LITERATURE REVIEW

Anil K Jain et.al, in their landmark paper of pattern recognition gave an overview of this field as the primary goal of pattern recognition is supervised or unsupervised classification. Among the various frameworks in which pattern recognition has been traditionally formulated, the statistical approach has been most intensively studied and used in practice. More recently, neural network techniques and methods imported from statistical learning theory have been receiving increasing attention. The design of a recognition system requires careful attention to the following issues: definition of pattern classes, sensing environment, pattern representation, feature extraction and selection, cluster analysis, classifier design and learning, selection of training and test samples, and performance evaluation. In spite of almost 50 years of research and development in this field, the general problem of recognizing complex patterns with arbitrary orientation, location, and scale remains unsolved. New and emerging applications, such as data mining, web searching, retrieval of multimedia data, face recognition, and cursive handwriting recognition, require robust and efficient pattern recognition techniques. The objective of this review paper is to summarize and compare some of the well-known methods used in various stages of a pattern recognition system and identify research topics and applications which are at the forefront of this exciting and challenging field [1].

Scott C Newton et.al, most real data structures encountered in speech and image recognition and in medical and many other decision making tasks are quite complex in nature and rather difficult to organize for designing autonomous and optimal control and recognition systems. This paper presents a modular, unsupervised neural network architecture which can be used for clustering and classification of complex data sets. The adaptive fuzzy leader clustering (AFLC) architecture is a hybrid neural-fuzzy system which learns on-line in a stable and efficient manner. The system uses a control structure similar to that found in the adaptive resonance theory (ART-1) network to identify the cluster centers initially. The initial classification of an input takes place in a two stage process: a simple competitive stage and a distance metric comparison stage. The cluster prototypes are then incrementally updated by relocating the centroid positions from fuzzy C-means (FCM) system equations for the centroids and the membership values. The operational characteristics of AFLC and the critical parameters involved in its operation are discussed. The performance of the AFLC algorithm is presented through application of the algorithm to the Anderson Iris data and laser-luminescent fingerprint image data. The AFLC algorithm successfully classifies features extracted from real data, discrete or continuous, indicating the potential strength of this new clustering algorithm in analyzing complex data sets [2].

Timo Ahvenlampi et.al, studied about the clustering algorithm in process monitoring and control application and took a digester as a case study [3].

Ibrahim Masood et.al, studied about control chart pattern recognition which has become an active area of research since late 1980s. Much progress has been made, in which there are trends to heighten the performance of artificial neural network (ANN)-based control chart pattern recognition schemes through feature-based and wavelet-denoise input representation techniques, and through modular and integrated recognizer designs. There is also a trend to enhance it's capability for monitoring and diagnosing multivariate process shifts. However, there is a lack of literature providing a critical review on the issues associated to such advances. The purpose of this paper is to highlight research direction, as well as to present a summary of some updated issues in the development of ANN-based control chart pattern recognition schemes as being addressed by the frontiers in this area. The issues highlighted in this paper are highly related to input data and process patterns, input representation, recognizer design and training, and multivariate process monitoring and diagnosis. Such issues could be useful for new researchers [4].

S Kalyani and K S Swarup studied about the security assessment of power system using pattern recognition and fuzzy C Means algorithm [5].

J H Yang et.al, presented a pattern recognition system using statistical correlation coefficient. Pattern recognition system has been widely used for detecting unnatural patterns in a control chart. This statistical correlation coefficient method is a simple procedure and works without any training data set [6].

Zhe Song et.al proposed a data-mining approach is used to develop a model for optimizing the efficiency of an electric-utility boiler subject to operating constraints. Selection of process variables to optimize combustion efficiency is discussed. The selected variables are critical for control of combustion efficiency of a coal fired boiler in the presence of operating constraints. Two schemes of generating control settings and updating control variables are evaluated. One scheme is based on the controllable and non-controllable variables. The second one incorporates response variables into the clustering process. The process control scheme based on the response variables produces the smallest variance of the target variable due to reduced coupling among the process variables. An industrial case study and its implementation illustrate the control approach developed in this paper [8].

