IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Density-Based Support Vector Machine with Outlier Filters for Real-Time Analysis of Rocks and Structures

Yukti Varshney[1] Shivendra Pratap Singh[2], Nitin Saini[3], Shreyance Sharma[4]
Moradabad Institute of Technology, Moradabad[1, 2]
Teerthanker Mahaveer University, Moradabad[3]
Bhagwant University, Ajmer[4]

Abstract: *Artificial Intelligence solve many real time issues and problems around the world. Real Time Analysis of Rocks and Structures is one such challenge. In many situations, it is necessary to know about rocks, building or iceberg structure. For example, in mining, rescue operation, sailing of ship, etc. The structure of rocks/ debris/ icebergs are obtained through advanced imaging systems and sensors. This paper introduces the application of AI algorithm for the analysis and prediction of the earth or rock structure.*

Keywords: Real-Time Bus Tracking, Passenger Information System GPS Integration, Mobile App, Data Security

I. INTRODUCTION

Images and patterns of rocks / debris / icebergs are got through advanced imaging system and sensors can be broughtinto the computer system and analyzed. I propose Machine learning and /or deep learning algorithm to analyze these structures and patterns, and provide required output with improved accuracy compared to existing methods. The functionality of these algorithms is to analyze the structure ofrocks and provide the required output for example, which partof the rock will soft and which part of the rock will be hard based on various parameters and patterns. The system will also use previously learned/analyzed data to predict the current output. Such system will be useful in providing easy path amidst of hard rock while mining, rescue operation, etc….All operations are to be done in real time.

A. Literature Survey

Support Vector Machines (SVM) is a machine learning algorithm used for classification purpose. It is one of the bestof its class. Another name for Support Vector Machines is Support Vector Network (SVN)

Support Vector Network (SVM) is the work of Cortes and Vapnik in 1995 [1]. It has many sophisticated features in dealing classification of patterns involving data of high dimension and scales small. Support Vector Network is a two-class classification model. Support Vector Network uses the concept of Vapnik– Chervonenkis dimension theory. It reduces the structural risk by statistical learning theory.

Support Vector Network produces highest interval in the feature space in linear classification. The aim of the model isto find the highest interval.

In advancement in machine learning, combination of the algorithm [2] and Support Vector Network is also being used as this technique may improve the training efficiency in a great scale. Comparative study [3] used Principal Component Analysis and Support Vector Network for Classifying datasets. Accuracy achieved by this method is 90.24% and 66.8% on two different datasets. Accuracy of classification method used in [4] is 82.7% to predict that people fall on the floor. To compare various medical datasets algorithms used in [5], and finally predicted result with 84% accuracy.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

II. METHODOLOGY

For both classification and regression problems Support Vector Network can be applied. SVM efficiently operates on more than variable that are continuous in nature. These variables are also categorical in nature. Support Vector Network identify various data classification by creating a hyper plane. This hyper plane is created in a space of multiple dimension. Support Vector Network adopts iterative method and generates hyper plane that is optimal. Hence the error is minimized. Main purpose of Support Vector Network is to determine a maximum marginal hyper plane (MMH) that partition the dataset into classes in best manner.

SVM is one of the best machine learning technique for performing classification. It's a machine learning algorithm under supervised category. Classifying data into different classes is one of its main applications. SVM get trained on a set of labelled data. Though SVM is used for classification problems, SVM is also applied for regression problems. Thisis one of the main advantages. For classification SVM placesa hyper plane boundary that decides and separates any two classes. Other application of Support Vector Network are Object Detection and image classification.

In this paper density /color values a rock structures is used that are collected from upper half and lower half of rock dataset for Classifying them using Support Vector Network.

Support Vectors

Data points, which are nearest to the hyper plane are calledSupport Vectors. Support Vectors will acts as separating line based on margin calculation. The classifier is constructed based on these much relevant points

Hyper plane

A Set of objects having different class memberships isseparated by a decision plane called a hyper plane.

Margin

The distance by which the two lines are apart from each other on the closest classes is called a margin. By measuring the length of a line drawn at 90^0 from the line to support vectors or closest points is used to calculate the above gap. A margin is a good margin, if the margin is larger between the classes. A margin is a bad margin, if the margin between the classes is small margin.

How does Support Vector Network work?

The main aim is to classify dataset in an efficient and accurate way. Margin is the distance between the either points. The aim is to consider a hyper plane with the highest margin between support vectors. Support Vector Network searches for the MMH plane as follows:

Create hyper planes to differentiate the dataset into classesaccurately.

Select the correct hyper plane with the highest classification from the either nearest data points.

