International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 22 (2018) pp. 15951-15955

Delineation of Management Zones in Precision Agriculture using Different Clustering Algorithms

Mrs. Krishna Priya C B[1], Dr. S. Venkateswari[2]

[1]Research Scholar, Noorul Islam Center for Higher Education, thacalay" former Nooral Islam university, India.
E-mail: krishnahemand@gmail.com

[2]Assistant Professor, Department Computer Science
L N Govt. College, Ponneri, Thiruvallur, Tamilnadu, India.
E-mail: venki.vasan@gmail.com

Abstract

Now a days precision agriculture is making a good benchmark in the field of Crop productivity. To enhance the rate of yield productivity spatial data related to soil, topology and other geo informatics details needs to be performed. Characteristics of the soil will effects the agricultural activities. So different management zones in the farming area needs to be managed well. Data mining techniques can be used to differentiate the management zones in the precision agriculture. Here different clustering algorithms are used for depiction of management zones.

INTRODUCTION

Data mining is one of the important technique which can be applied in the field of data analysis. Data mining is the process of extracting the patterns from the given datasets. Different clustering algorithms are used for finding these patterns. Clustering in data mining is based on the process of forming similar items together.

To form clusters for a set of data different types of clustering method is used such as partitioning, Hierarchical etc.

Management zones are homogenous regions in precision agriculture, which can be efficiently managed to attain good yield productivity. Here three clustering algorithms k means, gk clustering and farthest algorithms are used for delineating the management zones in the precision agriculture. Using these three algorithms different clusters can be easily depicted with their memory and time management. These three algorithms are equally efficient to form clusters but there are some significant difference among them.

Agricultural productivity will be increased with the farming techniques such as satellite monitoring, precision farming etc. Such farming techniques definitely incorporate with advanced information technology. In such scenario precision agriculture has its own relevance. Importance of Precision farming is reflects in the all area of agriculture globally. Precision Agriculture is a farming method which using the advanced technology rather than conventional methods. A farmer can easily manage their crop under the controlled environment of precision agriculture. Precision Agriculture is also called satellite farming since it is managed with the data provided by the satellites technologies. A controlled farming environment will enhance the entire agriculture from the crop selection to the harvesting.

Management zones are the key factor in the area of satellite farming. In precision agriculture the entire cultivating area can be logically divided into management zones. These management zones are categorized based on the features of soil, crop, atmospheric characteristics etc. Management zones are the regions having the similar features. In this paper soil is taking as the important component of precision agriculture and describes the steps for managing the soil properties. Depiction of management zones in the case of soil characteristics will be enhancing the entire cultivations. Characteristics of soil can be divided in to management zones in terms of Organic matter, Texture, Colour, NPK (Nitrogen Phosphorus and potassium) etc. The farming area will not be equally distributed with this factors. So this will be difficult for the farmers to properly manage these zones

In this paper three important algorithms are used to find the clusters of the soil. Three algorithms are k means clustering, gk clustering and farthest first algorithm. These algorithms are implemented and compared against their time and memory needed to execute the input datasets. In the following manner the paper is arranged. Related work gives the literature study conducted and describes the different methods to be used.

The following are the organized form of this paper. Related Work gives a brief description about the related works in the area. Source and selection methods of the input data set is explained in the paragraph Input data set. Methodology gives the description of the techniques and algorithms used in this study. Discussion of the result and analysis is explained in Result analysis. Future enhancements regarding the study is discussed in paragraph conclusion.

DATA SET DESCRIPTION

Management zones in the agriculture is affected with the different factors physically and logically. Main components which affecting the precision agriculture are climate and soil. Important climatic features are Wind, Temperature, Altitude and rain. Meanwhile soil plays an important role in agriculture. Important soil attributes includes soil organic carbon, NPK (Nitrogen, Phosphorus and Potassium), slope, Soil PH, Tair etc.

Managing these factors plays an important role in the Precision Agriculture.