Xian Xia Zhang et.al, proposed a new sensor-placement strategy is developed. Its main feature is to position the sensor by utilizing the main characteristics of spatial distribution. The key technique is to use a spatial-constrained fuzzy *c*-means algorithm to extract the characteristics of spatial distribution. For an easy implementation, a systematic sensor-placement design scheme in four steps (i.e., data collection, dimension reduction, data clustering, and sensor locating) is developed. Finally, control of a catalytic packed-bed reactor is taken as an application to demonstrate the effectiveness of the proposed sensor-placement scheme [9].

Jun Hai Zhai et.al, presents an overview of various algorithms which are classified into different category. The objective of this paper is to summarize and compare some of the well-known methods used in many applications [10].

OVERVIEW OF PATTERN RECOGNITION

Pattern recognition is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patterns. It is done now a day by the intelligent machines. Machine intelligence is the new upcoming area of research. In spite of almost 50 years of research, design of a general purpose machine pattern recognizer remains an elusive goal.

The term pattern recognition [11] encompasses a wide range of information processing problems of great practical significance, from speech recognition and the classification of handwritten characters, to fault detection in machinery and medical diagnosis. Pattern recognition is a field within the area of machine learning. Alternatively, it can be defined as the act of taking in raw data and performing an action based on the category of the data. As such it is a collection of methods for supervised learning. Pattern recognition aims to classify data (patterns) based on either a priori knowledge or on statistical information extracted from the patterns. The patterns to be classified are usually groups of measurements or observations, defining points in an appropriate multidimensional space [10]. Pattern recognition system consists of two-stage process. The first stage is feature extraction and the second stage is classification. Feature extraction is the measurement of population of entities that will be classified. This assists the classification stage by looking for features that allows fairly easy to distinguish between the different classes. Several different features have to be used for classification. The set of features that are used makes up a feature vector, which represents each member of the population. Then, pattern recognition system classifies each member of the population on the basis of information contained in the information vector. Pattern recognition is the scientific discipline whose goal is the classification of objects into a number of classes or categories [17]. Depending on the application, these objects can be images or signal waveforms or any other type of measurements that need to be classified. We will refer to these objects using the generic term patterns. Pattern recognition has a long history, but before 1960s it was mostly the output of theoretical research in the area of statistics. As with everything else, the advent of computers increased the demand for practical applications of pattern recognition, which in turn set new demands for further theoretical developments. As our society evolves from the industrial to its postindustrial phase, automation in industrial production and the need for information handling and retrieval are becoming increasingly important. This trend pushed pattern recognition to the high edge of today‘s engineering applications and research [23].

Pattern recognition is an integral part in most machine intelligent systems built for decision making. A complete pattern recognition system consists of a sensor that gathers the observations to be classified or described, a feature extraction mechanism that computes numeric or symbolic information from the observations, and a classification or description scheme that does the actual job of classifying or describing observations, relying on the extracted features [8]. The classification or description scheme is usually based on the availability of a set of patterns that have already been classified or described. This set of patterns is termed the training set, and the resulting learning strategy is characterized as supervised learning. Learning can also be unsupervised, in the sense that the system is not given an a priori labeling of patterns, instead it itself establishes the classes based on the statistical regularities of the patterns [7, 19, 21] .

