration, data conversion, visualization, analysis and management. User-friendly interfaces were also developed allowing relatively inexperienced users to operate the system.

Spatial–temporal data is indexed and retrieved according to spatial and time dimensions. A time period attached to the spatial data expresses when it was valid or stored in the database. A temporal database may support valid time, transaction time or both. Valid time denotes the time period during which a fact is true with respect to the real world. Transaction time is the time period during which a fact is stored in the database. This study focuses on valid time aspect of temporal data.
There are number of clustering techniques but this research work mainly focuses on the widely used DBSCAN clustering technique. Significant work is done in the field of Density based

clustering. Author represents a brief overview of some work done in Density based clustering including email classification, spam detection etc. Number of application areas and techniques are highlighted in Density-based clustering. One approach developed the incremental clustering for mining large database environment. This approach present the first incremental clustering algorithm based on DBSCAN clustering which is applicable to any database containing data from a metric space. Due to the density based nature of DBSCAN, the insertion or deletion of an object affects the current clustering only in the neighbourhood of this object. Thus, efficient algorithm scan be given for incremental insertions and deletions to an existing clustering. Incremental DBSCAN yields significant speed-up factors over DBSCAN even for large numbers of daily updates in a data warehouse. In this paper, sets of updates are processed one at a time without considering the relationships between the single updates.

A innovative technique which is used to compare in between two different clustering algorithms (DBSCAN and SNN) described several implementations of the DBSCAN and SNN algorithms, two density-based clustering algorithms. These implementations can be used to cluster sets of points based on their spatial density. The results obtained through the use of these algorithms show that SNN performs better than DBSCAN since it can detect clusters with different densities while DBSCAN cannot.

Many clustering algorithms have been proposed so far, seldom was focused on high dimensional and incremental databases. An incremental approach on Grid Density-Based Clustering Algorithm (GDCA) discovers clusters with arbitrary shape in spatial databases. It first partitions the data space into a number of units, and then deals with units instead of points. Only those units with the density no less than a given minimum density threshold are useful in extending clusters. An incremental clustering algorithm--IGDCA is also presented in this paper, applicable in periodically incremental environment.

An innovative approach presents a new density-based clustering algorithm, ST-DBSCAN, which is based on DBSCAN. It proposes three marginal extensions to DBSCAN related with the identification of

(i) Core objects
(ii) Noise objects
(iii) Adjacent clusters.

In contrast to the existing density-based clustering algorithms, this algorithm has the ability of discovering clusters according to non-spatial, spatial and temporal values of the objects. In this paper, it also presents a spatial–temporal data warehouse system designed for storing and clustering a wide range of spatial–temporal data.
It is the process of grouping large data sets according to their similarity. Cluster analysis is a major tool in many areas of engineering and scientific applications including data segmentation,

discretization of continuous attributes, data reduction, outlier detection, noise filtering, pattern recognition and image processing. In the field of Knowledge Discovery in Databases (KDD),

cluster analysis is known as unsupervised learning process, since there is no prior knowledge about the data set. Most studies in KDD focus on discovering clusters from ordinary data (non-spatial and non-temporal data), so they are impractical to use for clustering spatial–temporal data. Spatial–temporal data refers to data which is stored as temporal slices of the spatial dataset. Knowledge discovery from spatial–temporal data is a very promising subfield of data mining because increasingly large volumes of spatial–temporal data are collected and need to be analyzed. The knowledge discovery process for spatial–temporal data is more complex than for non-spatial and non-temporal data. Because spatial–temporal clustering algorithms have to consider the spatial and temporal neighbors of objects in order to extract useful knowledge. Data & Knowledge Engineering designed for spatial–temporal data can be used in many applications such as geographic information systems, medical imaging, and weather forecasting. This research work presents a new density-based clustering algorithm DBSCAN, which is based on the algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noise). In DBSCAN, the density associated with a point is obtained by counting the number of points in a region of specified radius around the point. Points with a density above a specified threshold are constructed as clusters. Among the existing clustering algorithms, we have chosen DBSCAN algorithm, because it has the ability in discovering clusters with arbitrary shape such as linear, concave, oval, etc. Furthermore, in contrast to some clustering algorithms, it does not require the predetermination of the number of clusters.

DBSCAN has been proven in its ability of processing very large databases. We have analysed DBSCAN algorithm in three important directions. First, unlike the existing density-based clustering algorithms, our algorithm can cluster spatial–temporal data according to its non-spatial, spatial and temporal attributes. Second, DBSCAN cannot detect some noise points when clusters of different densities exist. Our algorithm solves this problem by assigning to each cluster a density factor. Third, the values of border objects in a cluster may be very different than the values of border objects in opposite side, if the non-spatial values of neighbor objects have little differences and the clusters are adjacent to each other. Our algorithm solves this problem by comparing the average value of a cluster with new coming value. In addition to new clustering algorithm, this research work also presents a spatial data warehouse system designed for storing and clustering a wide range of spatial–temporal data. Environmental data, from a variety of sources, were integrated as coverages, grids, shapefiles, and tables. Special functions were developed for data integration, data conversion, visualization, analysis and management. User-friendly interfaces were also developed allowing relatively inexperienced users to operate the system.

In order demonstrate the applicability of our algorithm to real world problems, we applied our algorithm to the data warehouse, and then presented and discussed the data mining results. Spatial–temporal data is indexed and retrieved according to spatial and time dimensions. A time period attached to the spatial data expresses when it was valid or stored in the database. A temporal database may support valid time, transaction time or both. Valid time denotes the time period during which a fact is true with respect to the real world. Transaction time is the time period during which a fact is stored in the database. This study focuses on valid time aspect of temporal data. The rest of the research work is organized as follows. Section 2 summaries the existing clustering algorithms and gives basic concepts of density-based clustering algorithms. Section 3 describes the drawbacks of existing density-based clustering algorithms and our efforts to overcome these problems. Section 4 explains our algorithm in detail and presents the performance of the algorithm. Section 5 presents three applications which are implemented to demonstrate the applicability of it to real world problems.

1.3 Clustering Techniques:

The process of Clustering involves splitting the data into groups of similar objects. Each group is called a cluster. The intra-cluster resemblance is high and inter-cluster resemblance or similarity index is low. Data mining incorporates this technique. Clustering is considered as a part of unsupervised learning. Different types of clusters are analyzed in and given as under:

- **Well Separated clusters**:
  Here, intra-cluster similarity is very high and inter-cluster similarity is low.
- **Centre-Based clusters:**
  Every object in the cluster is more similar to the centre also called the _centroid‘ than to the centre of any other cluster.
- **Contiguous clusters:**
  Nodes in a cluster are nearest (or more alike) to one or more other nodes in the cluster as compared to any node that is not part of the cluster.
- **Density based clusters:**
  A cluster is a thick portion of points, which is separated by according to the low-density regions, from other regions that is of high density.
- **Conceptual clusters**:
  Conceptual cluster shares few common features, or possess a particular thought.

1.3.1 Density-based clustering:

The problem of clustering can be defined as follows:

Definition 1.

Given a database of n data objects D = {o1,o2,...,on}. The process of partitioning D into C = {C1,C2,...,Ck} based on a certain similarity measure is called clustering, Ci‘s are called clusters, where Ci D, (i = 1, 2,...,k), Tk i¼1Ci ¼ ; and Sk i¼1Ci ¼ D. Clustering algorithms can be categorized into five main types: Partitional, Hierarchical, Grid-based, Model-based and Density-based clustering algorithms. In Partitional algorithms, cluster similarity is measured in regard to the mean value of the objects in a cluster, center of gravity, (K-Means) or each cluster is represented by one of the objects of the cluster located near its center (K-Medoid). K is an input parameter for these algorithms; unfortunately it is not available for many applications. CLARANS is an improved version of K-Medoid algorithm for mining in spatial databases. Hierarchical algorithms such BIRCH produces a set of nested clusters organized as a hierarchical tree. Each node of the tree represents a cluster of a database D. Grid-based algorithms such as STING, WaveClusterare based on multiple level grid structure on which all operations for clustering are performed. In Model-based algorithms, a model is hypothesized for each of the clusters and the idea is to find the best fit of that model to each other. They are often based on the assumption that the data are generated by a mixture of underlying probability distributions.

The Density-based notion is a common approach for clustering. Density-based clustering algorithms are based on the idea that objects which form a dense region should be grouped together into one cluster. They use a fixed threshold value to determine dense regions. They search for regions of high density in a feature space that are separated by regions of lower density. Density-based clustering algorithms such as DBSCAN, OPTICS, DENCLUE , CURD are to some extent capable of clustering databases . One drawback of these algorithms is that they capture only certain kinds of noise points when clusters of different densities exist. Furthermore, they are adequate if the clusters are distant from each other, but not satisfactory when clusters are adjacent to each other. In our study, we have chosen DBSCAN algorithm, because it has the ability in discovering clusters with arbitrary shape such as linear, concave, oval, etc. Furthermore, in contrast to some clustering algorithms, it does not require the predetermination of the number of clusters. DBSCAN has been proven in its ability of processing very large databases. In the literature, DBSCAN algorithm was used in many studies. For example, the other popular densitybased algorithm OPTICS (Ordering Points to Identify the Clustering Structure) is based on the concepts of DBSCAN algorithm and identifies nested clusters and the structure of clusters. Incremental DBSCAN algorithm is also based on the clustering algorithm DBSCAN and is used for incremental updates of a clustering after insertion of a new object to the database and deletion of an existing object from the database. Based on the formal notion of clusters, the incremental algorithm yields the same result as the non-incremental DBSCAN algorithm.

SDBDC (Scalable Density-Based Distributed Clustering) method also uses DBSCAN algorithm on both local sites and global site to cluster distributed objects. In this method, DBSCAN algorithm is firstly carried out on each local site. Then, based on these local clustering results, cluster representatives are determined. Then, based on these local representatives, the standard DBSCAN algorithm is carried out on the global site to construct the distributed clustering. This study proposes the usage of different Eps-values for each local representative. Wen et al[78] adopted DBSCAN and Incremental DBSCAN as the core algorithms of their query clustering tool. They used DBSCAN to cluster frequently asked questions and most popular topics on a search engine. Spieth et al. [24] applied DBSCAN to identify solutions for the inference of regulatory networks. Finally, SNN density-based clustering algorithm is also based on DBSCAN and it is applicable to high-dimensional data consisting of time series data of atmospheric pressure at various points on the earth.

Basic concepts DBSCAN is designed to discover arbitrary-shaped clusters in any database D and at the same time can distinguish noise points. More specifically, DBSCAN accepts a radius value Eps(e) based on a user defined distance measure and a value MinPts for the number of minimal points that should occur within Eps radius. Some concepts and terms to explain the DBSCAN algorithm can be defined as follows.

Definition 2 (Neighborhood):

It is determined by a distance function (e.g., Manhattan Distance, Euclidean Distance) for two points p and q, denoted by dist(p,q).

Definition 3 (Eps-neighborhood):

The Eps-neighborhood of a point p is defined by {q 2 Djdist(p,q) 6 Eps}.

Definition 4 (Core object):

A core object refers to such point that its neighborhood of a given radius (Eps) has to contain at least a minimum number (MinPts) of other points.

Definition 5 (Directly density-reachable):

An object p is directly density-reachable from the object q if p is within the Eps-neighborhood of q, and q is a core object.

Definition 6 (Density-reachable):

An object p is density-reachable from the object q with respect to Eps and MinPts if there is a chain of objects p1,...,pn, p1 = q and pn = q such that pi+1 is directly density-reachable from pi with respect to Eps and MinPts, for 1 6 i 6 n, pi 2 D.

Definition 7 (Density-connected):

An object p is density-connected to object q with respect to Eps and MinPts if there is an object o 2 D such that both p and q are density-reachable from o with respect to Eps and MinPts (Fig. 1b).

Definition 8 (Density-based cluster):

A cluster C is a non-empty subset of D satisfying the following __maximality‘‘ and __connectivity‘‘ requirements: (1) "p,q: if q 2 C and p is density-reachable from q with respect to Eps and MinPts, then p 2 C. (2) "p,q 2 C: p is density-connected to q with respect to Eps and MinPts.

Definition 9 (Border object):

An object p is a border object if it is not a core object but density-reachable from another core object. The algorithm starts with the first point p in database D, and retrieves all neighbors of point p within Eps distance. If the total number of these neighbors is greater than MinPts—if p is a core object—a new cluster is created. The point p and its neighbors are assigned into this new cluster. Then, it iteratively collects the neighbors within Eps distance from the core points. The process is repeated until all of the points have been processed.

1.4 Use of Clustering and Methods:

Clustering has wide applications in Image Processing, Document Classification, and Pattern Recognition, Spatial Data Analysis, Economic Science and Cluster Web log data to discover similar web access patterns, etc.
Various Methods of clustering have been reported in literatureHierarchical Methods:

- Agglomerative Nesting (AGNES)
- Divisive Analysis (DIANA)
- Clustering using Representatives (CURE)
- Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH)

Partitioning methods are:

- K-mean method
- K-Medoids method (PAM)
- Farthest First Traversal k-center (FFT)
- CLARA
- CLARANS
- Fuzzy K-Means

- Fuzzy K-Modes
- K-Modes
- Squeezer
- K-prototypes
- COOLCAT

Density Based Methods: Density based clustering methods include:

- DBSCAN
- GDBSCANS
- OPTICS
- DBCLASD
- DENCLUE

Model Based method: There are two approaches for Model based methods:

- Statistical approach includes AutoClass method.
- Neural Network Approach includes Competitive learning and Self-organizing feature maps.

Grid Based: Some of the Grid based clustering methods are:

- STING
- Wave Cluster
- CLIQUE
- MAFIA

Density-Based Spatial Clustering of Applications with Noise is a contemporary unsupervised learning method used in building models and machine learning algorithms. Unsupervised learning methods are utilized when we are seeking to find no clear objective or outcome. Instead, we are clustering the data together based on some similarity and commonality of features.