Reference Number	PH	Co2	Tair	RH
1001	1.52	51.45	46.39	49.27
1002	2.11	51.76	45.74	49.97
1003	1.21	51.6	46.37	51.07
1004	1.57	51.85	49.2	47.98
1005	1.24	48.79	45.75	48.99
1006	1.77	44.86	48.56	45.78
1007	1.34	51.85	50.05	47.79
1008	2.52	52.31	51.04	54.44
1009	1.25	48.63	45.12	42.67
1010	1.52	42.93	42.64	49.39

Fig. 1: Input Data set

The above figure explains the sample input used in these paper.

Procedure Strategy

Three important algorithms are used to find the clusters formed are K means, gk clustering and farthest first algorithm. Clustering is an important concept in the data mining field of computer science. Clustering is the process of grouping the homogenous items together. It is easy to find out the similar trends in the dataset using clustering. Here a comparison is made with the three algorithms using the input soil dataset. After implementing the three algorithms a conclusion can be reached regarding the better productive algorithm

A.K Means clustering algorithm

K means clustering comes under semi supervised learning. This algorithm uses un labelled data as its input. K means algorithm try to find the clusters which has similar features. K number of clusters will be the output for the algorithm. Centroid of the cluster will helps to find which type data is used for clustering

METHODOLOGY OF THE K MEANS ALGORITHM

K means algorithm is based on a repeated refinement of the procedure. The procedure aims to find the k centroids of the data elements. Input data set for an algorithm should have an ‘n‘ number of clusters and dataset containing sample. The sample should be any particular entity with specific attributes. The algorithm repeatedly works on two steps.

1. Initialising the Data elements

   Similar features will be grouped together and formed centroids. The centroids formed will be the data elements with the similar features.

2. Redefining the centroids

   Recompilation of the centroids is calculated with some objective function. This process will be repeated until reach the stopping criteria. The stopping criteria may be there is no data points to change their centroids or no more data to be there to analyse the algorihm.

The two steps will be run repeatedly for finding the K number of clusters for the data set. Mainly this algorithm uses Euclidian distance for calculating the distance between the datapoints.

A. Gustafson–Kessel (GK) clustering algorithm

Gustafson –Kessel clustering algorihm is one of the important data mining algorithms which is implemented in the member ship problem. This algorithm is based on fuzzy clustering logic. Gustafson-Kessel Algorithm is used in the membership algorithms which is used to find out either one point is belongs to one group or any other groups.

GK clustering has the power to find the cluster centers. It help to find the mean matrix of that particular dataset. GK clustering is used to overcome the limitations of the k means algorithm. The algorithm is more efficient and useful than the k means clustering. If the dataset is extremely larger to find the cluster means gk clustering is widely used.

B. Farthest First Clustering

Data from the different viewpoints can be analysed using clustering in the data mining process. Different classification of clustering algorithms are used to build clusters in the data mining. Here Farthest First clustering algorithm are efficiently used to form clusters. The particular is used in large datasets and form s clusters efficiently. The farthest first algorithm is mainly used to non-uniform clusters mainly.

The first point of the dataset is selected arbitrarily to form the clusters. After that each distance from the point will be calculated for finding the cluster heads.

Assume that all selected points will be from “l” length at least from each other. And also all the points in the space will be at most “L” distance from the subset.

Farthest first algorithms are widely used in many applications such as travelling sales man problems. The algorithm is based on the greedy approach to find out the clusters in the finite space of the dataset. This algorithm is easily used to find out the non-uniform clusters.

As a first step in the algorithm is to initialise the sequence of points. If the points are not initialized the following steps will be repeated

1. Iterate the remaining points which is not selected and find out the maximum distance from the other points.
2. Delete the point which is not selected from the points and add it to last of the points.

The procedure will repeats until the list become empty. The algorithm has the ability to form the non-uniform clusters and also the time consuming is very less in the case of formation of clusters.

PROCESS FLOW OF THE SYSTEM

The following data flow diagram will describe the different steps used in this procedure.