Outliers

Outliers are data points that are not consistent with other points in the observed data set. In statistics an outlier is an observation that lies at an abnormal distance from other values in a random sample from a population [6]-[7]. The reason for occurrence of an outlier may be because of errors in coding or errors occurring in sensor input or recordings, abnormal cases, errors in measurements etc… Usually outliers occur when two distributions mix with each other. Two methods are there to handle outliers.

Outlier detection or removal added as a step of preprocessing

Developing a robust model that is not sensitive to the outliers.

Density based Support Vector Machine

Density based Support Vector Machine [DSVM] detects and removes the outliers before processing [8].

The main objective of DSVM is to reduce the influence of outliers and to increase the margin and create best generalization and to adjust the decision boundary accordingto the density of data sets. The computations complexity is decreased by DSVM by reducing the number of supportvectors. The input vectors in DSVM are in highest confidence area of the data set. When compared with other input vectorsDSVM input vectors are more informative.

The outlier points that are away from dense data points aredetected by DSVM. The method is as follows.

Dense data points are found first. These points are considered to be important and meaningful points. The otherpoints are not meaningful and not or less important and can be ignored. The area which is denser is determined by Euclidean distance between data points of one data set.

DSVM with Distance Formula

The geometric distance in two dimensional space from point a(x1, y1) to b(x2, y2) is given by the formula

$$D(a,b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

By this formula the D (a, b) for all the data points a, b of one class must be calculated.

Using this formula we have to find the distance of one pointwith all other points. For example Distance of point 1 and other points (point 2, point 2…..point n).

Next find the sum of all distances calculated in previous step. For example

d1= [D(1,2) + D(1,3)+…..+ D(1,n)]

In Next step find the average of above distances using the above sum d1.

The data points that are inside or outside the denser area isclassified using this average distance.

The points which are inside the average distance are important and meaningful points and they are considered for new training data set. The points that are outside are less important and meaning less and they are ignored and removedfrom training data set. Thus the meaning less points are filtered before the data set enters the analysis [9]-[10].

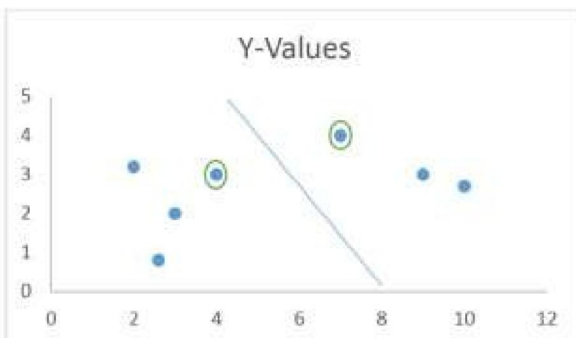


Fig 1 Data set output with Outliers, having small margin

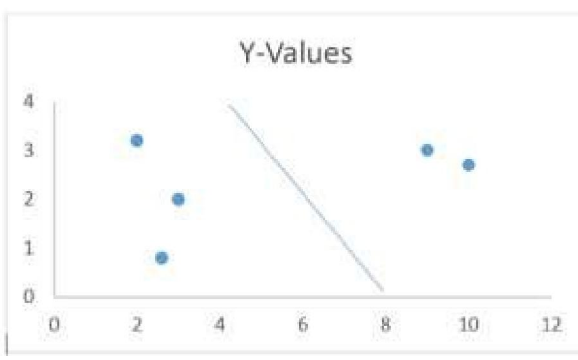


Fig 2 Data set output without Outliers, having biggermargin

III. IDENTIFYING EARTH OR ROCK TYPE AND/ORSTRUCTURE BY PHYSICAL PROPERTIES

Basic physical properties such as colour, shape, and hardness are used to identify the most common minerals in Earth's crust. The basis of mineral availability is different. Certain minerals developed due to same environmentconditions and may occur in same rock. Some mineral are created under different environment conditions and may occur in different rocks. Because of this property of similarity of minerals in rocks, some small occurrence of different mineral can be ignored even though they predict same colour or density. We can test few physical properties without testing many properties. And the results can be matched with the actual physical properties encountered in the field. We will find that the results are likely to match 90% with that of the original properties in the field.

The physical properties of a mineral are identified by its chemical composition and internal atomic structure. Hence diagnosis can be made using the properties. Examination of different physical properties of minerals can be carried out with varying level of property, including colour, formation style of crystal (or shape), hardness, shine, density, and cleavage or fracture.

The data collected for the above parameter are analysed with SVM algorithm.