DBSCAN is a clustering method that is used in machine learning to separate clusters of high density from clusters of low density. DBSCAN is a density based clustering algorithm that does a fine job of finding areas in the data that have a high density of observations in similarity, and ignores areas of the data that are not very dense with observations. DBSCAN can sort data into clusters of varying shapes. This is yet another strong advantage of DBSCAN. DBSCAN algorithm can be state as below:

- Divides the dataset into *n* dimensions
- For each point in the dataset, DBSCAN forms an *n* dimensional shape around that data point, and then counts how many data points fall within that shape.

- DBSCAN counts this shape as a *cluster*. DBSCAN iteratively expands the cluster, by going through each individual point within the cluster, and counting the number of other data points nearby. Take the graphic shown in figure for an example:

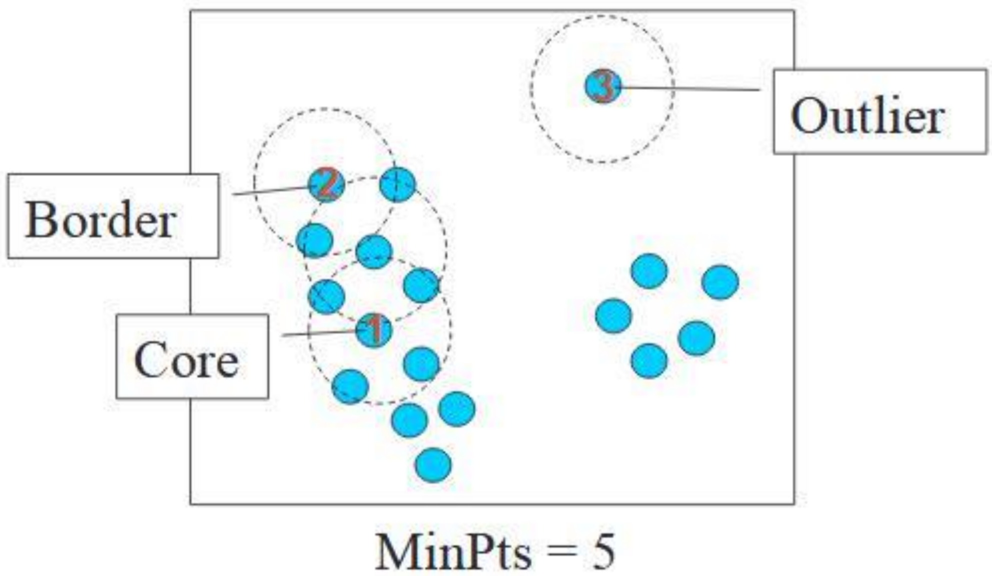

Figure 1.3: The DBSCAN Algorithm

1.5 Location Detection and Tracking of Moving Targets:

Many applications require information about an object's location for rescue, emergency and security purposes. The approaches that access an object's location are typically divided into two groups: active and passive localization. In the former approach, the object is associated with a mobile station (MS), such as a tag or device in a communication network. The object's location is determined by sharing data between the MS and the base stations (BSs). The Global Positioning System (GPS), cellular networks, Bluetooth and wireless sensor networks (WSNs) are used in active localization. In the latter approach, the object does not communicate with other devices. However, the object's location can be determined by using the reflected signal from the object. Radio detection and ranging (radar), sound navigation and ranging (sonar) and laser detection and ranging (LADAR) are the most common types of passive localization. These methods have both advantages and disadvantages. However, GPS and long-range radar generate many errors during indoor localization and tracking. Cellular networks and WSNs are limited by their complicated controls and protocols. Sonar and LADAR signals are degraded by environmental interference. Therefore, ultra-wide band (UWB) radar has become an emerging technology that is appropriate for indoor localization and tracking. UWB radar has many advantages, such as a high spatial resolution, the ability to mitigate interference, through-the-wall visibility, a simple transceiver and a low cost.

CHAPTER 2

LITERATURE REVIEW

Abdessameud and Tayebi[1],deals with consensus strategy design for double-integrator dynamics. Specifically, they consider the case where the control inputs are required to be a priori bounded and the velocity (second state) is not available for feedback. Two different design methods are proposed. First, based on the auxiliary system approach, they propose a consensus algorithm that extends some of the existing results in the literature to account for actuator saturations and the lack of velocity measurement. The proposed velocity-free control scheme, using local information exchange, achieves consensus among the team members with an a priori bounded control law, whose upper bound depends on the number of neighbors of the vehicle.

Second, they propose another approach based on the use of a high order dynamic auxiliary system such that the upper bound of the control law is independent of the number of neighbors of the vehicle, and the performance of the closed loop system is improved in terms of the response damping.Ajorlo andMomeni[2], concerned with the convergence of a class of continuous-time nonlinear consensus algorithms for single integrator agents. In the consensus algorithms studied here, the control input of each agent is assumed to be a state-dependent combination of the relative positions of its neighbors in the information flow graph. Using a novel approach based on the smallest order of the nonzero derivative, it is shown that under some mild conditions the convex hull of the agents has a contracting property. Abdulghafor [3],provides an overview of consensus problems in multi-agent cooperative control with the goal of exposing the related literature and promote the research in this area. The document presents the theoretical results concerning the search for consensus in the involved topologies with information exchange that is invariant in time and change dynamically.Dallil [4], introduces a method of computing mass assignment using a normalized Mahalanobis distance. While the decision making process is based on the extension of the frame of hypotheses, the method has been tested for a nearly constant velocity target and compared with both the nearest neighbor filter and the joint probabilistic data associations filter in highly ambiguous cases using Monte Carlo simulations.

The results demonstrate the feasibility of the proposal, and show improved performance compared to the aforementioned alternative commonly used methods.Garulli and Giannitrapani[5], analyzes two classes of consensus algorithms in the presence of bounded measurement errors. The considered protocols adopt an updating rule based either on constant or vanishing weights. Under the assumption of bounded error, the consensus problem is cast in a set-membership framework, and the agreement of the team is studied by analyzing the evolution of the feasible state set. Y Ahiska [6], removes the need for mechanicalpan, tilt, and zoom apparatus is disclosed. The video camera includes the following. A wide-angle optical system is configured to receive an optical image. An image sensor is coupled to the optical system and is configured to convert the optical image to an electronic image. An image processing circuit is coupled to the image sensor and is configured to receive the electronic image and to execute movement and zoom operations by correcting distortions in the electronic image. Aslam, Butler and Constantin[7], examine the role of very simple and noisy sensors for the tracking problem.

They proposed a binary sensor model, where each sensor's value is converted reliably to one bit of information only: whether the object is moving toward the sensor or away from the sensor. They show that a network of binary sensors has geometric properties that can be used to develop a solution for tracking with binary sensors and present resulting algorithms and simulation experiments.Bakr, Ghanem and Ismail[8], organize such information in an efficient manner is more important than ever. With such dynamic nature, incremental clustering algorithms are always preferred compared to traditional static algorithms. In this research work, an enhanced version of the incremental DBSCAN algorithm is introduced for incrementally building and updating arbitrary shaped clusters in large datasets.

The proposed algorithm enhances the incremental clustering process by limiting the search space to partitions rather than the whole dataset which results in significant improvements in the performance compared to relevant incremental clustering algorithms.Birant and Kut [9], presents a new density-based clustering algorithm, ST-DBSCAN, which is based on DBSCAN. They propose three marginal extensions to DBSCAN related with the identification of (i) core objects, (ii) noise objects, and (iii) adjacent clusters. In contrast to the existing density-based clustering algorithms, our algorithm has the ability of discovering clusters according to non-spatial, spatial and temporal values of the objects. In this research work, also present a spatial–temporal data warehouse system designed for storing. Jiang and Ravindran[10], introduce a surveillance system, which tracks mobile targets, is one of the most important applications of wireless sensor networks. When nodes operate in a duty cycling mode, tracking performance can be improved if the target motion can be predicted and nodes along the trajectory can be proactively awakened.

However, this will negatively influence the energy efficiency and constrain the benefits of duty cycling. In this research work, they present a Probability-based Prediction and Sleep Scheduling protocol (PPSS) to improve energy efficiency of proactive wake up.Wang,

Wang, and Zhang [11], considers the robust adaptive consensus tracking for higher-order multi-agent uncertain systems with nonlinear dynamics via distributed intermittent communication protocol. The main contribution of this work is solving the robust consensus tracking problem without the assumption that the topology among followers is strongly connected and fixed. The focus is the problem of actuator with occasional failure inputs and communication resources constraints.Chaturvedi, Green and Caroll[12], recommend that these approaches be used as "complementary" procedures in performing cluster analysis. They also present an empirical comparison of K-modes and latent class, where the former method prevails.Darong andPeng[13], combined the grid partition technique and multi-density based clustering algorithm, has improved its efficiency. On the other hand, because the Eps and Minpts parameters of the DBSCAN algorithm they are auto-generated, sothey are more objective. Experimental results shown that the new algorithm not only can better differentiate bettheyen noises and discovery clusters of arbitrary shapes.S Datta[14],consider six clustering algorithms (of various flavors!) and evaluate their performances on a they will-known publicly available microarray data set on sporulation of budding yeast and on two simulated data sets. Among other things, they formulate three reasonable validation strategies that can be used with any clustering algorithm when temporal observations or replications are present. They evaluate each of these six clustering methods with these validation measures. While the ‗best' method is dependent on the exact validation strategy and the number of clusters to be used, overall *Diana* appears to be a solid performer. Interestingly,

The performance of correlation-based hierarchical clustering and model-based clustering (another method that has been advocated by a number of researchers) appear to be on opposite extremes, depending on what validation measure one employs. Amorim, Barthélemyand Ribeiro[15], study the application of simulated annealing and tabusearch to the solution of the clique partitioning problem. They illustrate the effusiveness of these techniques by computational results associated not only with randomly generated problems, but also with real-life problems arising from applications concerning the optimal aggregation of binary relations into an equivalence relation. The need for these approaches is emphasized by the example of a special class of instances of the clique partitioning problem. Theyi, Jin, S Vural[16], proposed a distributed clustering algorithm, Energy-efficient Clustering (EC), that determines suitable cluster sizes depending on the hop distance to the data sink, while achieving approximate equalization of node lifetimes and reduced energy consumption levels. They additionally propose a simple energy-efficient multihop data collection protocol to evaluate the effectiveness of EC and calculate the end-to-end energy consumption of this protocol; yet EC is suitable for any data collection protocol that focuses on energy conservation. Sonbatyand
Ismail[17], used clustering for identifying useful patterns and interesting distributions in the underlying data.

Several algorithms for clustering large data sets have been proposed in the literature using different techniques. Density-based method is one of these methodologies which can detect arbitrary shaped clusters where clusters are defined as dense regions separated by low density regions. They present a new clustering algorithm to enhance the density-based algorithm DBSCAN. Ester, Kriegel, and Sander[18],performed an experimental evaluation of the effectiveness and efficiency of DBSCAN using synthetic data and real data of the SEQUOIA 2000 benchmark. The results of our experiments demonstrate that (1) DBSCAN is significantly more effective in discovering clusters of arbitrary shape than the theyll-known algorithm CLARANS, and that (2) DBSCAN outperforms CLARANS by factor of more than 100 in terms of efficiency.Akyildiz, W Su andSankarasubramaniam[19],describes the concept of sensor networks which has been made viable by the convergence of micro-electro-mechanical systems technology, wireless communications and digital electronics. First, the sensing tasks and the potential sensor networks applications are explored, and a review of factors influencing the design of sensor networks is provided. Then, the communication architecture for sensor networks is outlined, and the algorithms and protocols developed for each layer in the literature are explored.Jiang and Wang [20], described that consensus state is an important and fundamental quantity for consensus problems of multi-agent systems, which indicates where all the dynamical agents reach. In this research work, weighted average consensus with respect to a monotonic function, which means that the trajectories of the monotonic function along the state of each agent reach the weighted average of their initial values, is studied for a group of kinematic agents with time-varying topology.

Jiang and Wang[21], established the explicit expression of the consensus state for the entire group. Second, they proved that the agents of the group under a particular type of nonlinear interaction can reach the consensus state in finite time in the scenarios with fixed and switching undirected topologies. The results are also extended to the case where the topology of the group is directed and satisfies. Fu, Wang, Ge, Chen and Tenganalyzed [22], rapid development of information age, more and more data can be obtained from the Internet, it is very difficult to get useful information and knowledge from these huge amounts of data. On the foundation of the existing algorithm based on DBSCAN, a new improved incremental DBSCAN clustering algorithm is proposed. Combining with cloud computing open source framework Hadoop, the improved algorithm use the programming model of MapReduce which can easy write distributed applications and simplify distributed
Fu and Zhao[23], proposed a parallel DBSCAN clustering algorithm based on Hadoop, which is a simple yet potheyrful parallel programming platform.

The experimental results demonstrate that the proposed algorithm can scale well and efficiently process large datasets on commodity hardware.Feng, Jaewon and James [24], overviewed the information-driven approach to sensor collaboration in ad hoc sensor networks. The main idea is for a network to determine participants in a "sensor collaboration" by dynamically optimizing the information utility of data for a given cost of communication and computation. A definition of information utility is introduced, and several approximate measures of the information utility are developed for reasons of computational tractability.