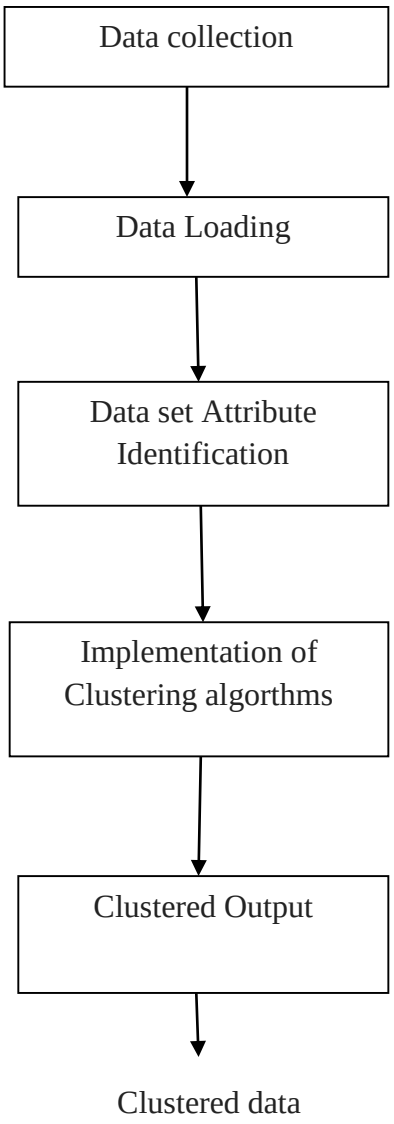

Fig. 2: Process flow of the system

The following are the different steps to be taken in the system.

i) Selection of data set

The paper is analyzing the management zones in the precision agriculture the input data is related to the soil features. The input data is taken from the attributes of the soil such as Co2_Flux_Exp, TSoil, and Soil_H2O etc.

ii) Preparation of Input data

Preparation of input data includes the pre-processing steps in the data mining. Pre processing helps to find the clean data from the selected input data. The dataset which is selected in the initial step may contain many impurities. So to remove or filter out all the impurities we need to use some pre-processing tools. The pre-processing tools such as WEKA tools helps to figure out the clean data. In this paper program codes are used to filter out the pre-processing data.

iii) Forming cluster heads using different algorithms

After the pre-processing step clustering algorithms are used to find out the different cluster heads. Cluster heads are formed using the following steps

1. Initiate start point, endpoint to denote the area of clustering
2. Start=0;
3. While(unclustered elements present)

Repeat

4. If (start~=0)
5. Start point = start point + th
6. Endpoint = endpoint- th

End if

7. Select randomly a point c present in the range from start time by using

The cluster heads formed will leads to the formation of clusters in the following steps.

iv) Absolute value formation

Input values are selected from the data set to find out the absolute value formation. Absolute value is find out to form the clusters easily.

v) Cluster formation

Clusters are formed using the different algorithms. The three algorithms such as K means, GK clustering and farthest first algorithms. Clusters are formed using the absolute value and cluster centres in the previous steps. Each clusters are formed with the cluster heads.

vii) Clustered data

Clusters are formed using the different algorithms. Clustered Management zones are formed using different algorithms. Steps are repeated for forming the clusters. Three algorithms are used and analysed the formed cluster features. The following are the procedure which is implemented in the paper.

- First pre-process the dataset
- After getting pre-processed dataset, we are going to the next loading GUI in order to load the datasets.
- Authentication and registration will be done before logging in the system with buttons Authentication, click for register, about us and contact us buttons.
- After authentication, we are redirecting into a page home.java.
- There we can choose our pre-processed dataset and form clusters of data using Farthest First Clustering method and displays all the output clusters in GUI.

ANALYSIS OF THE OUTPUT

A high configured machine is needed to run the application. Input data set can be retrieved with the help of WEKA tools or any other program codes. The application which is created is starts with the pre-processing activity to filter out the impurities of the data set.

The pre-processing activity should be done efficiently so that we can reduce the residual error .Hardware and software specification for the system used is given below.

Hardware Specification:

·Memory - 4 GB/ 8 GB RAM

. Processor- Intel Pentium

.Speed- Giga Hertz Speed

. Hard Disc Capacity- 1 TB

Software Specification:

·OS used in the System- Windows 7, 10

. Programming tools used-Java 1.8, MySQL, Net beans 8.2

Analysis of the input dataset is done using K means, GK clustering and Farthest First Algorithms. Total number of 695 instances are used to create the clusters. Each cluster is formed using the different algorithms.

After analysis clusters are formed using the data mining algorithms and the result of the performance is drawn as time and memory graph. The three algorithms are plot in the same graph with x and y coordinates.