There are several techniques available with kernel like, radial basis function (rbf), linear, polynomial and others. Most preferred kernel function is radial basis function. For non- linear hyper-plane we use radial basis function and polynomial. In the following example, we classified the rock data set based of density characteristics of rock by using linear kernel.

SVN Implementation

Example: A linear Support Vector Network kernel is used.Python language is used for implementation.

```
import numpy as npy
import matplotlib.pyplot as plot
from sklearn import m1, datasets

# import some data to play with earth = datasets.load_earth()
X = earth.data[:, :2]
# we only take two features. We could avoid this ugly
slicing by using a two-dim dataset y = earth.target

# we create an instance of SVM and fit out data. Our data is
not scaled as we want to plot the support vectors
C = 1.0  SVM regularization parameter
m1 = m1.SVC(kernel='linear', C=1,gamma=0).fit(X, y)

# create a mesh to plot in
x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1
h = (x_max / x_min)/100
xx, yy = npy.meshgrid(np.arange(x_min, x_max, h),
 npy.arange(y_min, y_max, h))

plot.subplot(1, 1, 1)
Z = m1.predict(np.c_[xx.ravel(), yy.ravel()])
Z = Z.reshape(xx.shape)
plot.contourf(xx, yy, Z, cmap=plt.cm.Paired, alpha=0.8)

plot.scatter(X[:, 0], X[:, 1], c=y, cmap=plot.cm.Paired)
plot.xlabel('Rock length')
plot.ylabel('Rock density')
plot.xlim(xx.min(), xx.max())
plot.title('Rock Analysis using SVM')
plot.show()
```

Fig 3 Code Segment

And the results are matched with the reference data and thetype of rock or structure is decided. Following figure shows results obtained for the selected data sets.

Sample Classification Charts

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

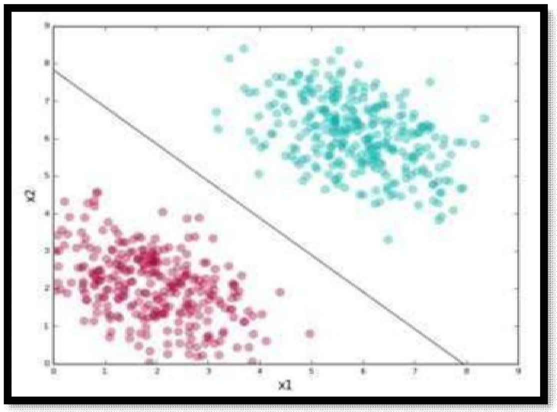
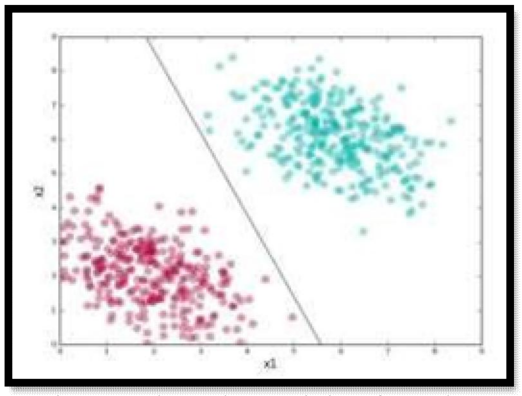


