International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-1, May 2019

Prediction of Clusters using SVM Algorithm in Delineation of Management Zones

Krishna Priya C B, S. Venkateswari

Abstract*: *Management zones in precision agriculture can be defined as the homogenous regions in the particular productivity area. To increase the yield productivity of the agriculture, it is important to manage these agriculture zones effectively. An important technique in analyzing these management zones are clustering. Using clustering it is easy to categorize the productivity area of the crop field. Clustering is a relevant topic in Data mining. Different clustering algorithms can be used to find out the clusters for the management zones. These clusters can be predicted using the algorithms such as Support vector Machine Algorithm.

Index Terms*: *Machine Learning, Support Vector Machine. Agriculture, Data mining.

I. INTRODUCTION

Machine Learning plays an important role in the analysis of data. Machine learning and Artificial Intelligence, both contributes the concepts in predicting the data clusters. Machine Learning has the ability to find out the data analysis automatically by using its techniques such as supervised learning as well as unsupervised learning. Machine learning helps the system to do the analysis automatically using algorithms. Machine learning techniques involves supervised, unsupervised and semi supervised learning. Supervised learning works on the known data sets and to produce future patterns. There will be provided a labelled set of information. But in the case of unsupervised learning no labelled data set is given for the analysis of data. So in unsupervised learning the techniques are used to form inferences from the unknown biases. Based on classification and regression data can be interpreted in the analysis. Classification depends on the dependent variables for processing to create the results. These dependent variables are categorical and unordered. Meanwhile regression is used to predict the output using the train data. Regression also use dependent variables which are either in continuous or ordered values. In this paper management zones are predicted with specified algorithms in the machine learning such as Support Vector machines. Support vector machine algorithms uses classification and regression techniques.

Revised Manuscript Received on 22 May 2019.
***** Correspondence Author
Mrs. Krishna Priya C B*, Research Scholar, Noorul Islam Centre for Higher Education, Kumaracoil-629180, Kanyakumari District, Tamil Nadu, India. E-mail: krishnahemand@gmail.com
Dr. S. Venkateswari, Department of Computer Science, L N Govt. College, Ponneri, Thiruvallur, Tamil Nadu, India.

Precision farming techniques are widely used in our country. This will improve the productivity in the agriculture sector. This type of farming is supported with management zones. The delineation of management zones can be performed with clustering algorithms such as K means, GK clustering and farthest first algorithm. Compare among these three algorithms farthest fist algorithm is more effective than other two. The created clusters can be predicted with machine learning algorithms such as support vector machines. Data used in this analysis is spatial data related to the management zones. Spatial data can be recognized as the information related to the spatial environment such as soil, land scape, humidity, pressure etc. The spatial data which is used in the processing will be collected from the controlled data environment using the Information technology tools.
Here support vector machine algorithm is used to classify the clusters. The clusters are basically management zones in the specified crop field. So these clusters formed will be treated well to produce high yield. The formed clusters can be used to produce high yield by applying soil fertility according to the needs. The crop productivity will be increased as a result of this proper management of the fields. Support vector algorithms in machine learning is used to predict the clusters in the management zones. The entire process for the analysis is described in the paper. The following are the organized form of this paper. The organized structure of the process is given as follows. In Related study given a brief description of the works. Description of the input is explained in the analysis of data. Methods and procedures explains the processing of the methods. Analysis of result is described in the Result and discussion. In the Conclusion explains the future works to be done in the specified area.

II. REVIEW OF LITERATURE

A. Review Stage

Machine learning techniques are best described by Huang, G., Huang et.al, [1]. Different techniques including classification and regression is explained in this study. Studies regarding the management zones in precision agriculture can be explained by the studies of T.A. Doerge and D.Rajesh [2], [3]. Various Site specific management zones are detail explained by T.A. Doerge [2]. Different maps for fertilizers are implemented with the data mining concepts. Management zone strategy has been studied with various examples. Related data mining clustering algorithms are referred by Garima Sehgal and [4] and Sharmila et.al [5].

Comparisons of different clustering algorithms helped to find out the suitable algorithm which can be used in the depiction of management zones.

