ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 1, March 2018

6th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '18)

13th-14th March 2018

Organized by

Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

Comparative Evaluation of Various Techniques for Fisher's Iris Classification

Mohan Prasath.M[1], Paul Nathan.K[2], Prakash.M[3], Praveen Kumar.D[4], Chidambaram.S[5]

UG Scholar, Department of ECE, Adhiyamaan College of Engineering, Hosur, TN, India[1,2,3,4]

Associate Professor, Department of ECE, Adhiyamaan College of Engineering, Hosur, TN, India[5]

ABSTRACT: Fisher's Iris data base is perhaps the best known database to be found in the pattern recognition literature. The data set contains 3 classes of 50 instances each, where each class refers to a type of Iris plant. One class is linearly separable from the other two; the latter are not linearly separable from each other. Cluster analysis plays a vital role in various fields in order to group similar data from the available database. There are various clustering algorithms available in order to cluster the data but the entire algorithm are not suitable for all applications. Classification and clustering are both used to group data (observations) into multiple groups where those in the same group are in some way "similar" and those in different groups are not. In this project, we envisages distinct classification techniques and assessed the effectiveness of the algorithms on Fisher's Iris dataset.

KEYWORDS: Fisher's Iris data, Pattern recognition, Cluster analysis, Classification

I. INTRODUCTION

The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by the British statistician and biologist Ronald Fisher in his research. The use of multiple measurements in taxonomic problems as an example of linear discriminant analysis. Two of the three species were collected in the Gaspe Peninsula "all from the same pasture, and picked on the same day and measured at the same time by the same person with the same apparatus". The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor).Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.The data base contains the following attributes:

1)sepal length in cm
2)sepal width in cm
3)petal length in cm
4)petal width in cm

The fifth attribute can be predicted which is the class attribute this means that each instance also includes an identifying class name, which are as follows: Iris Setosa, Iris Versicolour, and Iris Virginica. This forms the matrix of 150*3 and is used in the program. By examining the Sepal and Petal, The program can classify the IRIS plant. Sepal width has positive relationship with Sepal length and petal width has positive relationship with petal length.

II. RELATED WORK

There are so many experts research on iris flower dataset. R.A. Fisher first introduced this dataset in his famous paper 'The use of multiple measurements in taxonomic problems' [5].
In a seminar VitalyBorovinskiy demonstrated the solution of iris flower classification problem by 3 types of neural networks; multilayer perceptron, radial basis function network, and probabilistic neural network. The result indicates that multilayer perceptron performed better both during the training and testing [6].

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 1, March 2018

6th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '18)

13th-14th March 2018

Organized by

Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

The adaptation of network weights using Particle Swarm Optimization (PSO) as a mechanism to improve the performance of Artificial Neural Network (ANN) in the classification of IRIS dataset was proposed by Dutta D., Roy A., Reddy k [7].

Later, Patrick S. Hoey analysed the Iris dataset via two distinct statistical methods; He first plotted the dataset onto scatterplots to determine patterns in the data in relation to the Iris classifications. Furthermore, develop an application in Java that will run a series of methods on the dataset to extract relevant statistical information from the dataset. With these two methods, he made a concrete prediction about the dataset [8].

III. FISHER'S IRIS DATA SET

Standard dataset containing sepal and petal width and length for three types of Iris flowers, Setosa, Versicolor and Virginica. One species (Setosa) is relatively easy to classify, while the other two are much harder to separate. The data contains 150 observations, 50 from each species.

Fig 1 Iris Versicolor

Fig 2 Iris Virginica

Fig 3 Iris Setosa

Table 1: Description of three classes of Fisher's Iris data Set

S.No	Class	Number of Samples
1	Iris setosa	50
2	Iris virginica	50
3	Iris versicolor	50

IV. METHODS AND TECHNIQUES

CLASSIFICATION TECHNIQUE

Classification of remotely sensed data is used to assign corresponding levels with respect to groups with homogeneous characteristics, with the aim of discriminating multiple objects from each other within the image. A classification problem arises when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. Many problems in business, science, industry, and medicine can be treated as classification problems.