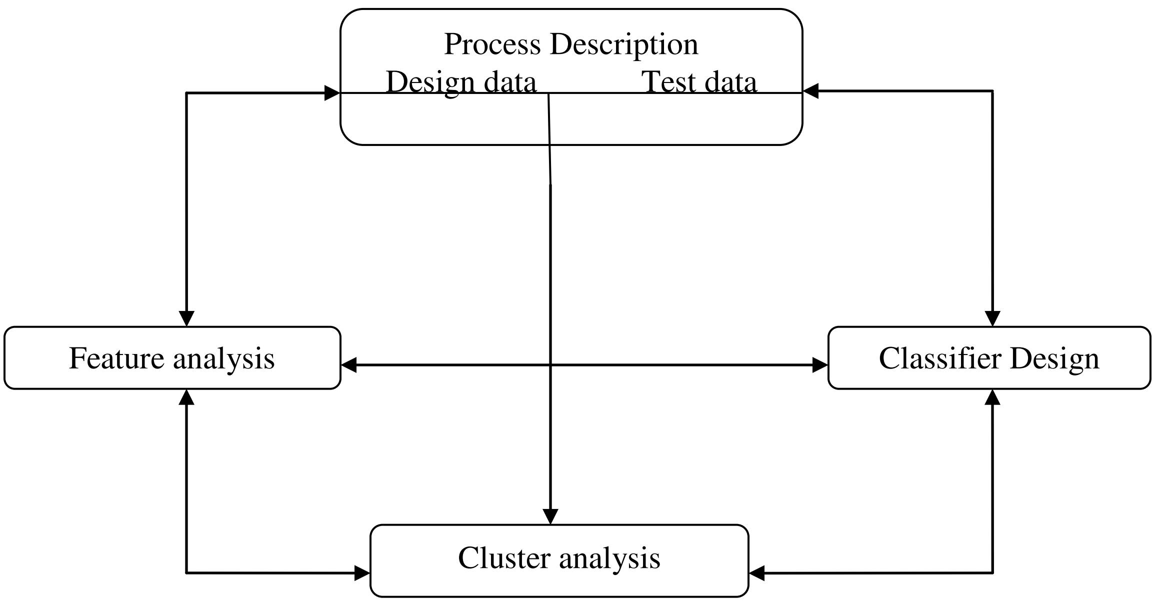

Figure 1: Different elements for pattern recognition

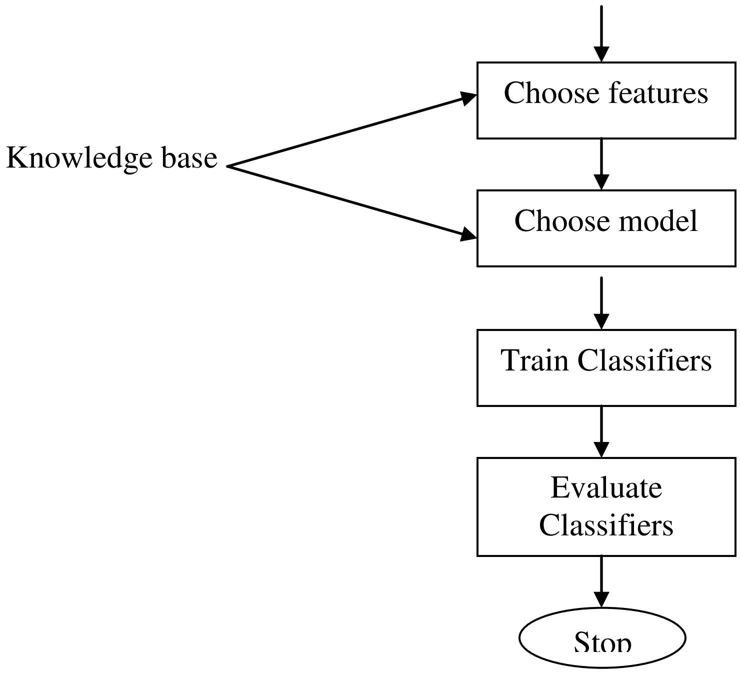

Figure 2. Flow chart for pattern recognition system

Applications

Pattern recognition is concerned with the automatic detection or classification of objects or events. Some examples of the problems to which pattern recognition techniques have been applied [literature survey]:

1. Automated analysis of medical images obtained form microscopes and CAT scanners.
2. Magnetic resonance images, nuclear medicine images, X-rays, and photographs
3. Automated inspection of parts on an assembly line
4. Human speech recognition by computers
5. Automatic grading of plywood, steel and other sheet material
6. Classification of seismic signals for oil and mineral exploration, and earthquake prediction
7. Selection of tax returns to audit, stocks to buy, and people to insure
8. Identification of people from fingerprints, hand shape and resize, retinal scans, voice characteristics, typing patterns and handwriting

9. Automatic inspection of printed circuits and printed character and handwriting recognition
10. Automatic analysis of satellite pictures to determine the type and condition of agricultural crops, weather conditions, snow and water reserves, and mineral prospects.
11. Selection of good prospects from a mail-order list
12. Classification of electrocardiograms into diagnostic categories of heart disease, detection of spikes in electroencephalograms, and other medical waveforms analyses.