B. Analysis of data

Data used in the paper is related to spatial data such as soil and its factors. Here soil is classified as management zones to improve the soil fertility. Clustering helps the analyst to easily understand the soil properties. Various Soil properties such as Co2 Flux, TSoil, PH, CO2a, Tair and RH have taken as soil parameters. Information regarding the soil has taken from the following site https://esdac.jrc.ec.europa.eu/ESDB_Archive/Soil_Data/Global.htm

These soil characteristics effects the soil productivity. So applying the fertilizers will enhance the soil productivity more. 79 records are used for analyzing the data. The sample input table is given below.

Table 1. The Soil Productivity

Reference Number	CO2_Flux_Exp	PH	CO2a	Tair	RH
1002	3.14	5.9	257	39.9	105.9
1003	3.54	3.2	117	39.9	105.9
1004	3.44	4.8	204	39.9	105.9
1005	3.22	4.2	198	39.9	105.9
1006	3.19	3.2	119	39.9	105.9
1007	3.38	4.2	190	39.9	105.9
1008	3.11	8.3	312	39.9	105.9
1009	3.14	7.8	262	39.9	105.9
1010	4.72	3.6	111	39.9	105.9
1011	3.5	6.9	366	38.9	106
1012	3.47	17.1	874	38.9	106

Soil fertility factors are given in the Table (1). Input data is organised in a tabular form. The unprocessed input may also contain impurities. So to avoid the impurities a pre-processing is needed. Pre-processing can be done with some program codes as well as WEKA tools. The data in the given bellow.

Table 2. gives the description about the records used.

Data Set Name	No of Attributes	No of Instances
Dataset 1	5	79

Here pre-processing has been done from the program segments instead of using the WEKA tools.

C. Methodology

The following part will discuss bout the procedures and analysis of the algorithm used in the paper. A supervised machine algorithm which is used in the entire process is Support Vector Machine algorithm

Support Vector Machine Algorithm is used for the classification and regression analysis in the data mining. In most of the case it is used with the classification processes. The algorithm plots each data element from the N dimensional space. Each point value will becomes the coordinates for the same. After performing the classification by getting a hyper plane, that will separates the two groups.

The following figure shows the concept.

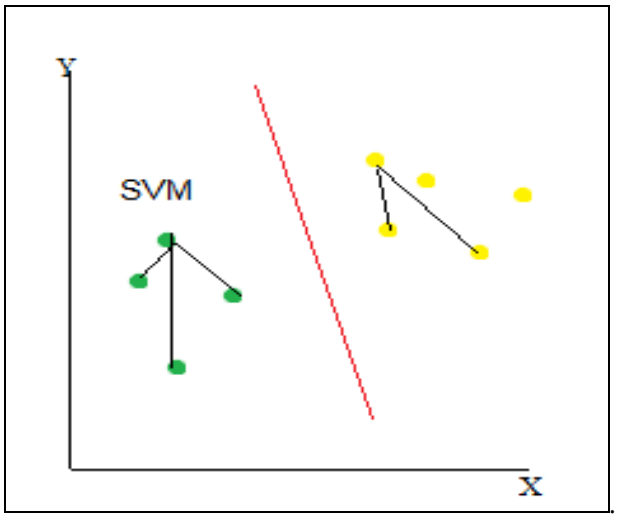

Fig 1. SVM Example

Support vectors defines the coordinates of the individual data value. These are used to differentiate the hyper-plane region of the classified data. SVM algorithm is also known as Support Vector Network. SVM is easy to handle the labelled data for training. So it is used in supervised learning. But in the unlabelled data un supervised learning is used to find out the future set of information

SVM is using support vectors for the analysis of data. Classification of the set of input is the common method in machine learning. This technique can be applied in the field of predicting the management zones in precision agriculture. SVM creates many hyper plane with the data, it is hard to choose which hyper plane separates the data in a large area. Linear classifier is used for the clustering process. We are given with a known labelled set of data, so it is easy to predict the groups with the SVM

III. DATA FLOW OF THE SYSTEM

The next section describes the process flow of the system

A. Selection of Input data

Data input set is loaded from the input file. This Input file will contain some unwanted data such as missing values, duplication etc.

