CHAPTER 1 INTRODUCTION

In past, only requirement was human intelligence but in the present era world is full of artificial intelligence and its challenges that are still to be met. One can say that ranging from vehicles to household accessories all are required to be artificially intelligent. Everyone need smart phones, smart cars and smart machines. Everyone needs all things intelligent. So, our research is also aimed to develop Artificial Intelligence that provides the capability to master these challenges that is ranging from normal methods for automated reasoning to interaction techniques. As computer had made its important place in our life so it has became very necessary to find different ways of not only human-computer interaction but human-computer smart interaction. The word Smart means artificially intelligent. Artificial intelligence is becoming a need of day to day life. In the domain of essentials, all the things are required to be artificial intelligent. To meet the requirement one important step is to find an intelligent way of computer & user interaction. Data collection & classification both are part of artificial intelligence.

Machine learning is considered as a subset of artificial intelligence. In the Machine learning process a computer learns from the given dataset. After learning from the dataset it predicts the results of new data. The new data is similar to the training data. Machine learning process is all about performing special tasks by generalizing from the examples. In machine learning the focuses is given to development of computer programs which are capable of teaching itself to grow and changing when exposed to the new data. These systems are capable to automatically learn from the given data. (V. Vapnik, 1995)

Machine learning algorithm identifies patterns based on different features and then makes predictions on new unclassified data based on the patterns which it learned earlier. It is basically an algorithm that approximates a system response with a learned function. Developing some general purpose algorithm of practical value is the main goal of machine learning research. It can be considered as learning by examples which mean that it can find out how to perform important tasks by generalizing from examples. As more and more data becomes available more complex problems can be easily solved.
Machine learning is about learning to do better in the future based on past experiences. Supervised and unsupervised learning methods are two approaches of Machine learning methods (V. Vapnik, 1995). Even the large databases of various organizations are automated which provide ease of access. But the next step of ease is to automate the statistical analysis of large complex datasets which can provide a data-driven basis for automated decision making and probabilistic reasoning. For all this machine learning and data classification is required. Machine learning applications can range from natural language and image processing to scientific imaging as well as computational neuroscience.

Classification is one of the most important tasks for different applications such as text categorization, tone recognition, image classification, micro-array gene expression, proteins structure predictions and data classification etc. Most of the existing supervised classification methods are based on traditional statistics, which can provide ideal results when sample size is tending to infinity. However, only finite samples can be acquired in practice.

1.1 Learning Machines

A common scenario in modern experimental physics is the so-called black box situation (Figure 1), where one has to face the complex data e.g. a detector but it is not evident what the data represent. The problem arises from the fact that there is no a priori knowledge of its probability distribution, so it can’t be distinguished e.g. signal and background.

“Figure 1. The Black box situation with no knowledge of the system F (X)”

This lack of knowledge is often solved by using Monte Carlo-based simulations of the experiment, and by doing this one can achieve samples of a priori classified data with high credibility. It
is desirable to observe for instance model the detector output of the supposedly rare phenomenon (e.g. neutrinos), and then assume that the vast majority of our real data is background. It could model signals from two or more different phenomenon to be able to separate between them. If then one could extrapolate our Monte Carlo knowledge to be valid for all data as the approximated probability distribution of the experiment and could actually directly classify the data, i.e. it is in a white box situation.

A learning machine is an algorithm that approximates a system response with a learned function. Neural networks (NN) and Support vector machines (SVM) are the most common designs. They are closely related to each other, and differ only in the way they learn. When a learning machine is being trained, it is given pairs \((x_i, y_i)\) of data and classification, where \(x \in \mathbb{R}^n\) and \(y \in \{-1, 1\}\), e.g. \(y = 1\) for signal and \(y = -1\) for background. It then tries to come up with (learns) a function \(d(x, w)\) that mimics the system for all pairs, so that \(d(x_i, w) \rightarrow y_i\), see Figure 2. The training data pairs can come from simulations and detector samples.

![Flow chart for Training of a Learning Machine](image)

“Figure 2. Flow chart for Training of a Learning Machine.”

The error of the learned function can be expressed in three parts using the squared error function
\[Ed, y (d, y) = \text{E}_y [(y - E(y))^2] + \text{Ed} [(d - E(d))^2] + (E(y) + E(d))^2\]

= Noise (y) + Variance(d) + Bias2(d). (1)

There are two main learning methods for learning machines.

- To keep the variance fixed and to minimize the bias.
- To keep the bias fixed (very low) and to minimize the variance.

The first method is implemented by NNs, and the second one is by SVMs. Machine learning is a process in which a computer learns from dataset which is given to it and then it is able to predict the results of new data which is similar to the training data. Machine learning is all about figuring out how to perform important tasks by generalizing from examples. It is a subarea of artificial intelligence. It focuses on development of computer programs that can teach itself for growing and changing when exposed to new data. These systems automatically learn from the data.