Fig.4. Hardness characteristics of Samples



-

Fig. 5. Hardness characteristics of Samples

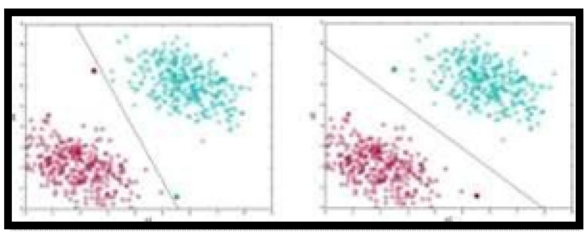


Fig.6. Hardness characteristics of Samples

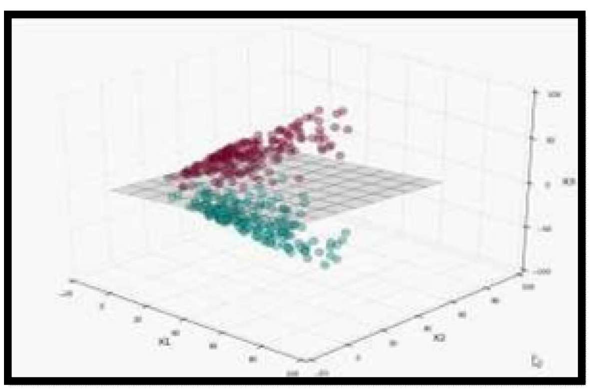


Fig.7. Three Dimensional- One Plane -Hardness

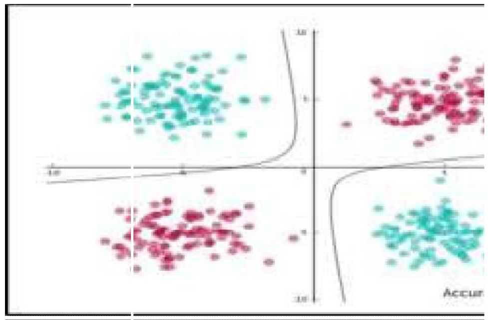


Fig. 8. Non- Separable Dataset- Three Dimensional -3Plane

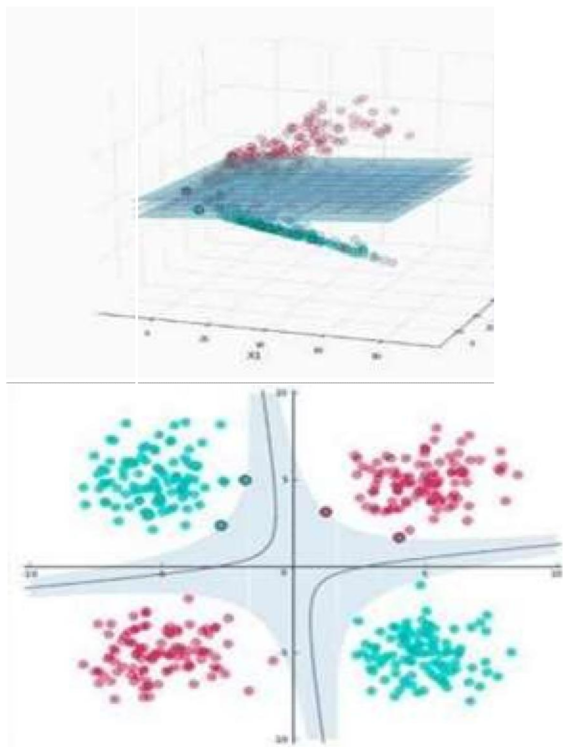


Fig. 9. Non- Separable Dataset- Three Dimensional-3 Plane

Results

From the above charts we can conclude the characteristics of rock under study. For example in the above charts the green plot shows low density of rock structure and red plot shows high denser rock characteristics. With these we can predict the hardness of the rock. And we can structure the hard portion of rock based on the predictions. In figure 7 the chart gives the prediction that the upper half of the rock structure is denser (harder) than the lower half of the rock structure. The predictions provide results with accuracy of 85% when compared with the actual characteristics of the rock.

IV. FUTURE WORKS

Here we have used only compared features of rock structures collected from to parts of the rock and generated the predictions. Similarly we can take several features on different parts of the rocks through various sensor values and can use different machine learning and/or deep learning algorithms to classify and/or predict and conclude the structures of the rock more accurately.

V. CONCLUSION

The results shows the Artificial Intelligence Techniques had provided predictions of rock characteristics with high accuracy. Thus application of above Artificial Intelligence algorithm enhances analyzing and predicting the structure of earth or rock in real time which will be helpful in mining, rescue operations, sailing of ships, etc….

REFERENCES

[1] Cortes, C.; Vapnik, V. Support-vector networks v. Mach. Learn. 1995 20, 273–297.
[2] Wold, S.; Esbensen, K.; Geladi, P. Principal Component analysis. Chemom. Intell. Lab. Syst. 1987, 2, 37–52.
[3] Bouzalmat, A.; Kharroubi, J.; Zarghili, A. Comparative Study of PCA, ICA, LDA using SVM Classifier.
[4] J. Emerg. Techol. Web Intell. 2014 6.
[5] Hai, N.T.; An, T.H. PCA-SVM Algorithm for Classification of Skeletal Data-Based Eigen Postures. Am. J. Biomed. Eng. 2016, 6, 47—158.
[6] V. Cherkassky and F. Mulier, Learning from Data, Concepts, Theory and Methods, IEEE Press 2007.
[7] S. Theodoridis and K. Koutroumbas, Pattern Recognition, Academic Press 1999.
[8] Z. Nazari, D. Kang, and H. Endo, Density Based Support Vector Machines, The 29th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), pp.1- 3, 2014.
[9] D. Ripley, Robust Statistics, M.Sc. in Applied Statistics MT2004.
[10] Y. Dodge, The Concise Encyclopedia of Statistics, Springer 2008.