B. Pre processing

Pre-processing is an important step in the data mining. Pre-processing helps to filter out the original result set. This dataset can be used in the data processing which will lead to an accurate output. Pre-processing can done with the help of using either WEKA tools or program codes.

C. Application of Supervised learning

Here we are using the soil data for processing. The data used will undergo algorithms which supports Machine learning. As we are using labelled data it is easy to generate the classified groups.

D. SVM Selection

Input values in the dataset can be transformed into an absolute value. This will make the computation easier. The following Math function is used to find the absolute value.
Absolute distance: Math. Abs ()
The pre-processed values will create the coordinates of their own since they are support vectors in nature.
The algorithm uses the following steps.
Declare the values for the data

While not reached **End Do**
Select first node from the set randomly
{Processing each node}
While no.of visited node< no.of nodes
Then continue the processing
If there are unvisited nodes then select node randomly
Else if there are visited neighbours
Select **one node randomly**
Else
Select one node randomly
End While
The above mentioned steps can be drawn in a flow chart as follows

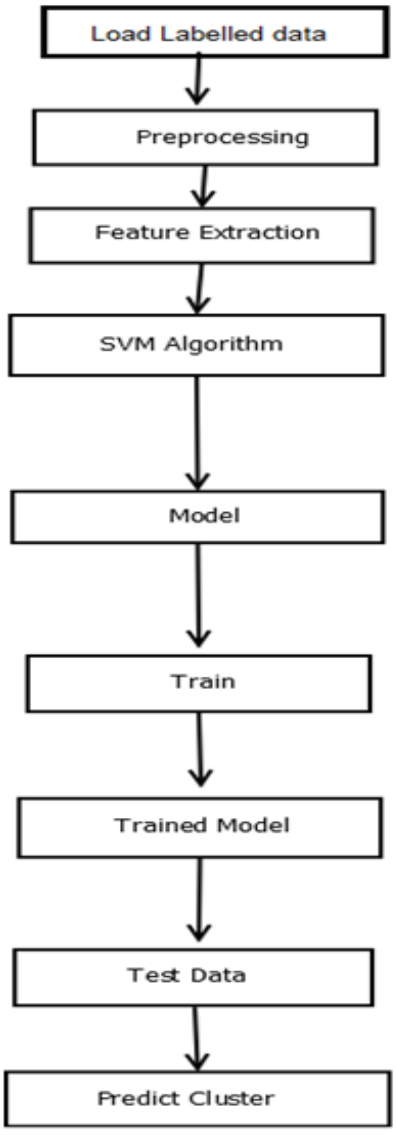

Fig 2. Process flow of the System

E. Support Vector Machine Algorithm

SVM are used in the linear classification of data analysis. It uses a hypothesis space of linear functions which is trained with a learning algorithm. No of data space can be large for SVM. Both simple and large input can easily classify with the algorithm. Two differentiate two classes of data there can form many hyper plane. But to find the largest margins of these separation is difficult. In such scenario SVM are used well. These hyper planes can form, the boundaries for the data space. The falling of data in either side of the plane can be consider as other groups.it can be linear, or multidimensional according to the data character set.
As a result of the SVM algorithm greedy permutation of points can be clustered together since the clusters formed will not be in uniform. Hence these algorithms can be used for large scale data analysis such as spatial data. SVM algorithm is similar to Naive Bayes algorithm. Using the support vector machines management zones in the precision agriculture is easily managed. Non uniformed data can easily be clustered using this algorithm since it is based on greedy method.

IV. RESULT ANALYSIS

As an initial step data is pre-processed with program segments. The system which is used in the processing should support the latest technologies. The system should have these hardware requirements:
Screen resolution of 1024x768 for proper and complete viewing of screens.
Intel Pentium IV or V
•4GB/8 GB RAM
The software requires the support of the following software for the database and other requirements.

- Operating System : Windows 7 64 bit
- Front End tools : Java 1.8
- Backend : Mysql
- IDE : Netbeans 8.2

The memory used and time taken to generate the clustering has been analysed.
The graph which shows the time and memory is given below. In the graph the red line represents the Farthest first clustering algorithm.