Pattern recognition is studied in many fields, including psychology, ethnology, forensics, marketing, artificial intelligence, remote sensing, agriculture, computer science, data mining, document classification, multimedia, biometrics, surveillance, medical imaging, bioinformatics and internet search [13]. Pattern recognition helps to resolve various problems such as: optical character recognition (OCR), zip-code recognition, bank check recognition, industrial parts inspection, speech recognition, document recognition, face recognition, gait recognition or gesture recognition, fingerprint recognition, image indexing or retrieval, image segmentation (by pixels classification). One can find the evolution of pattern recognition; this enables the reader to establish a categorization of the existing pattern recognition system according to the used methodology and the application [16]. Pattern recognition is also applied in more complex fields like data mining (DM) also called knowledge-discovery in databases (KDD). This emerging topic includes the process of automatically searching large volumes of data for patterns such as association rules. A pattern that is interesting (according to a user-imposed interest measure) and certain n enough (again according to the user's criteria) is called knowledge. The output of a program that monitors the set of facts in a database and produces patterns in this sense is discovered knowledge"[3, 11, 19, 25].

Pattern recognition in robotics

The applications of pattern recognition in robotics are permanent. With visual pattern recognition systems, a robot may acquire the ability to explore its environment without user intervention; it may be able to build a reliable map of the environment and localize itself in the map: this will help the robot achieve full autonomy. Examples of robots using visual pattern recognition approaches are the Sony's AIBO ERS-7, Yaskawa's SmartPal, and Phillips' iCat. In robotics, visual serving or visual tracking is of high interest. For example visual tracking allows, robots to extract themselves the content of the observed scene as a human observer can do it by changing his different perspectives and scales of observation.

Pattern recognition in biometrics

The biometric authentication takes increasing place in various applications ranging from personal applications like access control to governmental applications like biometric passport and fight against terrorism. In this applications domain, one measures and analyses human physical (or physiological or biometric) and behavioral characteristics for authentication (or recognition) purposes. Examples of biometric characteristics include fingerprints, eye retinas and irises, facial patterns and hand geometry measurement, DNA (Deoxyribonucleic acid). Examples of biometric behavioral characteristics include signature, gait and typing patterns. This helps to identify individual people in forensics applications [7, 16, 23].

OVERVIEW OF STATISTICAL CLASSIFICATION

Statistical Classification

Statistical classification is a procedure in which individual items are placed in to groups based on quantitative information on one or more characteristics (features, traits) inherent in the item and based on the training set of previously labeled items. Statistical classification is far more efficient way of data classification on the basis of its nature.

Based on the literature survey and other studies, it is found that few important and widely used types of data classifiers are as following.

1. Linear classifier
   (a) Fisher's Linear Discriminant Method (FLDM)
   (b) Logistic Regression Tecnique (LRT)
   (c) Naïve Bayes Classifier (NBC)
   (d) Perceptron Models (PM)
   (e) Support Vector Machine (SVM)
2. Quadratic Classifier (QC)
3. K- nearest neighbor
4. Boosting
5. Decision Tree
6. Neural Networks (NN)
7. Bayesian Network (BN)
8. Hidden Markov Model (HMM)

