Depiction of Management zones in Precision agriculture Using Farthest first Clustering Algorithm

[1] Mrs. Krishna priya C B, [2] Dr. S. Venkateswari
[1] Assistant Professor, Academy of Climate Change Education and Research, Kerala Agricultural University, Thrissur,Kerala,India
[2] Assistant Professor, Dept.Computer Science, L N Govt College,Ponneri, Thiruvallur,Tamilnadu

Abstract : Agriculture is making a good contribution to the economy of any country. So technical activity which enhancing the agricultural purpose will be a boon to the farmers.Using the term management zones a farmer can easily define any particular agricultural area. Clustering of spatial data will be the basement for this activity.Farthest first clustering algorithm is used for finding and managing the particular agricultural zones. Properly managing the agricultural zones will enhance the crop productivity.

Introduction

Data mining is the process ofextracting information from a large data sets. The patterns can easily be extracted with the help of different clustering methods such as partitioning, hierarchical, density based etc. Clustering is the process of grouping the same set or behaviour of elements together.Clustering helps to find the group of similar elements and also it is adaptable to changes as well. The Farthest first traversing clustering algorithms is similar to k means clustering algorithms in which itis used the technique of partitioning of the clusters.Farthest first traversing algorithms has several applications including travelling sales man problem. Here, to implement management zones in agriculture, the farthest first algorithm is used.

Management zones are the key terms used in precision farming. Precision farming is widely used in the country. In precision farming latest information Technology is used by the experts and give the proper advice to the farmers regarding their field. In such experiences spatial database is used for analysing and processing the management zones. Spatial data means information which is collected in the form of information related to the space such as soil,plant,diseases of the crop etc. In precision agriculture input data will be in the form of spatial data which is collected with the help of Information Technology Tools.

Management zones are the regions of the particular field in which homogenous data is uniformly distributed. So handling this management zones are very much important while fertilizing the land in precision agriculture. Equally distributed soil fertility may enhance the crop productivity. NPK (Combination of Nitrogen,Phosphorus and Potassium), Soil organic carbon, soil moisture etc. will together constitute the soil fertility. So distributing these parameters become important in soil fertility. So treating each region of the field as management zones of homogenous characters. So the technique of clustering is applied while delineating the management zones. Here the farthest first traversalalgorithm is used to manage such agriculture zones.

In this paper the Farthest first traversing algorithms are used to find the clusters of the soil.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.

Related Work

Delineation of management zones are best described by E.A.Speranza [1].Soil electric conductivity of the soil is used for the study. Using GK clustering algorithms geometric means of the clusters are efficiently depicted. The fuzzy C means algorithm and GK clustering algorithms are compared with soil parameters. 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 referredbyGarima 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.

Input data set

To manage the agricultural zones spatial data base is needed. Previous studies had takenNPK (Nitrogen, Phosphorus and Potassium) as one of the inputs for soil fertility. NPK is considered as the high crop producing fertilizer consisting of Nitrogen, Phosphorus and Potassium. Data set for the soil has been acquired from the site

https://esdac.jrc.ec.europa.eu/ESDB_Archive/Soil_Data/Global.htm

Other soilcharacteristics such as Co2, Soil PH,Tair and RH are taken as Input dataset. 83 instances are used for making clusters for finding the management zones. The input data set is given below

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

Fig (1) Sample Input

The figure (1) shows the sample input data set with the five attributes for the soil fertility. Analysing those attributes with their reference number.The initial inputs may consists of non-uniformed, unwanted data. So a filtering process may needed as an initial step. To pre-process the dataset a supporting software called weka tools is used. The entire contents of an input dataset are as follows.

Data Set Name	No Of Attributes	No of Instances
Dataset 1	5	83

Fig (2) Contents of input dataset

One dataset withfive attributes such as Co2, SoilPH, Tair and RH are taken as Input dataset. After preprocessing the resulting data set is free from anomaly and replicated data.

Methodology

This section describes the steps and procedure taken for implementing the best clustering algorithm to find the clusters with optimum result.

A. Process Flow of the System

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

i) Data selection

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

ii) 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.

iii) Cluster Head Formation

After Pre-processing, cluster head has to be formed.

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

These cluster heads are helpfulto form the clusters in the next steps.

iv) Finding the Absolute value

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.