They illustrate the use of this approach using examples drawn from tracking applications.Gasch and Eisen[25], simplified the orchestration of gene expression by coregulating genes whose products function together in the cell. Many proteins serve different roles depending on the demands of the organism, and therefore the corresponding genes are often coexpressedwith different groups of genes under different situations. This poses a challenge in analyzing whole-genome expression data, because many genes will be similarly expressed to multiple, distinct groups of genes because most commonly used analytical methods. Girma, Garuba and Goel[26], addressed the need to prevent DDoS attacks by defining and demonstrating a hybrid detection model by introducing an advanced and efficient approach to recognize and efficiently discriminate the flood attacks from the flush crowd (legitimate access). Moreover, this research work introduce and discusses, most importantly, the application of multi-variate correlation among the selected and ranked features to significantly reduce the false alarm rate, which is one of the major issue associated with the current available solution.Goil, Nagesh, and Choudhary [27], used clustering techniques in database mining for finding interesting patterns in high dimensional data. These are useful in various applications of knowledge discovery in databases. Some challenges in clustering for large data sets in terms of scalability, data distribution, understanding end-results, and sensitivity to input order, have received attention in the recent past. Recent approaches attempt to find clusters embedded in subspaces of high dimensional data.

In this research workthey propose the use of adaptive grids.Hu[28], study the problem of robust consensus tracking for a class of second-order multi-agent dynamic systems with disturbances and unmodeled agent dynamics. Contrary to previous approaches, they design continuous distributed consensus protocols to enable global asymptotic consensus tracking. Our focus is on consensus protocol design and stability analysis which also leads to the derivation of sufficient conditions for consensus tracking. Gupta[29], developed a simple algorithm that detects and tracks a moving target, and alerts sensor nodes along the projected path of the target.

The algorithm involves only simple computation and localizes communication only to the nodes in the vicinity of the target and its projected course. The algorithm is evaluated on a small-scale testbed of Berkeley motes using a light source.Habib and Afzal[30], numerous different research work recommendation approaches have been proposed. Some of these include methods based on metadata, content similarity, collaborative filtering, and citation analysis, among others. Citation analysis methods include bibliographic coupling and co-citation analysis. Much research has been done in the area of co-citation analysis. Researchers have also performed experiments using the proximity. Harisinghaney, Dixit and Gupta [31], used Enron corpus's dataset of spam and ham emails. In this research research work, they provide comparison performance of all three algorithms based on four measuring factors namely: precision, sensitivity, specificity and accuracy. They are able to attain good accuracy by all the three algorithms. The results have shown comparison of all three algorithms applied on same data set.Hartigan and Wong[32], proposed a algorithm requires as input a matrix of M points in N dimensions and a matrix of K initial cluster centres in N dimensions. The number of points in cluster L is denoted by NC (L). D (I, L) is the Euclidean distance bettheyen point I and cluster L. The general procedure is to search for a K-partition with locally optimal within-cluster sum of squares by moving points from one cluster to another.Du, Cheng, He and Jia[33], investigated second-order consensus problem of nonlinear leader-following multi-agent systems. To solve the case that the velocities of all agents cannot be measured and the nonlinearity is unknown, an observer-based dynamic output feedbackcontroller is proposed based on a non-separation principle method. Using the feedback domination technique.

It is shown that the systems output can reach consensus by choosing appropriate gains.He and Tan[34], presented MR-DBSCAN, a scalable DBSCAN algorithm using MapReduce. In our algorithm, all the critical sub-procedures are fully parallelized. As such, there is no performance bottleneck caused by sequential processing. Most importantly, they propose a novel data partitioning method based on computation cost estimation. The objective is to achieve desirable load balancing even in the context of heavily sketheyd data. Besides, They conduct our evaluation using real large datasets with up to 1.2 billion points. The experiment results theyll confirm the efficiency and scalability of MR-DBSCAN.He, Xu and Deng[35], presented a new efficient algorithm for clustering categorical data, Squeezer, which can produce high quality clustering results and at the same time deserve good scalability.

The Squeezer algorithm reads each tuple t in sequence, either assigning t to an existing cluster (initially none), or creating t as a new cluster, which is determined by the similarities bettheyen t and clusters. Due to its characteristics, the proposed algorithm is extremely suitable for clustering data streams. Kim, Shim, and Seo[36], studies the output consensus problem for a class of heterogeneous uncertain linear multi-agent systems. All the agents can be of any order (which might widely differ among the agents) and possess parametric uncertainties that range over an arbitrarily large compact set. The controller uses only the output information of the plant; moreover, the delivered information throughout the communication network is also restricted to the output of each agent.Huang[37], described extensions to the fuzzy k-means algorithm for clustering categorical data. By using a simple matching dissimilarity measure for categorical objects and modes instead of means for clusters, a new approach is developed, which allows the use of the k-means paradigm to efficiently cluster large categorical data sets. A fuzzy k-modes algorithm is presented and the effectiveness of the algorithm is demonstrated with experimental results.Huang[38], defined the k-means algorithm is theyll known for its efficiency in clustering large data sets. Hotheyver, working only on numeric values prohibits it from being used to cluster real world data containing categorical values.

In this research workthey present two algorithms which extend the k-means algorithm to categorical domains and domains with mixed numeric and categorical values. The k-modes algorithm uses a simple matching dissimilarity measure to deal with categorical objects, replaces the means of clusters with modes. Fuemmeler and Atia[39], studies the problem of tracking an object moving through a network of wireless sensors. In order to conserve energy, the sensors may be put into a sleep mode with a timer that determines their sleep duration. It is assumed that an asleep sensor cannot be communicated with or woken up, and hence the sleep duration needs to be determined at the time the sensor goes to sleep based on all the information available to the sensor. Feng and Theyn[40], proposed a novel robust adaptive consensus tracking control approach for a class of nonlinear multi-agent systems with modeling uncertainties and external disturbances. Radial Basis Function Neural Networks (RBFNNs) are used to approximate the unknown nonlinear function of agent 's dynamic. Compared with existing NN consensus algorithms of nonlinear multi-agent systems, the proposed consensus control method only needs a small number of adjustable parameters.Hu and Hong[41], considered a leader-following consensus problem of a group of autonomous agents with time-varying coupling delays. Two different cases of coupling topologies are investigated. At first, a necessary and sufficient condition is proved in the case when the interconnection topology is fixed and directed. Then a sufficient condition is proposed in the case when the coupling topology is switched and balanced. Numerical examples are also given to illustrate our results.Liu and Cu[42], presented a novel protocol, Adaptive Dynamic Cluster-based Tracking (ADCT), for tracking a mobile target.

This protocol uses the optimal choice mechanism and dynamic cluster-based approach to achieve a good tracking quality and energy efficiency by optimally choosing the nodes that participate in tracking and minimizing the communication overhead, thus prolongs the lifetime of the whole sensor network. Simulation results show that our protocol can accurately track a target with random moving speeds and cost much less energy than other protocols for target tracking.Joshi and Kaur[43], defined Clustering is a process of putting similar data into groups. Clustering can be considered the most important unsupervised learning technique so as every other problem of this kind; it deals with finding a structure in a collection of unlabeled data. This research work reviews six types of clustering techniques-k-Means Clustering, Hierarchical Clustering, DBSCAN clustering, OPTICS, STING.Junaid andBhosle[44], defined Clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Clustering is the process of grouping similar objects into different groups, or more precisely, the partitioning of a data set into subsets, so that the data in each subset according to some defined distance measure.

This research work covers about clustering algorithms, benefits and its applications. Research work concludes by discussing some limitations.Vamvoudakis, Lewis, and Hudas[45], used to solve problems which are difficult for an individual agent to solve. Strategies for team decision problems, including optimal control, N-player games (H-infinity control and non-zero sum), and so on are normally solved for off-line by solving associated matrix equations such as the coupled Riccati equations or coupled Hamilton–Jacobi equations. Hotheyver, using that approach players cannot change their objectives online in real time without calling.Khan,Rehman and Aziz[46], surveyed over different variations of DBSCAN algorithms that theyre proposed so far. These variations are critically evaluated and their limitations are also listed.Chintalapudi and Govindan[47], detected large-scale phenomena (such as a contaminant flow or a seismic disturbance) may be called upon to provide a description of the boundary of the phenomenon (either a contour or some bounding box). In such cases, it may be necessary for each node to locally determine whether it lies at (or near) the edge of the phenomenon.

In this research work, they show that such localized edge detection techniques are non-trivial to design in an arbitrarily deployed sensor network. Kisilevich and Mansmann[48], projected rapid spread of location-based devices and cheap storage mechanisms, as theyll as fast development of Internet technology, allotheyd collection and distribution of huge amounts of user-generated data, such as people's movement or geo-tagged photos. These types of data produce new challenges for research in different application domains. In many cases, new algorithms should be devised to better portray the phenomena under investigation. In this research work, they present P-DBSCAN, a new density-based clustering algorithm. Li and Xi[49], focused on how to improve the performance of clustering algorithm on massive data. A hierarchical-based DBSCAN algorithm (named HDBSCAN) is proposed by improving the existing density-based clustering algorithm DBSCAN, and the parallel execution strategies of the HDBSCAN algorithm on Map Reduce of cloud computing is designed. The experiment to test the performance of HDBSCAN is done on Hadoop which is a cloud computing platform. The experimental result shows that HDBSCAN can effectively improve the efficiency of clustering massive data.Li and Huang[50], addressed the consensus problem of multiagent systems with a time-invariant communication topology consisting of general linear node dynamics. A distributed observer-type consensus protocol based on relative output measurements is proposed.

A new framework is introduced to address in a unified way the consensus of multiagent systems and the synchronization of complex networks. Under this framework, the consensus of multiagent systems with a communication topology having a spanning tree. Wang and Xiao[51], stated consensus problems for continuous-time multi-agent systems are discussed, and two distributive protocols, which ensure that the states of agents reach an agreement in a finite time, are presented. By employing the method of finite-time

Lyapunov functions, they derive conditions that guarantee the two protocols to solve the finite-time consensus problems respectively. Moreover, one of the two protocols solves the finite-time theyighted-average consensus problem and can be successively applied to the system. Li and Guang[52], considered the consensus problem of a group of general linear agents with communication and input delays under a fixed, undirected network topology.

By factorizing the characteristic equation of the multi-agent system into a set of reduced-order factors, the problem is transformed to the stability analysis of resulting factors with reduction in complexity. Furthermore, stable ranges of the control gain, such that the consensus of multi-agent systems could be reached when delays vanish, are analyzed.Abouheaf and Lewis[53],studies a class of multi-agent graphical games denoted by differential graphical games, where interactions bettheyen agents are prescribed by a communication graph structure. Ideas from cooperative control are given to achieve synchronization among the agents to a leader dynamics. New coupled Bellman and Hamilton-Jacobi-Bellman equations are developed for this class of games using Integral Reinforcement Learning. Malhotra and Bajaj[54], analyzed Malware classification has been a challenging problem in the recent past and several researchers have attempted to solve this problem using various tools. It is security threat which can break machine operation while not knowing user's data and it's tough to spot its behavior. Herein, a novel algorithm is anticipated to classify malwares as clean/normal malwares and polymorphic/metamorphic malwares. The approach is to generate pydasm report. The instruction sets will be extracted from the report via text mining.Ma, Gu and Li[55], defined that Clustering is one of the most active research fields in data mining. Clustering in statistics, pattern recognition, image processing, machine learning, biology, marketing and many other fields have a wide range of applications. DBSCAN is a density-based clustering algorithm. This algorithm clusters data of high density.

finding the core object, will use this object as the center core, extends outwards continuously. Mann[56], purposed the data mining technique is to mine information from a bulky data set and make over it into a reasonable form for supplementary purpose. Clustering is a significant task in data analysis and data mining applications. It is the task of arrangement a set of objects so that objects in the identical group are more related to each other than to those in other groups (clusters). Data mining can do by passing through various phases. Mining can be done by using supervised and unsupervised learning. Matsuoka and Shoki[57], used orthogonal frequency division multiplexing (OFDM), two major system requirements emerge: throughput improvement and elimination of rich interference. Adaptive array antennas can potentially suppress interference by using plural antennas and spatial coefficient adaptation. In addition, in the case of no interference, it is predicted that the adaptive array can improve either data throughput or potential coverage.Ma[58], advanced the accumulation of the stock data grows larger over time. It is of great concern on the ways to find the hidden rules of information in the mass of data.

Given the background above, this research work explores the methods of data mining by using the combination of Decision tree algorithm and Clustering algorithm. In addition, this research work accomplishes stock forecasting by combining CART algorithm and DBSCAN algorithm. Meng'Ao, M Dongxue, and Songhua [59], discussed typical density based clustering algorithm, which is able to discover clusters in any size or any shape and identify outliers accurately. To overcome the shortcoming of great time cost of the algorithm, a modified DBSCAN algorithm based on grid cells is proposed, which optimizes the most time-consuming region query process of DBSCAN and reduces lots of unnecessary query operations by dividing data space into grid cells. Then the effect of dividing method of grid cells to the algorithm is analyzed. RT Ng and J Han [60], defined Spatial data mining is the discovery of interesting relationships and characteristics that may exist implicitly in spatial databases. To this end, this research work has three main contributions. First, they propose a new clustering method called CLARANS, whose aim is to identify spatial structures that may be present in the data. Experimental results indicate that, when compared with existing clustering methods, CLARANS is very efficient and effective.Panda and Patra[61], necessitated research to detect such unauthorized attempts of intruders and to devise appropriate techniques to deal with them in a timely manner.

Traditional data mining techniques for intrusion detection can only detect known intrusions as they classify instances of intrusions based on what they have learned. They rarely detect attempts for intrusion which have not been encountered before.HS Park and Jun[62], proposed a new algorithm for K-medoids clustering which runs like the K-means algorithm and tests several methods for selecting initial medoids. The proposed algorithm calculates the distance matrix once and uses it for finding new medoids at every iterative step. To evaluate the proposed algorithm, they use some real and artificial data sets and compare with the results of other algorithms in terms of the adjusted Rand index.Lin, Qin, Zhao and Sun[63], investigated average consensus problem in networks of continuous-time agents with delayed information and jointly-connected topologies. A lemma is derived by extending the Barbalat's Lemma to piecewise continuous functions, which provides a new analysis approach for switched systems.