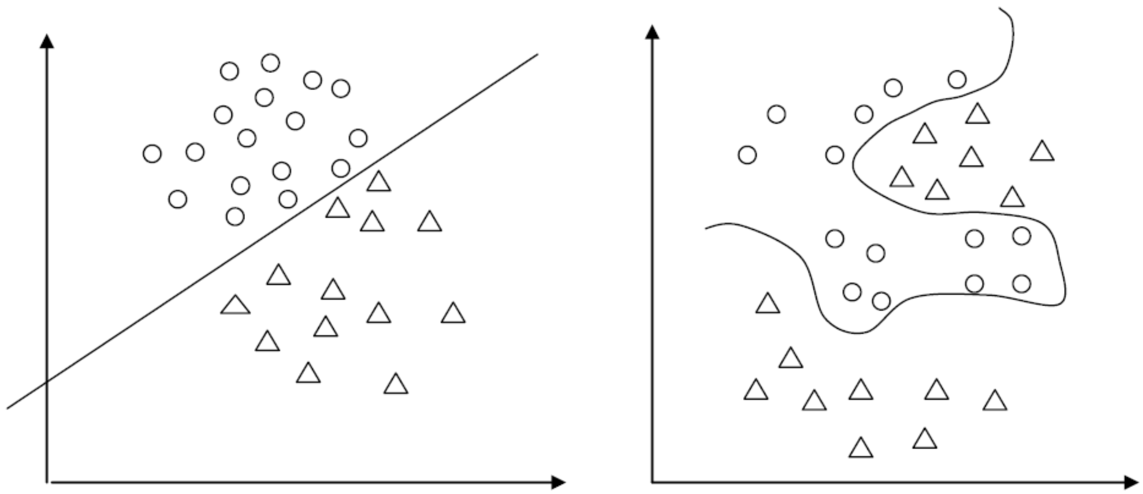

Figure 3: Linear and non-linear classification (Sourcre: PHI)

Data classification efficiency shelters on the samples taken for training of data clustering algorithms. No matter which classification or decision rule is used, it must be trained using the available training samples. As a result, the performance of a classifier depends on both the number of available training samples as well as the specific values of the samples. Samples are to be selected and created very carefully. At the same time, the goal of designing a recognition system is to classify future test samples which are likely to be different from the training samples. Data classification techniques need to be optimized. Optimizing a classifier to maximize its performance on the training set may not always result in the desired performance on a test set. Data classifier should be generalized in nature. The generalization ability of a classifier refers to its performance in classifying test patterns which were not used during the training stage. A poor generalization ability of a classifier can be attributed to any one of the following factors [11, 20]:

1. The number of features is too large relative to the number of training samples
2. The number of unknown parameters associated with the classifier is large
3. A classifier is too intensively optimized on the training set

Overtraining has been investigated theoretically for classifiers that minimize the apparent error rate (the error on the training set). Ideally, this error should be zero.

Feature selection and classification are essentially same. Once a feature selection or classification procedure finds a proper representation, a classifier can be designed using a number of possible approaches. In practice, the choice of a classifier is a difficult problem and it is often based on which classifier(s) happen to be available, or best known, to the user [22]. We identify three different approaches to designing a classifier. The simplest and the most intuitive approach to classifier design is based on the concept of similarity: patterns that are similar should be assigned to the same class. So, once a good metric has been established to define similarity, patterns can be classified by template matching or the minimum distance classifier using a few prototypes per class. The choice of the metric and the prototypes is crucial to the success of this approach. In the nearest mean classifier, selecting prototypes is very simple and robust; each pattern class is represented by a single prototype which is the mean vector of all the training patterns in that class. More advanced techniques for computing prototypes are vector quantization and learning vector quantization, and the data reduction methods associated with the one-nearest neighbor decision rule (1-NN), such as editing and condensing. The vector quantization needs to be done very carefully. The most straightforward 1-NN rule can be conveniently used as a benchmark for all the other classifiers since it appears to always provide a reasonable classification performance in most applications. The classifier techniques have dealt with in this work. Further, as the 1-NN classifier does not require any user-specified parameters (except perhaps the distance metric used to find the nearest neighbor, but Euclidean distance is commonly used), its classification results are implementation independent. Euclidean distance is a measure of error in the data classification. In many classification problems, the classifier is expected to have some desired invariant properties. An example is the shift invariance of characters in character recognition; a change in a character's location should not affect its classification [13, 16]. If the preprocessing or the representation scheme does not normalize the input pattern for this invariance, then the same character may be represented at multiple positions in the feature space. Normalization is done with respect to the maximum value in a given sate of data values. These positions define a one-dimensional subspace. As more invariants are considered, the dimensionality of this subspace correspondingly increases. Dimensionality of the subspace should be optimal. Template matching or the nearest mean classifier can be viewed as finding the nearest subspace. The second main concept used for designing pattern classifiers is based on the probabilistic approach. It involves conjectures. The optimal Bayes decision rule (with the 0/1 loss function) assigns a pattern to the class with the maximum posterior probability. It is binary in the nature. This rule can be modified to take into account costs associated with different types of misclassifications. Misclassification is the basis for rules modification. For known class conditional densities, the Bayes decision rule gives the optimum classifier, in the sense that, for given prior probabilities, loss function and class-conditional densities, no other decision rule will have a lower risk (i.e., expected value of the loss function, for example, probability of error). Error probability should be as minimum as possible. If the prior class probabilities are equal and a 0/1 loss function is adopted, the Bayes decision rule and the maximum likelihood decision rule exactly coincide. The decision rules are often non-derivable. In practice, the empirical Bayes decision rule, or plug-in rule, is used: the estimates of the densities are used in place of the true densities. These density estimates are either parametric or nonparametric. Commonly used parametric models are multivariate Gaussian distributions for continuous features, binomial distributions for binary features, and multinormal distributions for integer-valued (and categorical) features [7, 19]. A critical issue for Gaussian distributions is the assumption made about the covariance matrices. If the covariance matrices for different classes are assumed to be identical, then the Bayes plug-in rule, called Bayesnormal- linear, provides a linear decision boundary. Linearization leads to simplicity in the design. On the other hand, if the covariance matrices are assumed to be different, the resulting Bayes plug-in rule, which we call Bayes-normal-quadratic, provides a quadratic decision boundary. The order of the decision boundary is kept as small as possible. In addition to the commonly used maximum likelihood estimator of the covariance matrix, various regularization techniques are available to obtain a robust estimate in small sample size situations and the leave-one-out estimator is available for minimizing the bias. Biases are adhoc consideration in the classification design.