CLUSTER ANALYSIS TECHNIQUE

Cluster analysis groups the given data objects based on only information found in the data and describes the objects and their relationships. The data objects which have the maximum similarity within a group and the

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 1, March 2018

6th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '18)

13th-14th March 2018

Organized by

Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

greater the difference between the groups are, the better or more distinct the clustering. Clustering is an effective technique for exploratory data analysis, and has found applications in a wide variety of areas.

CONFUSION MATRIX

A confusion matrix is a table that is often used to **describe the performance of a classification model** (or "classifier") on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing.

- There are two possible predicted classes: "yes" and "no". If we were predicting the presence of a disease, for example, "yes" would mean they have the disease, and "no" would mean they don't have the disease.
- The classifier made a total of 165 predictions (e.g., 165 patients were being tested for the presence of that disease).
- Out of those 165 cases, the classifier predicted "yes" 110 times, and "no" 55 times.
- In reality, 105 patients in the sample have the disease, and 60 patients do not.

✓ **True Positives (TP):** These are cases in which we predicted yes (they have the disease), and they do have the disease.
✓ **True Negatives (TN):** We predicted no, and they don't have the disease.
✓ **False Positives (FP):** We predicted yes, but they don't actually have the disease. (Also known as a "Type I error.")
✓ **False Negatives (FN):** We predicted no, but they actually do have the disease. (Also known as a "Type II error.")

	Predicted: NO	Predicted: YES	
Actual: NO	TN=50	FP=10	60
Actual: YES	FN=5	TP=100	105
	55	110	

Fig. 4 Confusion Matrix Example

V. SVM CLASSIFIER

The SVM classifier is a binary classifier that maximizes the margin. The separator hyperplane is parallel to the margin planes and is midway between the planes. Each margin plane passes through point(s) of the learning set that belong to a particular class and is closest to the margin plane of the other class. The distance between the margin planes is called the margin. Note that multiple pairs of margin planes are possible with different margins. The SVM algorithm finds the maximum margin separating hyperplane. The points from each class that determine the margin planes are called the support vectors (SVs). We can set up an optimization problem by directly maximizing the margin (geometric margin).

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 1, March 2018

6th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '18)

13th-14th March 2018

Organized by

Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

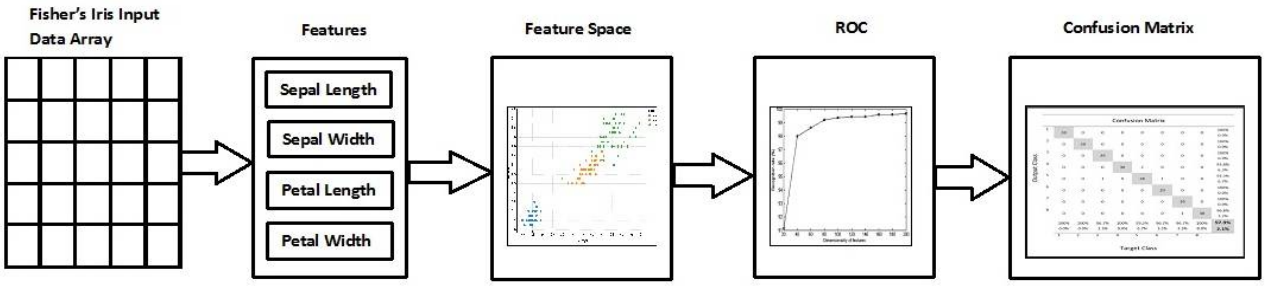

Fig. 5 Flow diagram of the proposed system

VI. RESULTS AND DISCUSSION

Experimental work is done through MATLAB programming language. The iris flower have various species ,in that three species namely Iris setosa, Iris virginica and Iris versicolor are taken for clustering based on the available data sets provided by Fisher's Iris data set. The fifty data sets classify the length and width of sepals, petals of three species commonly. One of the clusters contains Iris setosa and the other cluster contains both Iris virginica and Iris versicolor and is not separable without the species information.so this experimental results proves the efficiency by clustering all the three species with time complexity. The overall classification accuracy of three Iris classes with various versions of support vector machine classifiers are reported in Table 2. In the following table the values of different classifiers obtained by supplying different sets of input data into the chosen models of SVM and their outputs are presented. The simulation was carried out using Intel ® core ™ i5-7200U CPU @ 2.50 GHz PC with 8GB RAM.