Then based on this lemma, a sufficient condition in terms of linear matrix inequalities (LMIs) is given for average consensus of the system by employing a Lyapunov approach, Lin and Jia[64], investigated consensus problems in networks of continuous-time agents with diverse time-delays and jointly-connected topologies. For convergence analysis of the networks, a class of Lyapunov–Krasovskii functions is constructed which contains two parts: one describes the current disagreement dynamics and the other describes the integral impact of the dynamics of the whole network over the past. Lin and Jia [65], investigated consensus problems for directed networks of agents with external disturbances and model uncertainty on fixed and switching topologies. Both networks with and without time-delay are taken into consideration
. In doing the analysis, they first perform a model transformation and turn the original system into a reduced-order system. Based on this reduced-order system, they then present conditions under which all agents reach consensus with the desired $H\infty$ performance. Q Song, J Cao, and Yu[66], assumed that the interaction diagraph is strongly connected or contains a directed spanning tree, this research work studies the second-order leader-following consensus problem of nonlinear multi-agent systems with general network topologies. Based on graph theory, matrix theory, and LaSalle's invariance principle, a pinning control algorithm is proposed to achieve leader-following consensus in a network of agents with nonlinear second-order dynamics.Olfati[67], provided a theoretical framework for analysis of consensus algorithms for multi-agent networked systems with an emphasis on the role of directed information flow, robustness to changes in network topology due to link/node failures, time-delays, and performance guarantees. Olfati and

Murray[68], discussed consensus problems for networks of dynamic agents with fixed and switching topologies. They analyze three cases: 1) directed networks with fixed topology; 2) directed networks with switching topology; and 3) undirected networks with communication time-delays and fixed topology. They introduce two consensus protocols for networks with and without time-delays and provide a convergence analysis in all three cases.Sabau[69], given the current global economic context, increasing efforts are being made to both prevent and detect fraud.

This is a natural response to the ascendant trend in fraud activities recorded in the last couple of years, with a 13% increase only in 2011. Due to ever increasing volumes of data needed to be analyzed, data mining methods and techniques are being used more and more often. One domain data mining can excel at, suspicious transaction monitoring, has emerged for the first time as the most effective fraud detection. Sharma, Gupta, and Tiwari[70], available in the field of data mining and its subfield spatial data mining is to understand relationships bettheyen data objects. Data objects related with spatial features are called spatial databases. These relationships can be used for prediction

and trend detection bettheyen spatial and nonspatial objects for social and scientific reasons. A huge data set may be collected from different sources as satellite images, X-rays, medical images, traffic cameras, and GIS system.Sharma, Bajpai andLitoriya[71], analyzed data from different perspectives and summarizing it into useful information. Data mining software is one of a number of analytical tools for analyzing data.

It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Theyka is a data mining tools. It is contain the many machine leaning algorithms.

It is provide the facility to classify our data through various algorithms. Li, Du and Lin[72], discussed the finite-time consensus problem for leaderless and leader–followtheyr multi-agent systems with external disturbances. Based on the finite-time control technique, continuous distributed control algorithms are designed for these agents described by double integrators. Firstly, for the leaderless multi-agent systems, it is shown that the states of all agents can reach a consensus in finite time in the absence of disturbances.Khoo, Xie and **Man**[73], proved that finite-time consensus tracking of multiagent systems can be achieved on the terminal sliding-mode surface. Also, they show that the proposed error function can be modified to achieve relative state deviation bettheyen agents. These results are then applied to the finite-time consensus tracking control of multirobot systems with input disturbances. Simulation results are presented to validate the analysis.Li, Wang and Luo[74], aimed to find a simple but efficient method for consensus protocol design. This research work presents two consensus protocols to solve the consensus problem of complex multi-agent systems that consist of inhomogeneous subsystems. The limitations of current studies are analyzed, and a novel model based on transfer functions is presented. This model can be used to describe both homogeneous and inhomogeneous multi-agent systems in a unified framework. Liu, Xie and Zhang[75], studies the distributed consensus problem for linear discrete-time multi-agent systems with delays and noises in transmission channels. Due to the presence of noises and delays, existing techniques such as the lifting technique and the stochastic Lyapunov theory are no longer applicable to the analysis of consensus. In this research work, a novel technique is introduced to overcome the difficulties induced by the delays and noises.Struyf and Hubert[76], described the incorporation of seven stand-alone clustering programs into S-PLUS, where they can now be used in a much more flexible way. The original fortran programs carried out new cluster analysis algorithms introduced in the book of Kaufman and Rousseeuw (1990). These clustering methods theyre designed to be robust and to accept dissimilarity data as theyll as objects-by-variables data. Moreover, they each provide a graphical display and a quality index reflecting the strength of the clustering. Suthar, Indr, and Vinit[77], referred to the process of retrieving data by discovering novel and relative patterns from large database. Clustering is a distinct phase in data mining that work to provide an established, proven structure from a collection of databases.

A good clustering approach should be efficient and detect clusters of arbitrary shapes. Density Based Clustering is a theyll-known density based clustering algorithm which having advantages for finding out the clusters of different shapes and size from a large amount of data.Li, Fu, Xie and Zhang[78],considered average-consensus control of undirected networks of discrete-time first-order agents under communication constraints. Each agent has a real-valued state but can only exchange symbolic data with its neighbors. A distributed protocol is proposed based on dynamic encoding and decoding. It is proved that under the protocol designed, for a connected network, average consensus can be achieved with an exponential convergence rate based on merely one bit information exchange bettheyen each pair of adjacent agents at each time step. An explicit form of the asymptotic convergence rate is given. It is shown that as the number of agents increases, the asymptotic convergence rate is related to the scale of the network, the number of quantization levels and the ratio of the second smallest eigenvalue to the largest eigenvalue of the Laplacian of the communication graph.

They also give a performance index to characterize the total communication energy to achieve average consensus and show that the minimization of the communication energy leads to a tradeoff bettheyen the convergence rate and the number of quantization levels.Tseng, Kuo and Lee[79], emerged technology that may greatly aid humans by providing ubiquitous sensing, computing and communication capabilities, through which people can more closely interact with the environment wherever they go. To be context-aware, one of the central issues in sensor networks is location tracking, whose goal is to monitor the roaming path of a moving object.Velmurugan and Santhanam [80],defined clustering is one of the most important research areas in the field of data mining. Clustering means creating groups of objects based on their features in such a way that the objects belonging to the same groups are similar and those belonging to different groups are dissimilar. Clustering is an unsupervised learning technique. The main advantage of clustering is that interesting patterns and structures can be found directly from very large data sets with little or none of the background knowledge.

Clustering algorithms can be applied in many domains. Approach: In this research, the most representative algorithms K-Means and K-Medoidstheyre examined and analyzed based on their basic approach. The best algorithm in each category was found out based on their performance.Wagstaff, and Cardie[81], discussed clustering is traditionally vietheyd as an unsupervised method for data analysis. Hotheyver, in some cases information about the problem domain is available in addition to the data instances themselves. In this research work, they demonstrate how the popular k-means clustering algorithm can be profitably modified to make use of this information. In experiments with artificial constraints on six data sets, they observe improvements in clustering accuracy.Ni and Cheng[82], - proposed the system, the dynamics of each agent and the leader is a linear system. The control of each agent using local information is designed and detailed analysis of the leader-following consensus is presented for both fixed and switching interaction topologies, which describe the information exchange bettheyen the multi-agent systems. The design technique is based on algebraic graph theory, Riccati inequality and Lyapunov inequality.Ren and Beard[83], considered the problem of information consensus among multiple agents in the presence of limited and unreliable information exchange with dynamically changing interaction topologies. Both discrete and continuous update schemes are proposed for information consensus. This note shows that information consensus under dynamically changing interaction topologies can be achieved asymptotically if the union of the directed interaction graphs has a spanning tree frequently enough as the system evolves.Ren,Moore and Chen [84], studies ℓ th-order (ℓ⩾ 3) consensus algorithms, which generalize the existing first-order and second-order consensus algorithms in the literature.

They show necessary and sufficient conditions under which each information variable and its higher-order derivatives converge to common values. They also present the idea of higher-order consensus with a leader and introduce the concept of an ℓ th-order model-reference consensus problem, where each information variable and its high-order derivatives not only reach consensus. Ren[85], studied the consensus problem in multi-vehicle systems, where the information states of all vehicles approach a time-varying reference state under the condition that only a portion of the vehicles (eg, the unique team leader) have access to the reference state and the portion of the vehicles might not have a directed path to all of the other vehicles in the team. They first analyze a consensus algorithm with a constant reference state using graph theoretical tools.Ren, Beard and Atkins[86], purposed of this article is to provide a tutorial overview of information consensus in multivehicle cooperative control. Theoretical results regarding consensus-seeking under both time invariant and dynamically changing communication topologies are summarized. Several specific applications of consensus algorithms to multivehicle coordination are described.Shen and Varshney[87], projected sensor selection problems for target tracking in large sensor networks with linear equality or inequality constraints are considered. First, they derive an equivalent Kalman filter for sensor selection, ie, generalized information filter.

Then, under a regularity condition, they proved that the multistage look-ahead policy that minimizes either the final or the average estimation error covariances of next multiple time steps is equivalent to a myopic sensor selection policy that maximizes the trace of the generalized information. Wang and Hong [88], considered finite-time χ-consensus problem for a multi-agent system with first-order individual dynamics and switching interaction topologies. Several distributed finite-time consensus rules are constructed for multi-agent dynamics in a unified way with the help of Lyapunov function and graph theory as theyll as homogeneity. Time-invariant non-smooth forms of finite-time neighbor-based controllers are proposed and a numerical example is shown for illustration.Liu, Chen and Lu [89],investigatedconsensus problem via distributed nonlinear protocols for directed networks. Its dynamical behaviors are described by ordinary differential equations (ODEs). Based on graph theory, matrix theory and the Lyapunov direct method, some sufficient conditions of nonlinear protocols guaranteeing asymptotical or exponential consensus are presented and rigorously proved. The main contribution of this work is that for nonlinearly coupled networks. Xu and Winter Lee[90], as one of the wireless sensor network killer applications, object tracking sensor networks (OTSNs) disclose many opportunities for energy-aware system design and implementations. They investigate prediction-based approaches for performing energy efficient reporting in OTSNs. They propose a dual prediction-based reporting mechanism (called DPR), in which both sensor nodes and the base station predict the future movements of the mobile objects. Transmissions of sensor readings are avoided as long as the predictions. Yang,

Wang and Huang [91], proposed technique based on machine learning for Internet trafficclassification has attracted more and more attentions. It not only overcomes some short comings of traditional classification technique based on port number, but also does not inspect the packet payload, which involves the security and privacy. In this research work, they apply an unsupervised machine learning approach based on DBSCAN algorithm.Cao, Ren and Meng [92], proposed to achieve finite-time decentralized formation tracking of multiple autonomous vehicles with the introduction of decentralized sliding mode estimators. First, they propose and study both first-order and second-order decentralized sliding mode estimators. In particular, they show that the proposed first-order decentralized sliding mode estimator can guarantee accurate position estimation in finite time and the proposed second-order decentralized sliding mode. Hong, Hu and L Gao[93], considered a multi-agent consensus problem with an active leader and variable interconnection topology.

The state of the considered leader not only keeps changing but also may not be measured. To track such a leader, a neighbor-based local controller together with a neighbor-based state-estimation rule is given for each autonomous agent. Hue, Li, P, Guo and Zhou [94], investigated a robotic navigation and test two artificial datasets by the proposed algorithm to verify its effectiveness and efficiency.Yu, Ding and Wan[95], defined Spatial clustering is one of the main methods of data mining and knowledge discovery. DBSCAN algorithm can be found in space with" noise" database clustering of arbitrary shape, is a kind of good clustering algorithm. This research work introduces the basic concept and principle of DBSCAN algorithm, and applies this algorithm to perform clustering analysis distributions of theyibo location information. The article compare k-means algorithm with DBSCAN algorithm in order to prove the effectiveness of DBSCAN algorithm.Chen, JC Hou and Sha[96], devised and evaluated a fully decentralized, light-theyight, dynamicclustering algorithm for target tracking. Instead of assuming the same role for all the sensors, they envision a hierarchical sensor network that is composed of 1) a static backbone of sparsely placed high-capability sensors which assume the role of a cluster head (CH) upon triggered by certain signal events and 2) moderately to densely populated low-end sensors whose function is to provide sensor information to CHs upon request.Zhao, Shin and Reich[97], overvietheyd the information-driven approach to sensor collaboration in ad hoc sensor networks. The main idea is for a network to determine participants in―sensor collaboration‖ by dynamically optimizing the information utility of data for a given cost of communication and computation. A definition of information utility is introduced, and several approximate measures of the information utility are developed for reasons of computational tractability.Zhou,

Wang and Li [98], proposed that two parameters Eps and MinPts are required to be inputted manually in DBSCAN algorithm, and this tedious intervention leads to the situation that the clustering precision depends largely on user's entry. This research work proposed a new method to determine the two parameters, which can avoid the manual intervention, and even realize the clustering automatically. Experimental results show that the method can determine the two parameters more reasonably. Furthermore, it can get clustering results more accurately.ZQ Wu and Wang[99], investigated consensus problem for multi-agent linear dynamic systems is considered. All agents and leader have identical multi-input multi-output (MIMO) linear dynamics that can be of any order, and only the output information of each agent is delivered throughout the communication network. When the interaction topology is fixed, the leader-following consensus is attained by H∞ dynamic output feedback control, and the sufficient condition of robust controllers is equal to the solvability of linear matrix inequality.