OVERVIEW OF DATA CLUSTERING

Clustering

Cluster analysis or clustering [7, 12] is the assignment of objects into groups (called clusters) so that objects from the same cluster are more similar to each other than objects from different clusters. Often similarity is assessed according to a distance measure. 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. Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible. Depending on the nature of the data and the purpose for which clustering is being used, different measures of similarity may be used to place items into classes, where the similarity measure controls how the clusters are formed. Some examples of measures that can be used as in clustering include distance, connectivity, and intensity.

In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters [18].

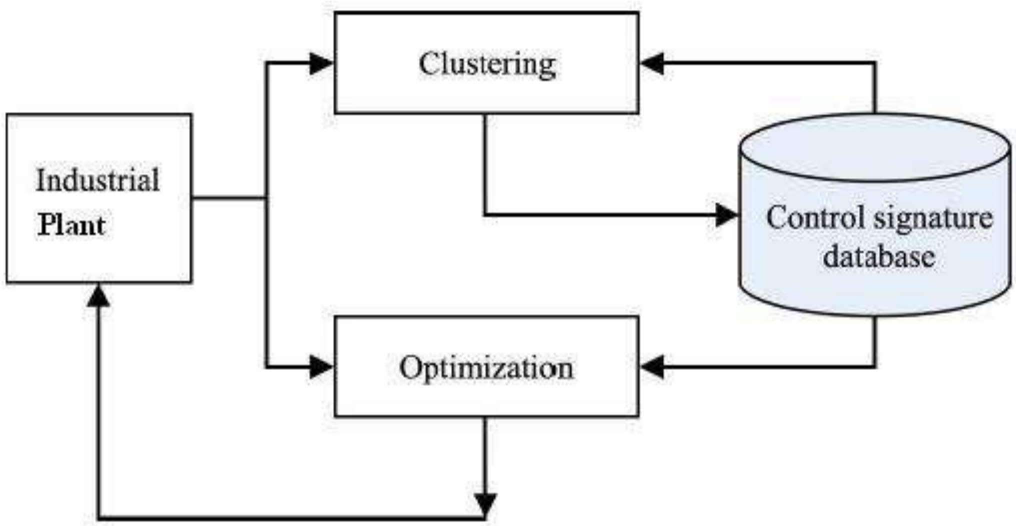

Figure 4: Cluster based efficiency monitoring of industrial plant (source: IEEE Transactions)

Applications

Recently, clustering has been applied to a wide range of topics and areas. Uses of clustering techniques can be found in pattern recognition and classic disciplines as psychology and business. This makes clustering a technique that merges and combines techniques from different disciplines such as mathematics, physics, math-programming, statistics, computer sciences, artificial intelligence and databases among others [21, 22].