Table 2:Fisher's Iris data Set(SVM Classifiers)

	Iris Setosa	Iris Versicolor	Iris Virginica	OA
L -SVM	100%	90%	96%	95.3%
Q- SVM	100%	92%	98%	96.7%
C- SVM	100%	94%	92%	95.3%
FG- SVM	96%	88%	98%	94.0%
MG- SVM	100%	96%	94%	96.7%
CG- SVM	100%	98%	90%	96.0%

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 1, March 2018

6th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '18)

13th-14th March 2018

Organized by

Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

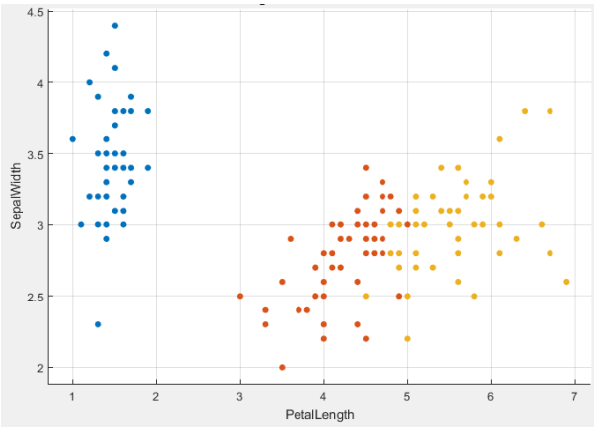

Fig 6: Feature Space with Sepal Width and Sepal Length

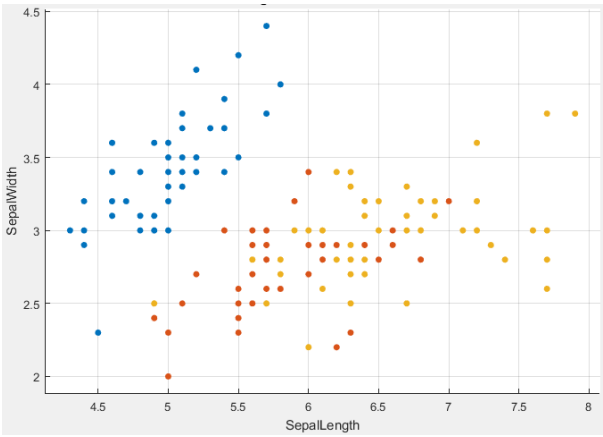

Fig 5: Feature Space with Sepal Width and Petal Length

VII. CONCLUSION

This project presents the comparative assessment of distinct support vector machine classifiers for Iris data classification. Experimental results shows that Q-SVM and MG-SVM has a very good performance compared to other variants. It is reported that the class Iris Setosa and Iris Virginica has been classified with high overall accuracy than the class of Iris Versicolor because of the good feature attributes in the training data. Furthermore, we also analyzed the overall performance of our algorithm by using three different species of iris datasets as initial input for clustering. The experimental results proved the effectiveness of the classification algorithm.

REFERENCES

[1] V.Leela, K.Sakthipriya,R.Manikandan, "A Comparative Analysis Between K-mean And Y-means Algorithms in Fisher's Iris DataSets." International Journal of Engineering and Technology (IJET).

[2] R.A.Abdulkadir1, Khalipha A. Imam, M.B. Jibril, "Simulation of Back Propagation Neural Network for Iris Flower", Classification American Journal of Engineering Research, e-ISSN: 2320-0847 p-ISSN: 2320-0936 Volume-6, Issue-1, pp-200-20.

[3] The dataset of Iris flower is available in the given link source http://archive.ics.uci.edu/ml/datasets/Iris

[4] The detail study about a Iris flower is available in the given link https://en.wikipedia.org/wiki/Iris_flower_data_set

[5] F. R. A., "The use of multiple measurements in taxonomic problems," Annual Eugenics, vol. 7, no. II, pp. 179-188, 1936.

[6] V. Borovinskiy, "Classification of Iris data set," University of Ljubljana, Ljubljana, 2009.

[7] D. Dutta, A. Roy and K. Choudhury, "Training Artificial Neural Network using Particle Swarm Optimization Algorithm," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 3, pp. 430-434, 2013.

[8] P. S. Hoey, "Statistical Analysis of the Iris Flower Dataset," University of Massachusetts atLowell, Massachusetts, 2004