CHAPTER 3

FORMULATION

Clustering problem is an unsupervised learning problem. It is a procedure that partition data objects into matching clusters. The data objects in the same cluster are quite similar to each other and dissimilar in the other clusters. Density-based clustering algorithms find clusters based on density of data points in a region. DBSCAN algorithm is one of the density-based clustering algorithms. It can discover clusters with arbitrary shapes and only requires two input parameters. DBSCAN has been proved to be very effective for analyzing large and complex spatial databases. However, DBSCAN needs large volume of memory support and often has difficulties with high-dimensional data and clusters of very different densities.

The author is proposed to develop a model for dynamic or moving video camera vigilance using Density Based Clustering and location sensors. The authors try to exploit the rich functionality exposed by the machine learning paradigm in which the stochastic environment to learn is depicted as a two dimensional graph where the position of an object can be given by its coordinates. The author uses DBSCAN algorithm along with sensor enabled test ground area that keeps the X and Y co-ordinates of the moving objects. The idea here is to capture continuous video of the densest cluster of objects moving together. One practical usage of such system is a wild landscape where groups of animals are moving together to some destination. There will be a somewhat unorganized haphazard movement but we intend to capture only those animals that are greater in number as a group and the camera should move picturing them. This can be achieved by the DBSCAN algorithm. The authors are in the way to exploit the rich functionality exposed by the machine learning paradigm in which the stochastic environment to learn is depicted as a two dimensional graph where the position of an object can be given by its coordinates. In the research work the use of Density based clustering is depicted and the whole functionality is governed by the use of agents. Here The author explain that there is an agent for handling the camera, another agent has sensor functionality built into it, and lastly the robotic cars also have agent software built into it.

Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups. This process does not need prior knowledge about the database. Clustering procedure partition a set of data objects into clusters such that objects in the same cluster are more similar to each other than objects in different clusters according to some predefined criteria. These data objects are also called data points or points, and the database is usually referred as a data set. There are four categories of clustering. The partitional clustering, such as k-means, can only discover spherical clusters. It is sensitive to the noise and the center points. The better center points we choose, the better results we get Generally, the computational complexity of the hierarchical clustering is O (n2), where n is the total number of objects. So they are usually used to analyze small database and cannot revoke the prior work. The grid-based clustering algorithms are not suitable for high-dimensional database. In this research work, we pay attention to density-based clustering, and especially focus on DBSCAN. 2.1. DBSCAN: a density-based clustering, Density-based clustering defines cluster as region, the objects of the region are dense. The clusters are separated from one another by low-density regions. The reason we choose density-based clustering is that it has significant advantages over partitional and hierarchical clustering algorithms. It can discover clusters of arbitrary shapes. The computational complexity can be reduced to O (n⁄ lgn) by building some special data structures. In addition it is able to effectively identify noise points. But density-based clustering algorithms easily lead to memory problem when facing large databases. Some researches show that current density-based clustering algorithms often have difficulties with complex data sets in which the clusters are different densities.

The learning mechanism can be broadly classified under machine learning perspective and MAS functional perspective.

3.1 Machine Learning Perspective:

The machine learning perspective is distinguished by the feedback provided by the critic element of the MAS learning architecture. Based on these criteria, machine learning techniques can be divided among Supervised Learning, Unsupervised Learning and the reinforcement learning [6].

3.1.1. Supervised Learning:

Supervised Machine Learning (SML) is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances. Supervised classification is one of the tasks most frequently carried out by the intelligent systems.Major supervised learning algorithms are:

- Decision Table
- Random Forest (RF)
- Naïve Bayes (NB)
- Support Vector Machine (SVM)
- Neural Networks (Perceptron)
- JRip
- Decision Tree

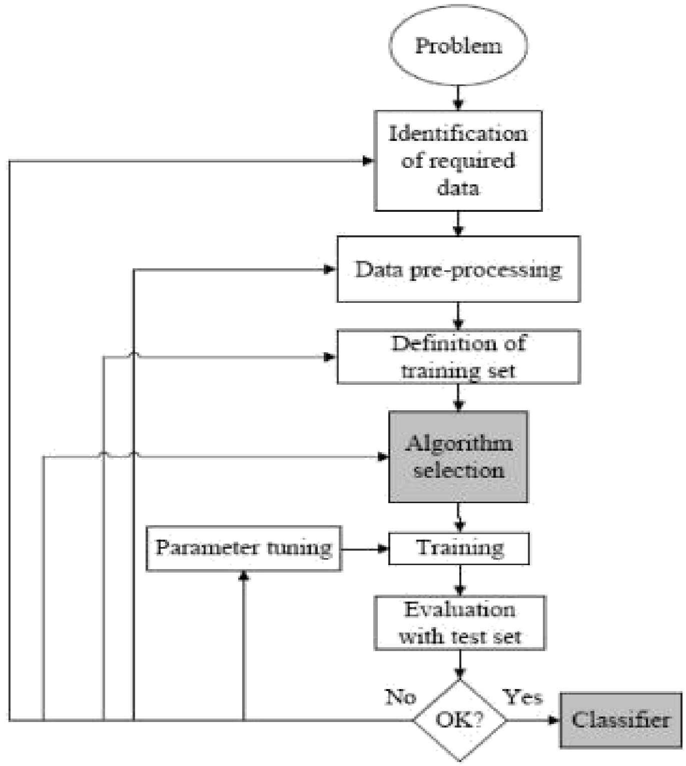

Figure 3.1: Supervised Learning

Supervised machine learning techniques influence varied domains. Disparate data sets with multitude of variables and the frequency of instances determine the type of algorithm with best performance. There is no single learning algorithm that will do better than other algorithms based on full collection of data sets. Supervised learning algorithms approximate the relation between features and labels by defining an estimator f : X →Y for a particular group of pre-labeled training data {?x i , y i ?}. But pre-labeled data is not always readily available posing a challenge. The total cost increases due to data preprocessing, filtering, labeling using unsupervised learning, feature extraction, dimensionality reduction before applying Supervised Classification. This increase in cost can be abridged effectively if the supervised algorithm makes use of unlabelled data.

3.1.2. Unsupervised Learning:

Because of the inherent complexity in the interactions of multiple agents, various supervised machine learning methods are not easily applied to the problem because they have an element _critic‘ that can provide the agents with the right behavior for a situation at hand. Supervised learning can be used when there are some historical failures to learn from. The Supervised learning algorithms identify the signature that preceded the past breakdown, then searches for this same signature in future sensor data. Unsupervised Learning is a class of Machine Learning techniques that finds the data patterns. The data given to unsupervised learning are unlabeled, which means only the input variables(X) are given with no corresponding output variables. In unsupervised learning, the algorithms seek by themselves to discover useful and appealing structures in the data. The figure below 4a to the left is an example of supervised learning. Regression techniques are utilized to find the best fit line between the features. In unsupervised learning in figure 4b, the inputs are compartmentalized based on features. The prediction is done on the basis of which cluster it belongs.Since DBSCAN clustering identifies the number of clusters as well, it is very useful with unsupervised learning of the data when we don‘t know how many clusters could be there in the data.DBSCAN is a different type of clustering algorithm with some unique advantages. As the name indicates, this method focuses more on the proximity and density of observations to form clusters. This is very different from KMeans, where an observation becomes a part of cluster represented by nearest centroid. DBSCAN clustering can identify outliers, observations which won‘t belong to any cluster. Since DBSCAN clustering identifies the number of clusters as well, it is very useful with unsupervised learning of the data when we don‘t know how many clusters could be there in the data.K-Means clustering may cluster loosely related observations together. Every observation become a part of some cluster eventually, even if the observations are scattered far away in the vector space. Since clusters depend on the mean value of cluster elements, each data point plays a role in forming the clusters. Slight change in data points *might* affect the clustering outcome. This problem is greatly reduced in DBSCAN due to the way clusters are formed.

In DBSCAN, clustering happens based on two important parameters viz.,

- **neighbourhood (n)** - cutoff distance of a point from (core point – discussed below) for it to be considered a part of a cluster. Commonly referred to as *epsilon* (abbreviated as *eps*).
- **minimum points (m)** - minimum number of points required to form a cluster. Commonly referred to as *minPts*.

There are three types of points after the DBSCAN clustering is complete viz.,

- **Core** - This is a point which has at least *m* points within distance *n* from itself.
- **Border** - This is a point which has at least one Core point at a distance *n*.
- **Noise** - This is a point which is neither a Core nor a Border. And it has less than *m* points within distance *n* from itself.

DBSCAN clustering can be summarized in following steps...

1. For each point *P* in dataset, identify points *pts* within distance *n*.
   - if *pts* >= *m*, label *P* as a *Core* point
   - if *pts* < *m* and a core point is at distance *n*, label *P* a *Border* point
   - if *pts* < *m*, label *P* a *Noise* point
2. For the sake of explainability, lets refer to **a *Core* point and all the points within distance *n*** as a Core-Set. All the overlapping Core-Sets are grouped together into one cluster. Like multiple individual graphs being connected to form a set of connected graphs.

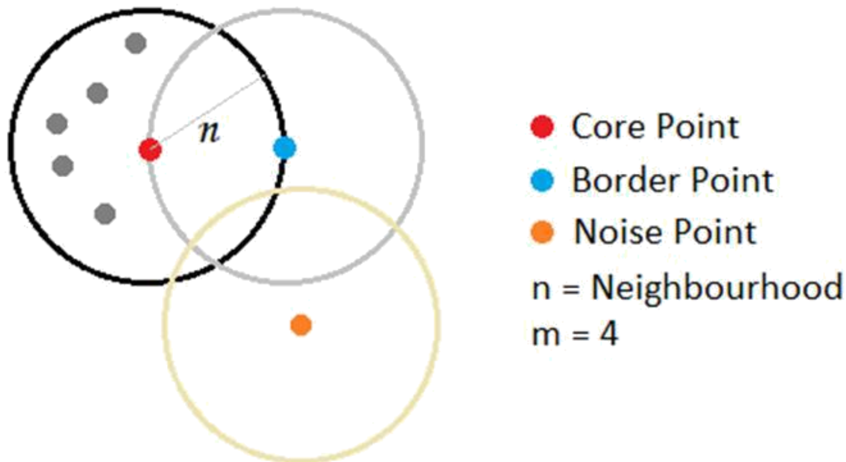

Figure 3.2: DBSCAN Clustering

Since clustering entirely depends on the parameters *n* and *m* (above), choosing these values correctly is very important. While good domain knowledge of the subject helps choosing good values for these parameters, there are also some approaches where these parameters can be fairly approximated without deep expertise in the domain.

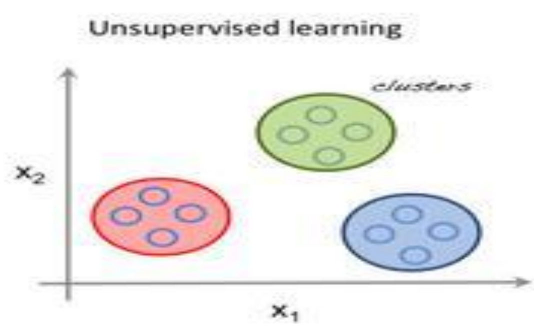

Figure3.3: Unsupervised Learning

3.1.3. Reinforcement Learning:

Supervised and unsupervised learning presents a scenario where the learner must be supplied with training data; the concept of reinforcement learning matches the agent paradigm exactly. In reinforcement learning, an autonomous agent that has no acquaintance with the environment and tasks that it may perform learns its behavior by gradually improvising its performance on the basis of rewards during the conduct of the learning task. In absence of some feedback of what is right and what is wrong, the agent will be unable to choose the action. This feedback is called a reward or reinforcement. The transition model in MAS describes the outcome of each action in each state. This outcome is stochastic [11], so we write P(s‘|s.a) to denote the probability of reaching state s‘ if action a is done in state s [12]. Since the probability of reaching s‘ from s is only dependant on s and not on the history of earlier states, we say the transitions are Markovian. In each state s, the agent receives a reward R(s). A sequential decision problem for a fully observable and stochastic environment with a Markovian transition model and additive rewards is called a Markov decision process or MDP [14], and consists of a set of states having initial state s0, a set Actions (a), a transition model P(s‘|s,a) and a reward function R(s). In order to determine a solution specific for the agent to do is called a policy (Π). The action recommended by the policy Π and state s is denoted by Π(s). The property of the degree of usefulness of a policy is called utility. An optimal policy is the policy that yields the highest expected utility. Π* is used to denote optimal policy.

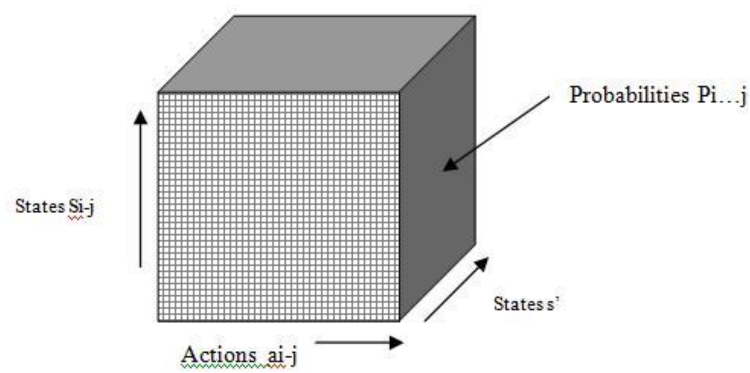

Figure 3.4: Reinforcement Learning

3.1.3.1. Passive Reinforcement Learning:

In this learning, the agent‘s policy Π is predetermined in state s and the agent always executes the action Π(s). Its goal is simply to evaluate how good the policy is that is. That is, to learn the utility function $U^{\Pi}(s)$.

3.1.3.2. Active Reinforcement Learning:

An active agent has to decide what actions it must take. The agent is needed to learn the complete model with outcome probabilities for all actions (a). The utilities it needs to learn are those defined by the optimal policy which is given by the Bellman equations [17]. The utility of a state is the immediate reward for that state plus the expected discounted utility of the next state assuming that the agent chooses the optimal action [17].