Types of Clustering

There are many different ways to express and formulate the clustering problem; as a consequence, the obtained results and its interpretations depend strongly on the way the clustering problem was originally formulated. For example, the clusters or groups that are identified may be exclusive, so that every instance belongs in only one group. Or, they may be overlapping, meaning that one instance may fall into several clusters. Or they may be probabilistic, whereby an instance belongs to each group depending on a certain assigned probability [7]. Or they may be hierarchical, such that there is a crude division of the instances into groups at a high level that is further refined into levels. Furthermore, different formulations lead to different algorithms to solve. If we also consider all the variations of each different algorithm proposed to solve each different formulation, we end up with a very large family of clustering algorithms [9, 11, 14, 24].

Cluster analysis is a very important and useful technique. The speed, reliability, and consistency with which a clustering algorithm can organize large amounts of data constitute overwhelming reasons to use it in applications such as data mining, information retrieval, image segmentation, signal compression and coding, and machine learning [3, 13]. As a consequence, hundreds of clustering algorithms have been proposed in the literature and new clustering algorithms continue to appear. Our work and study reveals that most of these algorithms are based on the following two popular clustering techniques:

1. Iterative square-error partitional clustering technique
2. Agglomerative hierarchical clustering technique

Hierarchical techniques are most pragmatic and these organize data in a nested sequence of groups which can be displayed in the form of a tree. Square-error partitional algorithms attempt to obtain that partition which minimizes the within-cluster scatter or maximizes the between-cluster scatter. To guarantee that an optimum solution has been obtained, one has to examine all possible partitions of the n d-dimensional patterns into K clusters (for a given K), which is not computationally feasible. The complete solution domain must be explored. So, various heuristics are used to reduce the search, but then there is no guarantee of optimality.

Partitional clustering techniques are used more frequently than hierarchical techniques in pattern recognition applications. Pattern recognition is essentially point to pint correlation.

Literature survey reveals that in cluster analysis, a user of a clustering algorithm should keep the following issues in mind:

1. Every clustering algorithm will find clusters in a given dataset whether they exist or not; the data should, therefore, be subjected to tests for clustering tendency before applying a clustering algorithm, followed by a validation of the clusters generated by the algorithm. Thus testing should be done inevitably.
2. There is no best clustering algorithm. Therefore, a user is advised to try several clustering algorithms on a given dataset. Often selection of data clustering is random.

The data classification needs validation also. Further, issues of data collection, data representation, normalization, and cluster validity are as important as the choice of clustering strategy. The problem of partitional clustering can be formally stated as follows:

1. Given n patterns in a d-dimensional metric space. The metric space should be as large as possible.
2. Determine a partition of the patterns into K clusters, such that the patterns in a cluster are more similar to each other than to patterns in different clusters. The selection of the value of K plays a pivotal role in the data classification efficiency.
3. The value of K may or may not be specified as it is dependent on the application requirements.

A clustering criterion, either global or local, must be adopted. Global is too meaningful. A global criterion, such as square-error, represents each cluster by a prototype and assigns the patterns to clusters according to the most similar prototypes. Local is too small in its scope. A local criterion forms clusters by utilizing local structure in the data. For example, clusters can be formed by identifying high-density regions in the pattern space or by assigning a pattern and its k nearest neighbors to the same cluster. Most of the partitional clustering techniques implicitly assume continuous-valued feature vectors so that the patterns can be viewed as being embedded in a metric space. These vectors are direction oriented. If the features are on a nominal or ordinal scale, Euclidean distances and cluster centers are not very meaningful, so hierarchical clustering methods are normally applied. It finds efficient solution even in those cases which can not be distinguished. The technique of conceptual clustering or learning from examples can be used with patterns represented by nonnumeric or symbolic descriptors. The objective here is to group patterns into conceptually simple classes. Concepts are defined in terms of attributes and patterns are arranged into a hierarchy of classes described by concepts.

The two most popular approaches to partitional clustering: square-error clustering and mixture decomposition have been dealt with. A square-error clustering method can be viewed as a particular case of mixture decomposition. We should also point out the difference between a clustering criterion and a clustering algorithm. Clustering criterion is the termination step in a clustering algorithm. A clustering algorithm is a particular implementation of a clustering criterion. In this sense, there are a large number of square-error clustering algorithms, each minimizing the square-error criterion and differing from the others in the choice of the algorithmic parameters. The square of the error is a more close measure of the error in the integrated manner.

