DATA SETS

For simulation purpose, we used six real data sets obtained from UCI Machine learning repository [24]. These datasets are outlined below:

i. IRIS Data: This is the well-known Anderson’s Iris data set. It contains a set of 150 vectors in four dimensions each representing one of three different species or classes of Iris flowers. The four features are sepal length, Sepal width, Petal length, and Petal width. The data set contains 50 instances of each of the three classes.

ii. Thyroid: Thyroid data set has 215 data patterns with five features distributed in three classes namely. Euthyroidism, hypothyroidism and hyperthyroidism.

iii. Wisconsin Breast Cancer (WBC): It has 699 samples distributed in two classes. There are 16 missing values, all in the bare nuclei feature. Hence, we have removed them. All reported results are computed on the remaining 683 data points. WBC data set is collected from Dr. Wolberg's clinical cases at the University of Wisconsin (Wolberg & Mangasarian 1990, Zhang 1992a). These were collected over a period of two and a half years, and the problem is to determine whether the tumors were benign or malignant based on data for each cancer patient. There are ten features, where one is the serial number, and nine are identifying characteristics: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitosis. The features are all linearly ordered but discrete in the range (1-10) and hence defined as continuous.

iv. Glass: Glass data set has 214 points with nine features distributed in six classes.
v. Wine: Wine data set consist of 178 samples in 13-dimension distributed in three classes. These data represent the results of chemical analysis of wines grown in a particular region of Italy but derived from three different cultivators. The analysis determined the quantities of 13 constituents found in each of the three types of wine.

vi. Ionosphere: The data represent autocorrelation functions of radar measurements. The task is to classify them into two classes denoting passages or obstruction in the ionosphere. There are 351 data patterns with 34 attributes distributed in two classes.

vii. Pima Indian diabetes (pima): It has 768 instances from the National Institute of Diabetes and Digestive and Kidney Diseases. The task is to determine whether the patient shows signs of diabetes according to World Health Organization criteria. All patients are females who live near Phoenix, Arizona. They are at least 21 years old and of Pima Indian heritage. There are eight continuous features: number of times pregnant, plasma glucose concentration, diastolic blood pressure, triceps skin fold thickness, 2-hour serum insulin, body mass index, diabetes pedigree function, and age.

These data sets are summarized in Table-6.3

...
List of Figures

Figure-3.1 Basic PNN Model
Figure-4.1 Biological Model of Neuron
Figure-4.2 Schematic representation of a mathematical model of neuron
Figure-4.3 Single layer Neural Network with n inputs and p outputs
Figure-4.4 Three-layer MLP Neural Network
Figure-4.5 Feed-forward Artificial Neural Nets
Figure-4.6 Basic architecture of PD
Figure-4.7 The building blocks of the PNN model.
Figure 5.1 General pseudo-code of Evolutionary Algorithm
Figure-5.2 Flow diagram of Optimization process of GA
Figure-5.3 Pictorial representation of selection, crossover and mutation operators
Figure-5.4 Roulette wheel selections of chromosomes.
Figure-5.5 Rank selections of chromosomes
Figure-5.6 Tournament selections of chromosomes.
Figure-5.7 Flow diagram of proposed scheme
Figure-5.8 Neighborhood topologies of PSO
Figure-5.9 Graphical representation of the g_best (left) and l_best (right) neighborhood topologies.
Figure-5.10 A pictorial representation of the Von Neumann neighborhood (top left), three-dimensional representation of the Von Neumann neighborhood (bottom left), the star (top right) and the pyramid (bottom right) neighborhoods.
Figure-5.11A Ants in a pheromone trail between nest and food.
Figure-5.11B An Obstacle interrupts the trail.
Figure-5.11C Ants find two paths to go around the obstacle.
Figure-5.11D A new pheromone trail is formed along the shorter path.
Figure-6.1  Pictorial presentation of comparative study on IRIS data set.
Figure-6.2  Pictorial presentation of comparative study on Wine data set
Figure-6.3  Pictorial presentation of comparative study on Glass data set
Figure-6.4  Pictorial presentation of comparative study on WBC data set
Figure-6.5  Pictorial presentation of comparative study on Ionosphere data set

Figure-6.6  Pictorial presentation of comparative study on Thyroid data set
Figure-7.1  Comparison of Classification Accuracies between proposed scheme and other schemes
Figure-7.2  Comparison of execution time between proposed and other schemes

•••
List of Tables

Table 5.1  Representations of the base 10 numbers 0 and 255 in different bases
Table-6.1  Results Comparison in different classifiers
Table-6.2  Details of the datasets and GA applied for feature selection
Table-6.3  Summary of Data Sets Used
Table-7.1  Description of the Data Set Used
Table-7.2  Parameters and population size used in GA
Table-7.3  A Comparison of Times of execution and respective Classification Accuracies obtained through proposed scheme with PNN, RCPNN with Gradient Descent and RCPNN with PSO.

...
List of Tables

Table 5.1 Representations of the base 10 numbers 0 and 255 in different bases

Table-6.1 Results Comparison in different classifiers

Table-6.2 Details of the datasets and GA applied for feature selection

Table-6.3 Summary of Data Sets Used

Table-7.1 Description of the Data Set Used

Table-7.2 Parameters and population size used in GA

Table-7.3 A Comparison of Times of execution and respective Classification Accuracies obtained through proposed scheme with PNN, RCPNN with Gradient Descent and RCPNN with PSO.
Derived Publications

This appendix provides a list of publications that has been published or is currently being reviewed, that were derived from the work introduced in this thesis.

Publications in Journal


2. Accepted a research paper “Investigating a novel GA-Based Feature Selection Method Using Improved KNN Classifiers” in the International Journal of Information and Communication Technology (IJICT) ISSN (Online): 1741-8070 - ISSN (Print): 1466-6642”

3. Submitted a research paper “An Evolutionary Feature Selection Technique Using Polynomial Neural Network” in the Elsevier which is under review status.

International Conferences


National Conferences