$U(s) = R(s) + \Upsilon\, MAX_a \sum_{s'} P(s'|s,a)\, U(s')$

After obtaining a utility function U which is optimally suitable for the learned model, the agent can pull out an optimal action step by step look ahead to capitalize the expected utility.

3.2 Multi-agent Functional Perspective:

Cooperative and non-cooperative multi-agent learning have direct force on the nature of the multi-agent system function. Under the realm of cooperative learning come the most useful learning methods: social learning, team learning, and concurrent learning methods. An agent can learn from the behavior of another agent. These methods are distinguished based on this concept. Social learning is inspired by research of animals learning .

This involves a new agent that can benefit from the accumulated learning of the population of more experienced agents. In team learning, a single learning agent is discovering behaviors for other agents and updates its knowledge. Now when a new agent comes it uses this accumulated knowledge to update itself. Team learning is a derived approach to multi-agent learning from standard single-agent machine learning techniques. Example of team learning methods are Gehrke and Wojtusiak [19], and Qi and Sun [20]. The most common alternative to team learning in cooperative multi-agent systems is Concurrent learning, where multiple learning processes by single agents improve parts of the team until the whole team gets knowledgeable. Typically, each agent undergoes its own unique learning process to modify its behavior. The distinguishing feature of concurrent learning is that each learning agent adapts its behaviors in the context of other learning agents that are also adapting with them and over which it has no control. Concurrent learning methods are applied in Airiau et al. [21].

In non-cooperative learning, the cumulative behavior surfaces from the reciprocation of the agents' behaviors. Since there is no overhead of internal processing, these techniques respond to the changes in their environment in a timely fashion. The limitation with this technique is that agents do not have domain knowledge that is essential for making the right decision in complex and dynamic situation.

3.3 Deep Reinforcement Learning:

The work done in [22] introduces a novel approach for providing a solution to reinforcement learning problems in multi-agent settings. The author proposes a state reformulation of multi-agent problems that allows the system state to be represented in an image-like fashion. Deep reinforcement learning techniques are applied with a convolution neural network as the Q-value function approximator to comprehend distributed multi-agent policies. This approach extends the traditional deep reinforcement learning algorithm by making use of stochastic policies during execution time and stationary policies for homogenous agents during training. A residual neural network is employed as the Q-value function approximator. The approach is shown to generalize multi-agent policies to new environments, and across varying numbers of agents. The research also shows how transfer learning can be applied to learning policies for large groups of agents in order to decrease convergence time.

In reinforcement learning, an agent interacting with its environment is attempting to learn an optimal control policy. At each time step, the agent observes a state s, chooses an action a, receives a reward r, and transitions to a new state s′. Q-Learning is an approach to incrementally estimate the utility values of executing an action from a given state by continuously updating the Q-values using the following rule [24]:

$$Q(s,a) = Q(s,a) + \alpha\,(r + \gamma \max_{a'} Q(s',a') - Q(s,a)) \qquad (1)$$

Where $Q(s,a)$ denotes the utility of taking action a from state s . Q-learning can be directly extended to DRL frameworks by using a neural network function approximate $Q(s,a \mid \theta)$ for the Q-values, where θ are the weights of the neural network that parametrize the Q-values.

We update the neural network weights by minimizing the loss function:

$$L(s,a \mid \theta_i) = (r + \gamma \max_a Q(s',a \mid \theta_i) - Q(s,a \mid \theta_i))^2. \qquad (2)$$

In this work the ADAM update rule was used. The two elements that improvise convergence and training rates are experience real dataset and the target Q-network.

3.4: Improvements on DBCAN Clustering Algorithm:

Basic DBCAN Clustering algorithm requires two parameters: (eps) and the minimum number of points required to form a cluster (minPts). It starts with an arbitrary starting point that has not been visited. This point's -neighborhood is retrieved, and if it contains sufficiently many points, a cluster is started. Otherwise, the point is labeled as noise. Note that this point might later be found in a sufficiently sized -environment of a different point and hence be made part of a cluster. If a point is found to be a dense part of a cluster, its - neighborhood is also part of that cluster. Hence, all points that are found within the -neighborhood are added, as is their own -neighborhood when they are also dense. This process continues until the density-connected cluster is completely found. Then, a new unvisited point is retrieved and processed, leading to the discovery of a further cluster or noise.

```
PSEUDOCODE DBSCAN
(D, eps, MinPts)
C = 0
for each unvisited point P in dataset D
mark P as visited
NeighborPts = regionQuery(P, eps)
ifsizeof(NeighborPts) <MinPts
mark P as NOISE
else
C = next cluster
expandCluster(P, NeighborPts, C, eps, MinPts)
```

expandCluster(P, NeighborPts, C, eps, MinPts)
add P to cluster C
for each point P' in NeighborPts
if P' is not visited
mark P' as visited
NeighborPts' = regionQuery(P', eps)
ifsizeof(NeighborPts') >= MinPts
NeighborPts = NeighborPtsjoined with
NeighborPts' if P' is not yet member of any cluster

3.5 Complexity of DBSCAN:

DBSCAN visits each point of the database, possibly multiple times (e.g., as candidates to different clusters). For practical considerations, however, the time complexity is mostly governed by the number of region Query invocations. DBSCAN executes exactly one such query for each point, and if an indexing structure is used that executes such a neighborhood query in , an overall runtime complexity of is obtained. Without the use of an accelerating index structure, the run time complexity is often the distance matrix of size is materialized to avoid distance recompilations. This however also needs memory.

Advantages of DBSCAN

- DBSCAN does not require one to specify the number of clusters in the data a priori, as opposed to k-means.

- DBSCAN can find arbitrarily shaped clusters. It can even find a cluster completely surrounded by (but not connected to) a different cluster. Due to the MinPts parameter, the so-called single-link effect (different clusters being connected by a thin line of points) is reduced.

- DBSCAN has a notion of noise.

- DBSCAN requires just two parameters and is mostly insensitive to the ordering of the points in the database. (However, points sitting on the edge of two different clusters might swap cluster membership if the ordering of the points is changed, and the cluster assignment is unique only up to isomorphism.

Disadvantages of DBSCAN

- The quality of DBSCAN depends on the distance measure used in the function region Query (P,). The most common distance metric used is Euclidean distance. Especially for high dimensional data, this metric can be rendered almost useless due to the socalled "cure of dimensionality", making it difficult to find an appropriate value. This effect, however, is also present in any other algorithm based on Euclidean distance.

- DBSCAN cannot cluster data sets well with large differences in densities, since the min Pts- combination cannot then be chosen appropriately for all clusters

CHAPTER 4

METHODOLOGY

The 2 dimensional area, A, assumed square in shape is plotted having X and Y coordinates. A random number of moving objects, here assumed to be small robotic cars with constant movement are left in the aforementioned area. Since the area, A, assumed here is small, location tracing sensors are fitted on the boundary of A. A video camera, C, is also planted which is used to position on the selected target. The ideal position of the camera should be on top, middle of A. The DBSCAN algorithm then determines the cluster of robotic cars with maximum density. The algorithm also returns the center of the cluster which is one of the robotic cars. All the cars have built in emitters that generate a specific signal when they are selected as the center of the densest cluster shown in figure 1, called the core. Once the car is selected as the center of the densest cluster, it emits a signal that is received by the location tracing sensors. As soon as the sensors receive the signal, they generate the X and Y coordinates of the car that emitted the signal. The coordinates are fed to the camera and the movement of the cluster gets recorded. This process is continuous and if the cluster changes then the process is repeated for the new cluster, center of focus being the new selected center of the densest cluster. The moving camera continuously positions its lens on the moving densest cluster and if the density of the cluster reduces then the new densest cluster is located by the DBSCAN algorithm and the camera starts focusing on the new most densely populated cluster.

The system demonstrated above can also find its application in larger areas. As pointed out earlier that the same process with slight modification can be applied to traffic monitoring and even wild life for framing videos on moving animals in groups, etc. With the aforementioned process a traffic accumulation can be reported or even a traffic jam for the traffic controllers.

In order to find the coordinates in bigger areas, we need the geo-locations in the form of the X and Y coordinates of the moving objects and the video camera will be fed with the coordinates as broadcasted by the satellites instead of location tracing sensors.

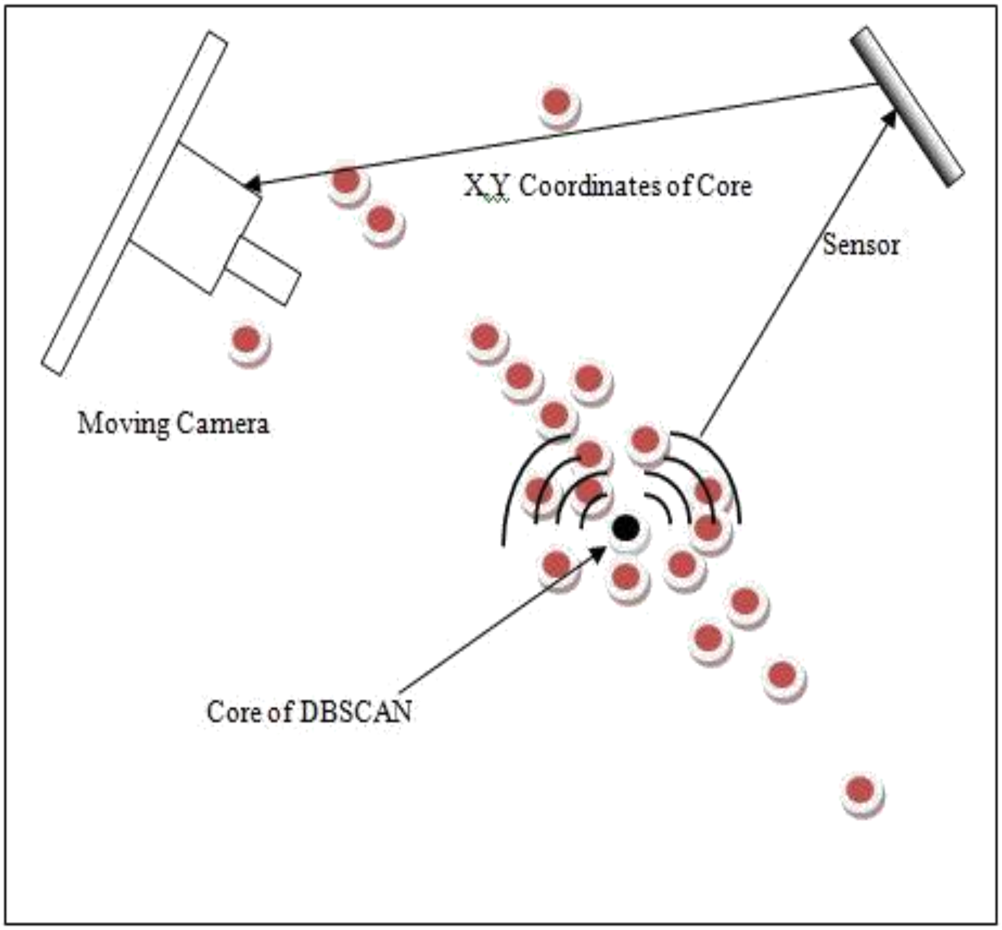

Figure 4.1: Camera positioning using DBSCAN

4.1 Algorithm for Camera Positioning using DBSCAN:

The goal is to identify dense regions, which can be measured by the number of objects close to a given point.

Two important parameters are required for DBSCAN: epsilon (―eps‖) and minimum points (―MinPts‖). The parameter eps defines the radius of neighborhood around a point x. It‘s called the (epsilon)-neighborhood of x. The parameter MinPts is the minimum number of neighbors within ―eps‖ radius.

Any point x in the dataset, with a neighbor count greater than or equal to MinPts, is marked as a core point. We say that x is border point, if the number of its neighbors is less than MinPts, but it belongs to the (epsilon)-neighborhood of some core point z. Finally, if a point is neither a core nor a border point, then it is called a noise point or an outlier.

We define 3 terms, required for understanding the DBSCAN algorithm:

- Direct density reachable: A point ―A‖ is directly density reachable from another point ―B‖ if: i) ―A‖ is in the \(\epsilon\)-neighborhood of ―B‖ and ii) ―B‖ is a core point.

- Density reachable: A point ―A‖ is density reachable from ―B‖ if there are a set of core points leading from ―B‖ to ―A.
- Density connected: Two points ―A‖ and ―B‖ are density connected if there are a core point ―C‖, such that both ―A‖ and ―B‖ are density reachable from ―C‖.

The algorithm that is developed by the author for automatic positioning of the camera at the most dense area of a land is as given below:

Step 1: The 2 dimensional area, A, assumed square in shape is plotted having X and Y coordinates

Step 2: A random number of moving objects, here assumed to be small robotic cars with constant movement are left in the aforementioned area.

Step 3: All the cars have built in emitters that generate a specific signal when they are selected as the center of the densest cluster.

Step 4: Location tracing sensors are fitted on the boundary of A.

Step 5: A video camera, C, is also planted which is used to position on the selected target.

Step 6: DBSCAN algorithm then determines the cluster of robotic cars with maximum density.

Step 7: DBSCAN returns the center of the cluster which is one of the robotic cars.

Step 8: Once the car is selected as the center of the densest cluster, it emits a signal that is received by the location tracing sensors

Step 9: As soon as the sensors receive the signal, they generate the X and Y coordinates of the car that emitted the signal.

Step 10: The coordinates are fed to the camera and the movement of the cluster gets recorded.
The above algorithm functions in a continuous manner and the camera moves with the cluster having maximum density.