PARTICLE SWARM INTELLIGENCE

It is now considered as a part of AI. Particle swarm optimization (PSO) algorithm is a kind of evolutionary computation technique developed by Kennedy and Eberhart in 1995. It is superior to GAs in many respects. It is similar to other population-based evolutionary algorithms in that the algorithm is initialized with a population of random solutions. It is inspired from swarm behavior of the living creature. Unlike most of the other population- based evolutionary algorithms, however, each candidate solution (called particle) is associated with a velocity and 'flies' through search space. PSO algorithm rapidly attracted researcher's attention and has been applied in neural network optimization, data clustering, engineering design, etc.

Hybridization leads to overall goodness of the algorithm. A hybrid cluster algorithm using partial swarm optimization can be proposed to improve the fuzzy C means clustering algorithm. Fuzziness refers to the ambiguity. As described in evolutionary techniques, PSO also uses a population of potential solutions to search the search space. However, PSO differs from other evolutionary algorithms such that there are no DNA inspired operators in order to manipulate the population. DNA characterizes a data. Instead, in PSO, the population dynamics resembles the movement of a birds flock while searching for food, where social sharing of information takes place and individuals can gain from the discoveries and previous experience of all other companions. Thus, each companion (called particle) in the population (called swarm) is assumed to ―fly over the search space in order to find promising regions of the landscape. The landscape refers to the solution space. In the case of minimizing a function, such regions possess lower function values than other visited previously. In this context, each particle is treated as a point in a D dimensional space, which adjusts its own ―flying according to its flying experience as well as the flying experience of other particles (companions). In my experiments, a version of this algorithm is used adding an inertia weight to the original PSO dynamics. PSO leads to faster convergence and practical solutions.

PROBLEM FORMULATION

Mostly the industrial plants are non linear and multiple input-multiple output MIMO in nature. In industrial plant there are thousands of sensors and transducers which sense different physical parameter in the field and provide the value of parameters (data) to the controller and the controller subsequently forwards all the data for the control room for record keeping and statistical process control application. In the control room all the data coming from the sensor are fused together which is called as multi sensor data fusion. After the multi sensor data fusion, different pattern recognition algorithm is used to recognize different form of data and pattern classification is used for classification of different sub groups of data. Data clustering algorithms are used for clustering the same group of data which can further optimized by different stochastic global optimization techniques and evolutionary algorithm can be used. Some of the example of these kind of algorithms are particle swarm optimization, genetic algorithm, stimulated annealing, ant colony optimization, genetic programming.

Different data classification algorithms and different clustering algorithm can be used for classification and clustering purpose. Some of the data clustering algorithms which can be used are fuzzy C means algorithm, fuzzy K means algorithm. Fuzzy C means algorithm is very popular as it is easy to use, straight forward and very efficient. However fuzzy C means algorithm is very sensitive to initialization and gets trapped very easily in local optima. The said clustering algorithms can be optimized using PSO algorithm which is a stochastic global optimization tool. A hybrid fuzzy clustering method based on fuzzy C means algorithm and fuzzy based particle swarm optimization method can be used for better clustering algorithm.

This research looks in to the different aspects of pattern recognition, pattern classification, data clustering algorithms and hybrid data clustering algorithms optimized using PSO. In any industrial plant the first step of information gathering is multi sensor data fusion. After this step in information processing steps, data classification and pattern recognition principles are used to classify and recognize the data. After classification and pattern recognition the similar data are clustered together with the help of different clustering algorithms. The clustering algorithms can be further optimized using population based global stochastic optimization tool like PSO algorithm.

The above steps if applied in to a data collection and information processing give a highly efficient method of collection and information processing. In industrial plant proper classification of process variable data and clustering of similar data is very important for the purpose of record keeping and for statistical process control applications.

This research looks in to information collection of process variable data which is vital for plant operation in real time, fusion of different data, recognition of the different types of process variable data, classification of subgroups of process variable data and clustering the same for the better arrangement of process variable data. The research also proposes a hybrid optimized clustering algorithm (combination of fuzzy C means and particle swarm optimization).