Machine learning algorithm identifies patterns based on different features and then makes predictions on new unclassified data based on the patterns which it learned earlier. It is basically an algorithm that approximates a system response with a learned function. Goal of machine learning research is to develop some general purpose algorithm of practical value. It can be considered as learning by examples which mean that it can find out how to perform important tasks by generalizing from examples. As more and more data becomes available more complex problems can be solved. Machine learning is about learning to do better in the future based on what was past experiences.

Machine learning is a cost-effective and efficient alternative to manual programming. Following are some machine learning problems:

- Optical character recognition
- Face detection
- Medical diagnosis
- Weather prediction
- Fraud detection
There are number of machine learning methods available such as neural networks, support vector machines (SVM), decision trees, KNN etc. They are closely related to each other and differ only in the way they learn. In machine learning, classification is a process which predicts that to which category a new observation most likely belongs based on the training data which contains classified data. Input data contains various instances which are the features related to data. For example, for classifying whether a person has dengue or not blotched-red skin, muscular-pain-articulation, temperature can be taken as features. All these features would help to classify whether a person has dengue or not. More the number of examples, better the result will be. A typical machine learning problem can be understood from the following figure. This figure describes how a problem is solved by a machine learning algorithm.

**Figure 3. Learning Problem**

Some labeled training examples are provided to a machine learning algorithm. Algorithm is then trained on this data. Some new examples are given to prediction rules. According to which, predicted classification is performed of the data.
Machine learning methods are defined under two approaches which are supervised and unsupervised learning methods. In supervised learning, a machine is given a training data in the first step and in the second step machine predicts results based on this training. Goal of supervised learning is to build a model of the distribution of class labels in terms of predicting features. Classifier is then used to assign class labels to the testing data. If instances are given with known labels then learning is supervised learning. In unsupervised learning no training data is provided which means instances are unlabeled. Machine tries to classify data based on the similarity found in the input data.

1.2 Classification

Classification is the problem of identifying to which category a new observation most likely belongs based on the training data which contains classified data. It generally refers to the categorization of data. In classification objects are categorize into fixed categories. In machine learning, classification is commonly referred to as supervised learning in which objects are labeled with appropriate class.

There are a number of approaches defined to perform classification. Classification procedure is divided into two steps training dataset and testing dataset. Success of classification depends on the quality of the data provided for training because it is the only input from which a learner learns. If data provided is inadequate or irrelevant then it will lead to misclassification of the result.

Various methods are available for classification to train the data for constructing effective dataset for processing. Processing is performed to divide the available data in various classes. Every classification problem differs in terms of algorithm implemented, tools used and size and type of data to be used. Classification is performed in a number of steps which includes data preprocessing, clustering, feature construction, feature selection, classification, Regression and visualization.

Classification is performed in two steps:

- **Training:**
  In this step, training of data is performed to create a model.
• Testing:

After training the system on given input, testing is performed. Testing classifies the data into different classes. Figure 4 describes the process of classification.

Some examples are given to system; it divides the given examples into various classes. In this section some available classification methods are described.

1.2.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) was first proposed by Vapnik and has since attracted a high degree of interest in the machine learning research community (Chang & Chih-Jen, 2001). Several recent studies have reported that the SVM (support vector machines) generally are capable of delivering higher performance in terms of classification accuracy than the other data classification algorithms. SVMs have been employed in a wide range of real world problems such as text categorization, hand-written digit recognition, tone recognition, image classification and object detection, micro-array gene expression data analysis, data classification. It has been shown that SVMs is consistently superior to other supervised learning methods. However, for some datasets, the performance of SVM is very sensitive to how the cost parameter and kernel parameters are set.
1.2.2 Rough Sets

Rough set theory, introduced by Zdzislaw Pawlak in the early 1980s is a new mathematical tool to deal with vagueness and uncertainty. This approach seems to be of fundamental importance to artificial intelligence (AI) and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, and knowledge discovery from databases, expert systems, decision support systems, inductive reasoning, and pattern recognition.

The rough set concept overlaps—to some extent—with many other mathematical tools developed to deal with vagueness and uncertainty, in particular with the Dempster-Shafer theory of evidence. The main difference is that the Dempster-Shafer theory uses belief functions as a main tool, while rough set theory makes use of sets—lower and upper approximations. Another relationship exists between fuzzy set theory and rough set theory. Rough set theory does not compete with fuzzy set theory, with which it is frequently contrasted, but rather complements it. In any case, rough set theory and fuzzy set theory are independent approaches to imperfect knowledge.

Furthermore, some relationship exists between rough set theory and discriminate analysis, Boolean reasoning methods, and decision analysis. One of the main advantages of rough set theory is that it does not need any preliminary or additional information about data, such as probability distribution in statistics, basic probability assignment in the Dempster-Shafer theory, or grade of membership or the value of possibility in fuzzy set theory.

1.2.3 Decision Trees

Decision trees are the tools in which rules are implemented in the form of tree structure. They classify instances by sorting them based on their feature values. Tree contains nodes and branches. Nodes of the tree represent the rules. Every branch of the tree has certain outcome, weight and probability. A decision tree can be considered as a series of nested if-else. A decision tree is constructed with a given set of instances and a target variable. Attribute with the highest information gain is chosen to be the splitting attribute for the root node and this process is repeatedly performed for all the other
attributes. Machine learning system will predict the value of target variable. Decision trees split data based on a single feature. Algorithms used with decision tree are- C4.5, ID3 and CART etc.