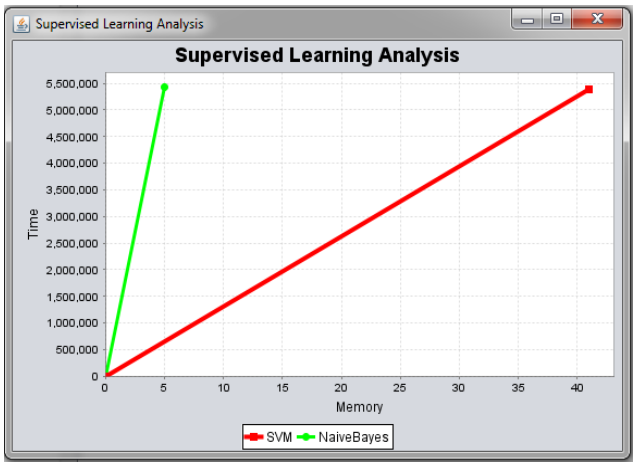

Fig 3. The SVM speed analysis

In memory analysis graph x axis represents the count of each iteration and the y axis represents the memory used in megabytes for the algorithm. Space complexity of the algorithm is showing an optimum growth, and this algorithm can be used to include more non uniform data.

In time analysis graph x axis represents the count of each iteration and the y axis represents the time taken by the algorithm in milliseconds.

The time taken to clustering is as follows.

The overall view of a dataset used in the clustering process is

Table 3. Output Result Set

Algorithm	No of Groups	Ratio of Each Groups	Unclassified Instance
SVM	5	150:219:199 :189:179	0

Five clusters are formed using the input dataset which is in the ratio given in the table. No Unclustered distance is there in the final stage of the processing

V. CONCLUSION

Here Support Vector Machine is used is for finding the management zones in the precision agriculture. Input data taken was the soil characteristics. Data is classified using the Support Vector Machines. Classification of data is done from a N dimensional space. As a future enhancement another algorithm can used to compare the efficiency of the Support Vector machine Algorithm. Management zones are easily classified with a linear space of data points.

REFERENCES

1. Trends in extreme learning machines: a review, by Huang, G., Huang, G., Song, S., & You, K. (2015). Neural Networks, (cited 323 times, HIC: 0 , CV: 0)
2. T.A. Doerge, “Management zone concepts”, Site-specific management guidelines. Atlanta, USA: Potash and Phosphate Institute, 1999. IEEE DOI 10.1109/eScience.2014.42
3. D.Rajesh, “Application of Spatial Data Mining for Agriculture”, International Journal of Computer Applications (0975 – 8887) Volume 15– No.2, February 2011
4. Garima Sehgal#1 Dr.Kanwal Garg, Comparison of Various Clustering Algorithms, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3074-3076
5. Sharmila, Mukesh Kumar, An Optimized Farthest First Clustering Algorithm, 2013 Nirma University International Conference on Engineering (NUiCONE)DOI:10.1109/NUiCONE.2013.6780070
6. A Review on Multi-Label Learning Algorithms, by Zhang, M., & Zhou, Z. (2014). IEEE TKDE, (cited 436 times, HIC: 7, CV: 91)
7. D Ramesh, B Vishnu Vardhan, “Data Mining Techniques and Application Agricultural Yield Data”. International Journal of Advanced Research in Computer and Communication Engineering, 2013.
8. E.A.Speranza “A Cluster-Based Approach to Support the Delineation”2014 IEEE 10th International Conference on science doi:10.1109/eScience.2014.42
9. E. Gustafson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In Proc. of IEEE CDC, 1979
10. A survey on feature selection methods, by Chandrasekhar, G., & Sachin, F. Int. J. on Computers & Electrical Engineering, (cited 279 times, HIC: 1, CV: 58)

AUTHORS PROFILE

Mrs. Krishna priya C B, Research Scholar, Noorul Islam Centre for Higher Education, Kumaracoil-629180, Kanyakumari District, Tamil Nadu, India.

Dr. S. Venkateswari, Assistant Professor, Department of Computer Science, L N Govt. College, Ponneri, Thiruvallur, Tamil Nadu, India.

.