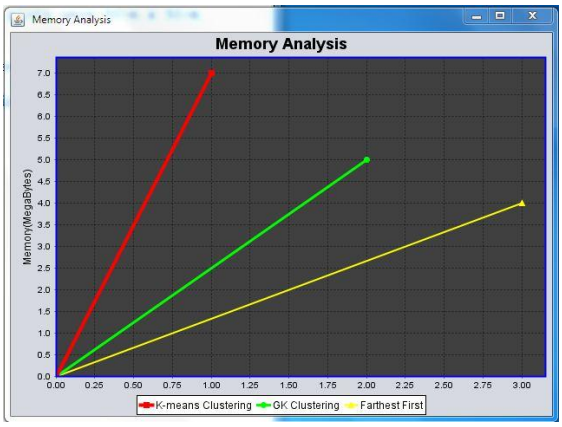

Fig. 3: The three algorithm memory analysis

In the above mentioned graph X axis denotes the each input for all the three algorithm and y axis represents the usage of the memory used in the process. From the figure itself it is understood that k means consumes memory a lot in the process while GK clustering uses less amount of memory. So in these case GK clustering uses the memory efficiently.

The below mentioned figure represents the time management for the three algorithms.

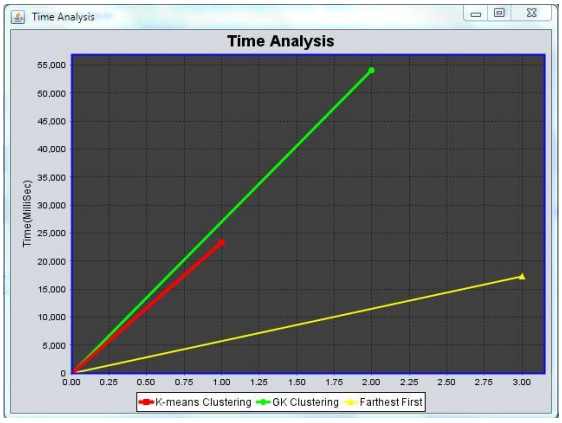

Fig. 4: The Three algorithm Time analysis

After getting the Time analysis graph it is very clear that farthest first algorithm is faster than the other two algorithms. Still the farthest first algorithm is helps to form the non-uniform clusters. This feature of farthest first will helps to manage the spatial factors in the precision agriculture.

On comparing with all the three algorithms from the graph, we concluded that the farthest first algorithm is the better one than others. On checking the time analysis GK clustering algorithm takes more time than the other algorithms and Farthest first takes the least. And also in the case of the memory analysis K means algorithm takes the more memory for operation than others and Farthest first takes the least.

CONCLUSION

The three algorithms are used to compare the analysis time and memory for constructing clusters in management zones. Management zones are the critical factor in the precision farming environment. So to manage the spatial factors effectively will increase the crop productivity.

Data mining is the trending technology which can be used efficiently in the field of depiction of management zones. More studies need to be there to find out the new trending contributions to enhance the agricultural sector.

REFERENCES

[1] E. Gustafson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In Proc. of IEEE CDC, 1979.[4] F. H¨oppner, F. Klawonn, R. Kruse, and T. Runkler.Fuzzy Cluster Analysis, Methods for classification, data analysis and image recognition.Wiley, 2000

[2] S.Nagaprasad,"Spatial Data Mining Using Novel Neural Networks for Soil Image Classification and Processing", International Journal of Engineering Science and Technology,2010. *doi*: 10.1007/s10844-006-9949-3

[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] T.A. Doerge, "Management zone concepts", Site-specific managementguidelines. Atlanta, USA: Potash and Phosphate Institute, 1999. IEEE DOI 10.1109/eScience.2014.42

[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] R. Dav´e. Characterization and detection of noise in clustering. Pattern Recognition Letters, 12:657–664, 1991.

[9] E.A.Speranza "A Cluster-Based Approach to Support the Delineation"2014 IEEE 10th International Conference on science doi:10.1109/eScience.2014.42

[10] R. Krishnapuram and J. Keller. A possibilistic approach to clustering. IEEE Trans. on fuzzy systems, 1:98–110, 1993. *DOI*: 10.1109/91.227387