Absoultedistance :Math. Abs()

v) Distance calculation

After finding the absolute value, distance between the values are found using the equation,

$$\text{Score}\,(X_i) = \sum_{j=1}^{m} f(x_{i,j}\,|D).$$

vi) An algorithm called the farthest first is used to find the Maximum distance between the values .The algorithm is explained in the next section.

vii) Cluster Output

As result of the clustering using the farthest first algorithm different clusters are generated. These clusters can be considered the management zones in the precision agriculture.Steps from iv to vi are done iteratively from the first instance to the last instance. Different clusters are formed using the farthest first algorithm. The above mentioned steps can be drawn in a flow chart as follows

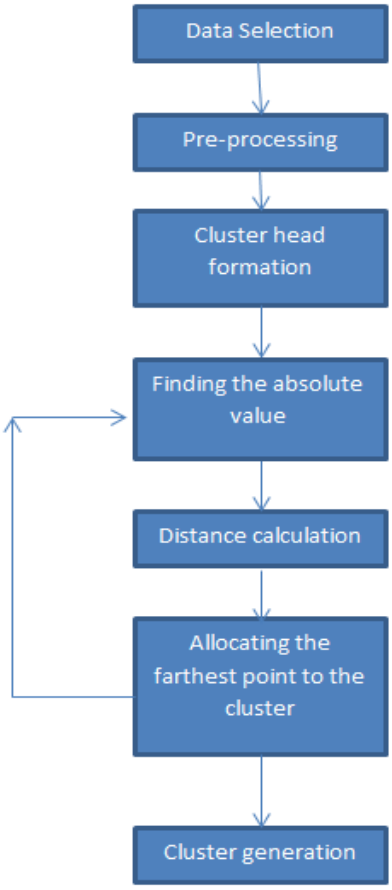

Fig (3) Architecture Diagram

B. Farthest first Algorithm

One of the fastest clustering algorithm is the farthest first algorithm. This algorithm is comes under the greedy method. As an initial step, k points are selected as cluster centers. The first point will be selected randomly. And the second point will be selected using the greedy method ie the point which is the farthest from the previously selected center.

The remaining enters will be selected in the same manner using greedy method. Remaining points will be added to the nearest centers.

The first point is selected deterministically,

For each Xi = [xi, 1, xi, 2… xi, m] in D that is described by m

Categorical attributes, we use f(xi,j |D)to denote thefrequency count of attribute value xi, j in the dataset. A score function is needed to analyse each point in the process which is given below:

Score $\sum j = 1, mf(xi, j\,|D)$

1. Farthest first traversal(D:input set k: integer) {
2. randomly select first center;
//select centers
3. for (I= 2,…,k) {
4. for (each remaining point)
{ calculate distance to the current center set; }

5. select the point with maximum distance as new center; }
//assign remaining points
6. for (each remaining point) {
9. calculate the distance to each cluster center;
10. put it to the cluster with minimum distance; }}

As a result of the farthest first 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. The Farthest first algorithm is similar to the k means clustering.

Using the farthest first algorithm 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.
Inter cluster distance is maximized while finding the farthest point from the center. This will cause the formation of non-uniform clusters.

Result Analysis

Input data set is pre-processed by the WEKA tools. The system which is used should have the good hardware and software configuration. The system has been build and run with the following configuration.
The system should have these hardware requirements:

•Screen resolution of 1024x768 for proper and complete viewing of screens.
•Intel Pentium IV 2.90GHz
•4GB 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
Analysed the input dataset which is described by the fig(1) and fig(2) with the help of the Farthest first algorithms, and found out the Farthest First algorithms is more faster due to its ability to include all elements to the clusters. 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.

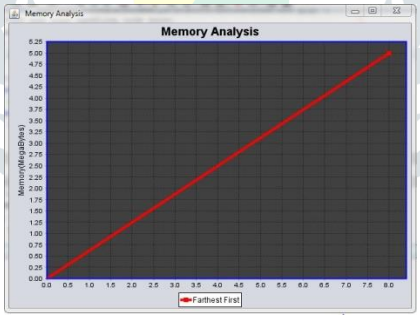

Fig (4) The farthest first algorithm memory analysis

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, as the Farthest first algorithm is used for large dataset. 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

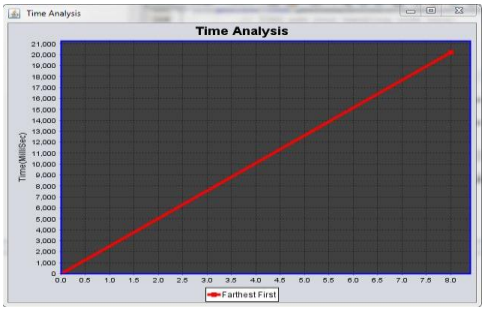

Fig 5) The Farthest first algorithm Time analysis

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

Algorithm	No Of Clusters	Ratio of each clusters	Unclustered instance
Farthest First	5	150:219:199:189:179	0

Fig (6) Output result set

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.

Conclusion

The Farthest first algorithm is used for finding the management zones in the precision agriculture. For the analysis Soil fertility characteristics are used as input dataset. Management zones can be considered as each cluster in the field. The clustering has done successfully with the help of the algorithm.

The algorithm is used for finding the farthest location from each cluster means. So the resulting clusters may include non-uniform data. The Farthest first algorithm is used for finding the clustering with large dataset.

References

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

[2]T.A. Doerge, "Management zone concepts", Site-specific managementguidelines. 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] 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

[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. 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.

[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