1.2.4 Bayesian Network

Bayesian network is a directed acyclic graph. Bayesian networks are also called as belief networks. In this graph links represents the dependencies between nodes and nodes of the graph represents events. The graphical analysis is defined under relationship analysis. Bayesian classification is basically used to reduce the error rate in performing classification.

These kinds of networks are defined under the conditional probability estimation for each variable. Estimation of inaccurate data is performed. Algorithms that are used with these networks are Pearl's message passing method, Clique triangulation, stochastic methods. All these take advantage of clusters present in the network.

1.2.5 Bayes Theorem

Bayes theorem is used in case of conditional probability. It would be right to say that Bayesian statistics is a new field produced by Bayes' theorem which is gradually transforming the field of statistics. Due to this a worthwhile potential for scientific research can be given. It is helping to boost up the probabilistic thinking. Bayes rule-based systems are not appropriate for uncertain reasoning because of the following reasons-

- Large storage space is required.
- Calculation time is large.
- Knowledge gathering is difficult
- Human beings are not good at this

Most of the existing supervised classification methods are based on traditional statistics, which can provide ideal results when sample size is tending to infinity. However, only finite samples can be acquired in practice. SVM, a powerful machine method developed from statistical learning and has made significant achievement in some field. Introduced in the early 90's, they led to an explosion of interest in machine learning. The foundations of SVM have been developed by Vapnik and are gaining popularity in field of machine learning due to many attractive features and promising empirical performance. SVM method does not suffer the limitations of data dimensionality and limited samples.
K-MAP is used to reduce the number of attributes by finding the relationships between attributes and required results. RSES is also used to find reducts which then applied to SVM to obtain better classification results. The aim of my work is efficient rule set generation using K-MAP, SVM and RST for classification of high dimensional dataset.

1. Representation of the information system – Information system is a set of objects represented in a data table. In which each row contains an object and each column represents a measurable attribute for the object.
2. Decision Systems – If an information system has an additional attribute called decision attribute, then it becomes decision system.
3. Indiscernibility – Objects in the decision system and information system may be indistinguishable from one another.

“Figure 5. Planning of Method”
4. Discretization – It is a simple preprocessing technique which is not directly related to rough set theory.

5. Lower and upper approximation – Lower approximation contains the set of objects that can be classified as members of a given set with certainty. Upper approximation contains the set of objects that can possibly be classified as members of a given set.

6. Reducts – Some attributes that don’t contribute to the classification result can be omitted such that the result remains intact. The set of remaining attributes is called a reduct.

7. Discernibility matrix – Work can be started by creating a symmetric matrix called discernibility matrix.

8. Discernibility function – From the discernibility matrix obtained a discernibility function can be immediately found.

9. Rule generation – After this, it is required to construct a set of rules from training data, so that new or unseen objects can be divided into known classes. These rules obtained can be minimized.

1.3 Description of Flow Diagram of Proposed System
Step 1: Training Data Set
A training data set has been taken as input in RSES for attribute reduction.

Step 2: Data Discretization of Conditional Attributes
Data discretization is a widely used data transformation procedure, which involves finding cuts in the data set, which divide in the data into intervals. If the data contains noncategorical attributes, they must be discretized in order to produce effective rules. Values lying within an interval are then mapped to the same value. Doing this reduces the size of the attributes value set and ensures that the rules that are mined are not too specific.

Step 3: Organizing Decision Table
Decision table is organized after data discretization.

Step 4: Reduction of Decision Table
Rough set based RSES is used to simplify the discretized decision table in order to obtain simpler classifications. One way to reduce the dimension of the data is to remove attributes
whose removal will not destroy the indiscernibility Relation. Attributes, which can be removed without affecting the system, are considered redundant. A minimal set of attributes after the redundancy removal is called reducts. Reduct preserves the degree of dependency and it cannot be reduced any further while still preserving the degree of dependency.

Thus, reduct is a minimal set of feature attributes which has the same ability to discern groups as when the full set of feature attributes is used.

**Step 5: New Training Data Set**

After discretization and attribute reduction, a new training data set is obtained.

**Step 6: Testing Data Set**

In this step, testing data set is used as input to RSES and step 1-5 is repeated.

**Step 7: Support Vector Machine Classification System**

A new training and testing data set applied to SVM for classification.

**Step 8: Output**

After applying SVM on new training and testing data set the classification accuracy is obtained.

**Advantages:**

- It is based on the original data only and does not need any external information, unlike probability in statistics or grade of membership in the Fuzzy set theory.
- It can reduce condition attribute and eliminate redundant information, but not reduce any effective information.
- It is a tool suitable for analyzing not only quantitative attributes but also qualitative ones.
- It reduces storage requirement.
- Reduce data set takes less execution time.