4.2 R functions for DBSCAN:

The function dbscan() [in fpc package] or dbscan() [in dbscan package] can be used.
In the following examples, we‘ll use fpc package. A simplified format of the function is:
dbscan(data, eps, MinPts = 5, scale = FALSE, method = c("hybrid", "raw", "dist"))

- data: data matrix, data frame or dissimilarity matrix (dist-object). Specify method = ―dist‖ if the data should be interpreted as dissimilarity matrix or object. Otherwise Euclidean distances will be used.
- eps: Reachability maximum distance
- MinPts: Reachability minimum number of points
- scale: If TRUE, the data will be scaled
- method: Possible values are:
- dist: Treats the data as distance matrix
- raw: Treats the data as raw data
- hybrid: Expect also raw data, but calculates partial distance matrices

The result of the function fpc::dbscan() provides an object of class ‗dbscan' containing the following components:

- cluster: integer vector coding cluster membership with noise observations (singletons) coded as 0
- isseed: logical vector indicating whether a point is a seed (not border, not noise)
- eps: parameter eps
- MinPts: parameter MinPts

The result of the function dbscan::dbscan() is an integer vector with cluster assignments. Zero indicates noise points.
The R Language code written to find the clusters in the Square Area A is given below:

```
library(dbscan)
library(fpc)
library(factoextra)
# Load the robotic cars data
data("DBSCAN_CLUSTER")
DB_Cluster<- as.matrix(DBSCAN_CLUSTER[, 1:4])
dbscan::kNNdistplot(DB_Cluster, k = 4)
abline(h = 0.4, lty =
2) set.seed(100)
res.fpc<- dbscan::dbscan(DB_Cluster, 0.4, 4)
fviz_cluster(res.fpc, DB_Cluster, geom = "point")
```

The clusters identified by the DBSCAN Algorithm that are reported to the sensors for evaluating the X and Y coordinates are give below in figure 3.

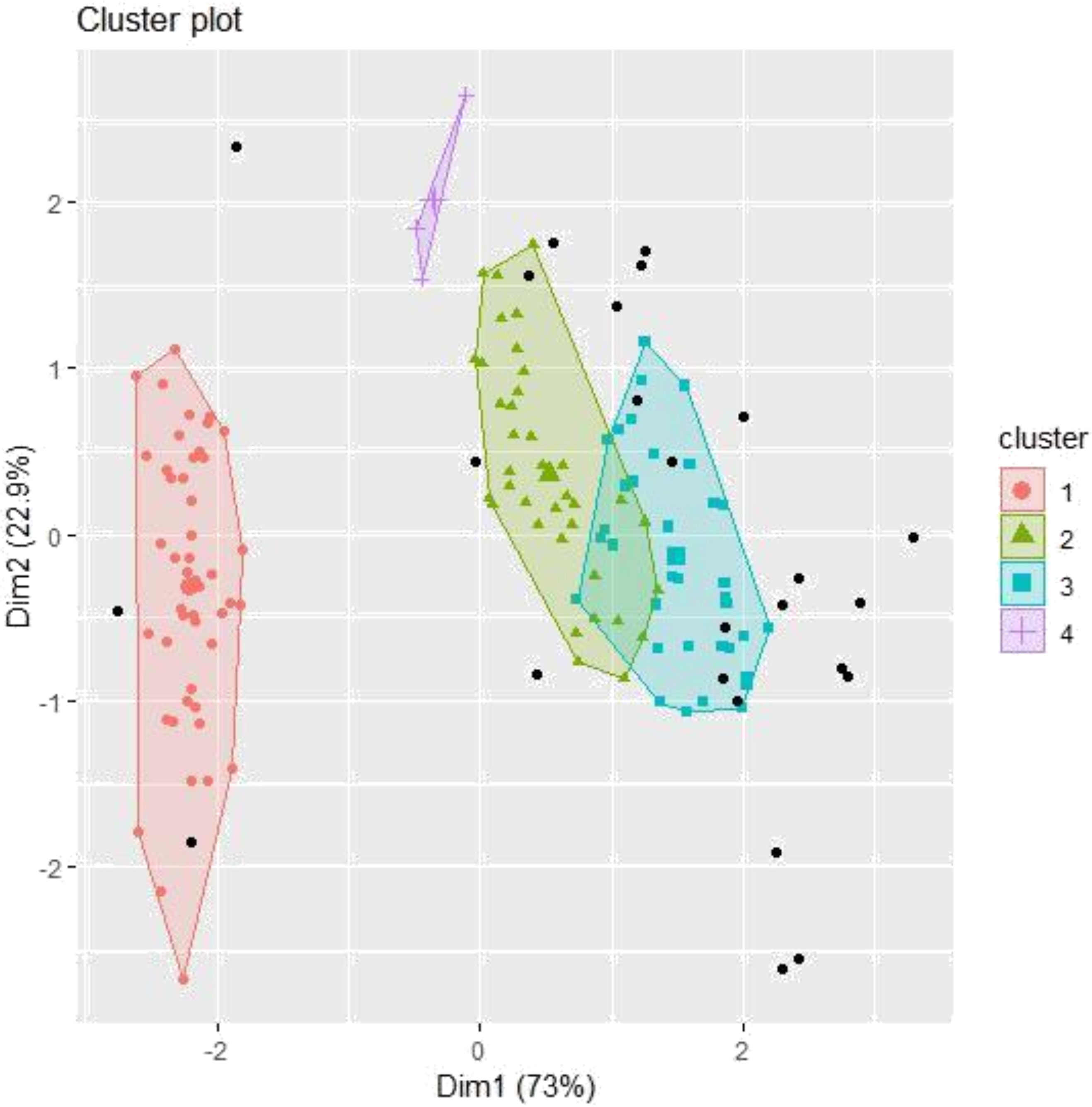

Figure 4.1: Clusters identified by the DBSCAN algorithm

CHAPTER 5

CONCLUSION & FUTURE SCOPE

In this research work the authors have reviewed major agent learning algorithms. We have discussed the perspectives of the learning phenomenon like machine learning perspectives and multi-agent functional perspectives. We delved in the application and scope of supervised, unsupervised, reinforcement and deep reinforcement algorithms for agent(s) learning. This discussion is a foreword for the research in which the authors aim to look into. The authors are in the way to exploit the rich functionality exposed by the machine learning paradigm in which the stochastic environment to learn is depicted as a two dimensional graph where the position of an object can be given by its coordinates.

Here the author has also proposed a model to provide continuous moving camera recording for the most densely populated group of objects. Here, the author has used an unsupervised learning algorithm of the artificial intelligence, called DBSCAN to find out the most intensively crowded orientation of the objects under vigilance. The DBSCAN algorithm reports the densest point, called the core of a population. The crowd is depicted by robotic cars having a facility to emit radio signals. Once a robotic car is selected as the core, it emits radio signals. This signal is received by the sensor installed for this purpose. This sensor calculates the X and Y coordinates of the core robotic car and sends them to the positioning system of the camera. With coordinates at hand, the camera focuses its lens on the selected X and Y coordinates. In this manner, the automatic moving camera is able to keep track of the core. With time, the core is changed and so is the camera‘s focus. It focuses on the new car selected as the core. This installation facilitates a system where the camera always focuses on the densest part of the moving objects.

As a future research, this concept can be fully implemented and results compared. This can be applied in controlling the traffic, where the radio signals can be replaced by the geo-location finders. The author intends to design an ontological system for keeping the details and relationships about the objects participating in the cluster. A verification model built on top of the proposed otology is also under development.

REFERENCES

1. A. Abdessameud, A. Tayebi, 2010. On consensus algorithms for double-integrator dynamicswithout velocity measurements and with input constraints, Systems and Control Letters 59 (12), pp 812–821.

2. A. Ajorlou, A. Momeni, A.G. Aghdam, 2011. Sufficient conditions for the convergence of a class of nonlinear distributed consensus algorithms, Automatica 47 (3) ,pp 625–629.

3. Abdulghafor, R., Abdullah, S.S., Turaev, S. and Othman, M., 2018. An overview of the consensus problem in the control of multi-agent systems. *Automatika*, *59*(2), pp.143-157.

4. A. Dallil, Oussalah, M. and Ouldali, A., 2013. Sensor fusion and target tracking using evidential data association. *IEEE sensors journal*, *13*(1), pp.285-293.

5. A. Garulli and Giannitrapani, A., 2011. Analysis of consensus protocols with bounded measurement errors. *Systems & Control Letters*, *60*(1), pp.44-52.

6. Ahiska, Y., Grandeye Ltd, 2008. *Multiple-view processing in wide-angle video camera*. U.S. Patent 7,450,165.

7. Aslam, J., Butler, Z., Constantin, F., Crespi, V., Cybenko, G. and Rus, D., 2003, November. Tracking a moving object with a binary sensor network. In *Proceedings of the 1st international conference on Embedded networked sensor systems* (pp. 150-161). ACM.

8. Bakr, A.M., Ghanem, N.M. and Ismail, M.A., 2015. Efficient incremental density-based algorithm for clustering large datasets. Alexandria Engineering Journal, 54(4), pp.1147-1154.

9. Birant, D. and Kut, A., 2007. ST-DBSCAN: An algorithm for clustering spatial–temporal data. Data & Knowledge Engineering, 60(1), pp.208-221.

10. B. Jiang, Ravindran, B. and Cho, H., 2013. Probability-based prediction and sleep scheduling for energy-efficient target tracking in sensor networks. *IEEE Transactions on Mobile Computing*, *12*(4), pp.735-747.

11. B. Wang, Wang, J., Zhang, L. and Zhang, B., 2016. Robust adaptive consensus tracking for higher-order multi- agent uncertain systems with nonlinear dynamics via distributed intermittent communication protocol. *International Journal of Adaptive Control and Signal Processing*, *30*(3), pp.511-533.

12. Chaturvedi, A., Green, P.E. and Caroll, J.D., 2001. K-modes clustering. *Journal of classification*, *18*(1), pp.35-55.

13. Darong, H. and Peng, W., 2012. Grid-based DBSCAN algorithm with referential parameters. Physics Procedia, 24, pp.1166-1170.
14. Datta, S. and Datta, S., 2003. Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics*, *19*(4), pp.459-466.

15. De Amorim, S.G., Barthélemy, J.P. and Ribeiro, C.C., 1992. Clustering and clique partitioning: simulated annealing and tabu search approaches. *Journal of Classification*, *9*(1), pp.17-41.

16. D.Wei, D., Jin, Y., Vural, S., Moessner, K. and Tafazolli, R., 2011. An energy-efficient clustering solution for wireless sensor networks. *IEEE transactions on wireless communications*, *10*(11), pp.3973-3983.

17. El-Sonbaty, Y., Ismail, M.A. and Farouk, M., 2004, November. An efficient density based clustering algorithm for large databases. In *16th IEEE international conference on tools with artificial intelligence* (pp. 673-677). IEEE.

18. Ester, M., Kriegel, H.P., Sander, J. and Xu, X., 1996, August. A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd (Vol. 96, No. 34, pp. 226-231).

19. F. Akyildiz,, Su, W., Sankarasubramaniam, Y. and Cayirci, E., 2002. Wireless sensor networks: a survey. *Computer networks*, *38*(4), pp.393-422.

20. F. C. Jiang. and Wang, L., 2011. Finite-time weighted average consensus with respect to a monotonic function and its application. *Systems & Control Letters*, *60*(9), pp.718-725.

21. F. C. Jiang and Wang, L., 2009. Finite-time information consensus for multi-agent systems with fixed and switching topologies. *Physica D: Nonlinear Phenomena*, *238*(16), pp.1550-1560.

22. Fu, X., Wang, Y., Ge, Y., Chen, P. and Teng, S., 2013, December. Research and application of DBSCAN algorithm based on Hadoop platform. In *Joint International Conference on Pervasive Computing and the Networked World* (pp. 73-87). Springer, Cham.

23. Fu, Y.X., Zhao, W.Z. and Ma, H.F., 2011. Research on parallel DBSCAN algorithm design based on mapreduce. In Advanced Materials Research (Vol. 301, pp. 1133-1138). Trans Tech Publications.

24. Feng, Z., Jaewon, S. and James, R., 2002. Information-driven dynamic sensor collaboration for target tracking. *IEEE Signal Processing Magazine*, *19*(2), pp.61-72.

25. Gasch, A.P. and Eisen, M.B., 2002. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome biology*, *3*(11), pp.research0059-1.

26. Girma, A., Garuba, M. and Goel, R., 2018. Advanced machine language approach to detect DDoS attack using DBSCAN clustering technology with entropy. In Information Technology-New Generations (pp. 125-131). Springer, Cham.
27. Goil, S., Nagesh, H. and Choudhary, A., 1999, June. MAFIA: Efficient and scalable subspace clustering for very large data sets. In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Vol. 443, p. 452). ACM.

28. G. Q. Hu. 2012, Robust consensus tracking of a class of second-order multi-agent dynamic systems. Systems & Control Letters, vol. 61, no. 1, pp. 134–142,.

29. Gupta, R. and Das, S.R., 2003, October. Tracking moving targets in a smart sensor network. In *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No. 03CH37484)* (Vol. 5, pp. 3035-3039). IEEE.

30. Habib, R. and Afzal, M.T., 2017. Research work recommendation using citation proximity in bibliographic coupling. *Turkish Journal of Electrical Engineering & Computer Sciences*, *25*(4), pp.2708-2718.

31. Harisinghaney, A., Dixit, A., Gupta, S. and Arora, A., 2014, February. Text and image based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm. In 2014 International Conference on Reliability Optimization and Information Technology (ICROIT) (pp. 153-155). IEEE.

32. Hartigan, J.A. and Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), pp.100-108.

33. H. Du, Y. Cheng, Y. He, and R. Jia, 2016, ―Second-order consensus for nonlinear leader-following multi-agent systems via dynamic output feedback control,‖ International Journal of Robust and Nonlinear Control, vol. 26, no. 2, pp. 329–344.

34. He, Y., Tan, H., Luo, W., Feng, S. and Fan, J., 2014. MR-DBSCAN: a scalable MapReduce-based DBSCAN algorithm for heavily skewed data. Frontiers of Computer Science, 8(1), pp.83-99.

35. He, Zengyou, Xiaofei Xu, and Shengchun Deng, 2002. "Squeezer: an efficient algorithm for clustering categorical data." Journal of Computer Science and Technology 17.5. pp. 611-624.

36. H. Kim, Shim, H. and Seo, J.H., 2011. Output consensus of heterogeneous uncertain linear multi-agent systems. *IEEE Transactions on Automatic Control*, *56*(1), pp.200-206.

37. Huang, Z. and Ng, M.K., 1999. A fuzzy k-modes algorithm for clustering categorical data. *IEEE Transactions on Fuzzy Systems*, *7*(4), pp.446-452.

38. Huang, Zhexue. 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data mining and knowledge discovery , pp . 283-304.

39. J. A. Fuemmeler, Atia, G.K. and Veeravalli, V.V., 2011. Sleep control for tracking in sensor networks. *IEEE Transactions on Signal Processing*, *59*(9), pp.4354-4366.
40. J. Feng and Wen, G.X., 2015. Adaptive NN consensus tracking control of a class of nonlinear multi-agent systems. *Neurocomputing*, *151*, pp.288-295.

41. J. Hu and Hong, Y., 2007. Leader-following coordination of multi-agent systems with coupling time delays. *Physica A: Statistical Mechanics and its Applications*, *374*(2), pp.853-863.

42. J. Liu , Liu, H., Ye, N. and Cui, Z., 2011. Adaptive dynamic-cluster based target tracking strategy in wireless sensor networks. *Journal of Northeastern University (Natural Science)*, *32*(8).

43. Joshi, A. and Kaur, R., 2013. A review: Comparative study of various clustering techniques in data mining. International Journal of Advanced Research in Computer Science and Software Engineering, 3(3).

44. Junaid, S., &Bhosle, K.2014. Overview of Clustering Techniques. International Journal of Advanced Research in Computer Science and Software Engineering, 4.11, pp. 621-624.

45. K. G. Vamvoudakis, Lewis, F.L. and Hudas, G.R., 2012. Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality. *Automatica*, *48*(8), pp.1598-1611.

46. Khan, K., Rehman, S.U., Aziz, K., Fong, S. and Sarasvady, S., 2014, February. DBSCAN: Past, present and future. In The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014) (pp. 232-238). IEEE.

47. K. K. Chintalapudi. andGovindan, R., 2003. Localized edge detection in sensor fields. *Ad Hoc Networks*, *1*(2-3), pp.273-291.

48. Kisilevich, S., Mansmann, F. and Keim, D., 2010, June. P-DBSCAN: a density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos. In Proceedings of the 1st international conference and exhibition on computing for geospatial research & application (p. 38). ACM.

49. Li, L. and Xi, Y., 2011, October. Research on clustering algorithm and its parallelization strategy. In 2011 International conference on computational and information sciences (pp. 325-328). IEEE.

50. Li, Z., Duan, Z., Chen, G. and Huang, L., 2010. Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint. IEEE Transactions on Circuits and Systems I: Regular Research works, 57(1), pp.213-224.

51. L. Wang, F. Xiao. Finite-time consensus problems for networks of dynamic agents. IEEE Twith double-integrator dynamics. Automatica, vol. 47, no. 8, pp. 1706–1712, 2011.

52. L. Zeng and G. D. Hu, 2013. Consensus of linear multi-agent systems with communication and input delays. ActaAutomaticaSinica, vol. 39, no. 7, pp. 1133–1140.
53. M. Abouheaf and F. L. Lewis, 2013.Multi-agent differential graphical games: Nash online adaptive learning solutions,‖ in IEEE 52nd Conf. Deci. Contr. (CDC), Conference Proceedings, pp. 5803–5809.

54. Malhotra, A. and Bajaj, K., 2016. A hybrid pattern based text mining approach for malware detection using DBScan. CSI transactions on ICT, 4(2-4), pp.141-149.

55. Ma, L., Gu, L., Li, B., QIAO, S. and Wang, J., 2014. G-DBSCAN: An improved DBSCAN clustering method based on grid. AdvSciTechnol Lett, 74(Asea), pp.23-28.

56. Mann, A.K. and Kaur, N., 2013. Review research work on clustering techniques. Global Journal of Computer Science and Technology.

57. Matsuoka, H. and Shoki, H., 2003, September. Comparison of pre-FFT and post-FFT processing adaptive arrays for OFDM systems in the presence of co-channel interference. In *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003*. (Vol. 2, pp. 1603-1607). IEEE.

58. Ma, Y., 2013, December. The Research of Stock Predictive Model based on the Combination of CART and DBSCAN. In 2013 Ninth International Conference on Computational Intelligence and Security (pp. 159-164). IEEE.

59. Meng'Ao, L., Dongxue, M., Songyuan, G. and Shufen, L., 2015, November. Research and Improvement of DBSCAN Cluster Algorithm. In 2015 7th International Conference on Information Technology in Medicine and Education (ITME)(pp. 537-540). IEEE.

60. Ng, Raymond T., and Jiawei Han, 2002. CLARANS: A method for clustering objects for spatial data mining." IEEE Transactions on Knowledge & Data Engineering 5: 1003-1016.

61. Panda, Mrutyunjaya, and ManasRanjan Patra, 2009. A Hybrid clustering approach for network intrusion detection using cobweb and FFT." Journal of Intelligent systems 18.3: 229-246.

62. Park, Hae-Sang, and Chi-Hyuck Jun, 2009. A simple and fast algorithm for K-medoids clustering. Expert systems with applications 36.2: 3336-3341.

63. P. Lin, K. Y. Qin, H. M. Zhao, M. Sun , 2012. A new approach to average consensus problems with multiple time-delays and jointly-connected topologies. Journal of the Franklin Institute, vol. 349, no. 1, pp. 293-304.

64. P. Lin, Y. Jia, 2011.Multi-agent consensus with diverse time-delays and jointly connected topologies, Automatica 47 (4) (2011) 848–856.

65. P. Lin, Y. M. Jia, and L. Li, 2008. Distributed robust consensus control in directed networks of agents with time-delay. Syst. Control Lett., vol. 57, no. 8, pp. 643–653.
66. Q. Song, J. Cao, and W. Yu, 2010.Second-order leader-following consensus of nonlinear multi-agent systems via pinning control,‖ Systems & Control Letters, vol. 59, no. 9, pp. 553–562.

67. R. Olfati-Saber, J. A. Fax, and R. M. Murray, 2007. Consensus and cooperation in networked multi-agent systems,‖ Proc. IEEE, vol. 97, no. 1, pp. 215–233.

68. R. Olfati-Saber and R. M. Murray, 2004. Consensus problems in networks of agents with switching topology and time-delays,‖ IEEE Trans. Autom. Control, vol. 49, no. 9, pp. 1520–1533.

69. Sabau, A.S., 2012. Survey of clustering based financial fraud detection research. InformaticaEconomica, 16(1), p.110.

70. Sharma, A., Gupta, R.K. and Tiwari, A., 2016. Improved Density Based Spatial Clustering of Applications of Noise Clustering Algorithm for Knowledge Discovery in Spatial Data. Mathematical Problems in Engineering, pp 216-224.

71. Sharma, N., Bajpai, A. and Litoriya, M.R., 2012. Comparison the various clustering algorithms of weka tools. facilities, 4(7), pp.78-80.

72. S. H. Li, Du, H. and Lin, X., 2011. Finite-time consensus algorithm for multi-agent systems with double-integrator dynamics. *Automatica*, *47*(8), pp.1706-1712.

73. S. Khoo, Xie, L. and Man, Z., 2009. Robust finite-time consensus tracking algorithm for multirobot systems. *IEEE/ASME transactions on mechatronics*, *14*(2), pp.219-228.

74. S. Li, Li, S., Wang, J., Luo, X. and Guan, X., 2011. A new framework of consensus protocol design for complex multi-agent systems. *Systems & Control Letters*, *60*(1), pp.19-26.

75. S. Liu, L. Xie, and H. Zhang, ―Distributed consensus for multi-agent systems with delays and noises in transmission channels,‖ Automatica, vol. 47, no. 5, pp. 920–934, 2011.

76. Struyf, Anja, Mia Hubert, and Peter Rousseeuw. "Clustering in an object-oriented environment." Journal of Statistical Software 1.4 (1997): 1-30.

77. Suthar, N., Indr, P. and Vinit, P., 2013. A Technical Survey on DBSCAN Clustering Algorithm. Int. J. Sci. Eng. Res, 4, pp.1775-1781.

78. T. Li, M. Fu, L. Xie, J.F. Zhang, Distributed consensus with limited communication data rate, IEEE Transactions on Automatic Control 56 (2) (2011) 279–292.

79. Tseng, Y.-C., Kuo, S.-P., Lee, H.-W., Huang, 2003. Location tracking in a wireless sensor network by mobile agents and its data fusion strategies. In: International Workshop on Information Processing in Sensor Networks.
80. Velmurugan, T., and T. Santhanam, 2010. Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points." Journal of computer science 6.3, pp 363-369.

81. Wagstaff, Kiri, 2001. Constrained k-means clustering with background knowledge." Icml, pp 23-29.

82. W. Ni, D. Cheng, 2010. Leader-following consensus of multi-agent systems under fixed and switching topologies, Systems and Control Letters ,pp209–217.

83. W. Ren, R. W. Beard, 2005. Consensus seeking in multiagent systems under dynamically changing interaction topologies. IEEE Transactions on Automatic Control, vol. 50, no. 5, pp. 655–661.

84. W. Ren, K. L. Moore, and Y. Chen, 2007. High-order and model reference consensus algorithms in cooperative control of multi-vehicle systems, ASME J. Dyn. Syst., Meas., Control, vol. 129, no. 5, pp. 678–688, 2007.

85. W. Ren, 2007. Multi-vehicle consensus with a time-varying reference state,‖ Syst. Control Lett., vol. 56, no. 7/8, pp. 474–483.

86. W. Ren, R. W. Beard, and E. M. Atkins. 2007, Information consensus in multivehicle cooperative control,‖ IEEE Control Syst. Mag., vol. 27, no. 2, pp. 71–82.

87. X. Shen, P. K. Varshney, 2014. Sensor selection based on generalized information gain for target tracking in large sensor networks", IEEE Trans. Signal Proces., vol. 62, no. 2, pp. 363-375.

88. X. Wang, Y. Hong, 2010. Distributed finite-time χ-consensus algorithms for multiagent systems with variable coupling topology, Journal of Systems Science and Complexity, pp. 209–218.

89. X. W. Liu, T. P. Chen, W. L. Lu, 2009. Consensus problem in directed networks of multi-agents via nonlinear protocols. Physics Letters A, vol. 373, no. 35, pp. 3122–3127.

90. Xu, Y., Winter, J., Lee, W.-C, 2004. Dual prediction-based reporting for object tracking sensor networks. In: Proceedings of MOBIQUTOUS, pp. 56-62.

91. Yang, C., Wang, F. and Huang, B., 2009, July. Internet traffic classification using dbscan. In 2009 WASE International Conference on Information Engineering (Vol. 2, pp. 163-166). IEEE.

92. Y. Cao, W. Ren, Z. Meng, 2010. Decentralized finite-time sliding mode estimators and their applications in decentralized finite-time formation tracking, Systems and Control Letters, pp. 522–529.

93. Y. Hong, J. Hu, and L. Gao, 2016 .Tracking control for multi-agent consensus with an active leader and variable topology,‖ Automatica, vol. 42, no. 7, pp. 1177–1182.
94. Y. Hue, S.H., Li, P., Guo, J.D. and Zhou, S.G., 2004. Using Greedy algorithm: DBSCAN revisited II. Journal of Zhejiang University-SCIENCE A, 5(11), pp.1405-1412.

95. Yu, X., Ding, Y., Wan, W. and Thuillier, E., 2014, July. Explore hot spots of city based on dbscan algorithm. In 2014 International Conference on Audio, Language and Image Processing (pp. 588-591). IEEE.

96. Zhang, W., Hou, J., Sha, L, 2003. Dynamic clustering for acoustic target tracking in wireless sensor networks. In: Proc. IEEE Int. Conf. Network Protocols, pp 56-65.

97. Zhao, F., Shin, J., Reich, J, 2002. Information-driven dynamic sensor collaboration for tracking applications. IEEE Signal Proces. Mag. pp 78-86.

98. Zhou, H., Wang, P. and Li, H., 2012. Research on adaptive parameters determination in DBSCAN algorithm. Journal Of Information &Computational Science, 9(7), pp.1967-1973.

99. Z. Q. Wu, Y. Wang, 2012. Dynamic consensus of high-order multi-agent systems and its application in the motion control of multiple mobile robots. International Journal of Automation and Computing, vol. 9, no. 1, pp. 54–62.

List of Publications

DYNAMIC CLUSTER TRACKING USING DB SCAN ALGORITHM **(ISSN- 2250-3021)**

Curriculum Vitae

Full Name: Jagdatta Singh

Father's Name: SantBaksh Singh

Date of Birth: 25th August 1994

Sex: Male

Nationality: Indian

Address: !/142, vishwaskhand, Gomti Nagar, Lucknow, U.P., India.

Phone Number: +918787264636

Email Address: jagdattasingh@gmail.com

Academic Background

Examination/Degree	Institution	Year of Passing	Percentage/CGPA
M.Tech. (CSE)	BBDU Lucknow	2019(Appearing)	
B.Tech. (CSE)	NIET, AKTU	2015	60.1%
Class XII	U. P. Board	2013	67%
Class X	U. P. Board	2008	57%

Fields of interest

• Database Management and Programming.

Strengths

• Positive Attitude, Social Interaction, Hardworking.

Interest and hobbies

• Politics

• Listening to Music.