Agronomy data mining forecast extraction by plotted data stream fractal clustering on live geo map.

Shivangi Mehta, V.V.P. Engineering College, Rajkot.

Abstract:
Agronomics is the branch of economics dealing with the distribution, management and productivity of agriculture land. Agronomy is the science of soil management and crop production. In agriculture, role of both for achieving maximum crop yield at minimum cost is vast and to create its forecast results by using different data mining techniques is a goal of research. Process of taking a decision is as complex as there are several factors affecting entire farming process. extraction of data set by using data mining on past enterprise data ware house bulky data regarding the crop management, soil management, production forecasting, commodity cost forecasting etc by using probable matching crops to people according to basic inputs like water availability in mm, average temperature, average soil fertility of farm, locality of farm, soil type, crop duration etc by certain calculations at backend can provide proposed decision support system which is very useful in agriculture system to assist farmers for selection of a crop for cultivation mapping using different ground parameters like soil type, PH-value of soil, average weather required, required water consumption, temperature range, etc. This system used to increase productivity of crops by providing basic information and the list of the crops. Using of on-line analytical processing and data mining enriches knowledge base with new agriculture information in the form of data table and lists in user-friendly way by using graphical selector in very easy format.

Introduction:

Agronomics is the branch of economics dealing with the distribution, management and productivity of agriculture land.

Agronomy is the science of soil management and crop production

Cluster is a group of objects that belongs to the same class. In other words, similar objects are grouped in one **cluster** and dissimilar objects are grouped in another **cluster.**

Clustering is an unsupervised machine learning method that attempts to uncover the natural groupings and statistical distributions of data.

Fractal Cluster It is a main component of the model on which all the other components like plotting area, graphical selector, Data Populator, Data Grid, Total crop production, utilization of resources are placed.

Data Stream Fractal Cluster is in working window now because, Huge amount of streaming data is generated in recent years in the field of agriculture for example land types, ph value of land, water level, crop management etc., Stream data flows in and out of a system continuously and with varying update rates. They are fast changing, massive, potentially infinite and unbounded. By the virtue of Data stream characteristics, stream clustering became challenging due to limited memory and real time query response requirements. Massive volumes of data should be handled with limited memory. We cannot scan such huge amount of data more than once. New concepts may keep evolving in data streams over time. Evolving concepts require data stream processing algorithms to continuously update their models to adapt the changes. Fractal dimension is a

powerful tool to describe self-similarity, and changes in the correlation dimension imply changes of data distribution, which can be used to indicate changes in data trends. Summaries of the processed data help in computing important statistics of new clusters with the arrival of new points. In data stream mining the proposed algorithm should handle online clustering meritoriously and should maintain the clusters considering the potentiality of clusters. Time plays a major role in data stream clustering as a data point belonging to a cluster in some time horizon can become an outlier in some other time horizon as most recent data plays an important role. Similarly, in a cluster where there are no data points in one time horizon may accommodate more points in another time horizon making it more progressive. Therefore, the proposed algorithm should maintain a balance not to reject the outliers at once, but should wait for some time to remove it as outlier aiming to find clusters of arbitrary shape. A large number of clustering algorithms for data streams have been proposed, where the similarity of the objects is defined with use of some distance measure or objective function. The proposed algorithm uses correlation fractal dimension for finding arbitrary shaped cluster and further improving the precision of clustering. **Data points are merged into a cluster who's Relative Change in the Fractal Dimension (RCFD) is less than a minimum threshold and if no points are added to any of the clusters within a stipulated time horizon, then they are considered as real outliers. This paper extends the prior work using the concept of fractal dimension and multi layered grid.** Weights are assigned to the clusters using damped window model and pruning is done based on the importance of the clusters, which improve the results of the clustering.

Big Data is extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

Problems in current scenario and real time solution or working of research, Agriculture is a very vast field in terms of different agriculture techniques, crop management, types of land, world level commodity market position etc. so farmers can't predict at the time of new crop process thatwhich one is the best commodity for his land and region. Even though most of the countries in the world based on rainy water for agriculture so this is very unpredictable at the time of year starting that how much rain will be occurs. For past data set, technical issue is end- user may not be able to use the clustering technique efficiently. Thus study is carried out in this paper focuses on how to provide the effective clustering model that provides the utmost flexibility to end-user to create the cluster according to his/her requirement and convenience in shape as per their requirement by using graphical selector. User can applied different data mining patterns on past data set with huge number of records and extract analysed data set as per the requirement and used this kind of analysed data set for forecasting related to crop management, commodity forecasting, soil management etc.

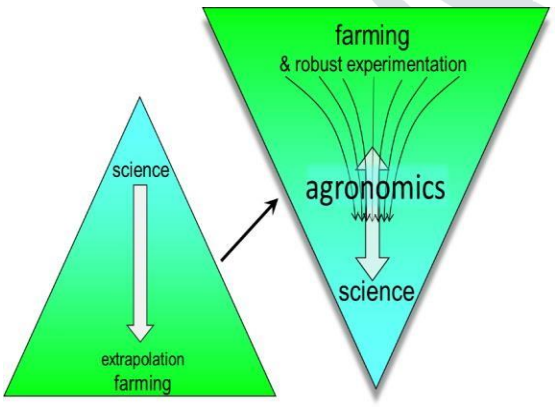

Figure a': role of agronomics and data science in the field of agriculture

Related Work

During the literature review papers related to fractal shape clustering was analyzed and observed. After the literature review authors have define the review findings in which it has been clearly seen the requirement of user customized fractal shape clustering model.

Literature Review

Dhiah Al-Shammary [1] proposed a fractal self-similarity model that provides a novel way of computing the similarity of SOAP messages. Fractal is proposed as an unsupervised clustering technique that dynamically groups SOAP messages. Lu- yong Wang [2] proposed a fractal clustering method to cluster genes using intrinsic (fractal) dimension from modern geometry. Fractal dimension is used to characterize the degree of self-similarity among the points in the clusters. The main idea of fractal clustering is to group points in a cluster in such a way that none of the points in the cluster changes the cluster's intrinsic dimension radically. The computation was performed with Hausdorff fractal dimension through the means of the box-counting plot algorithm, since it is the fastest and also robust enough. Lu-yong Wang [3] proposed a fractal clustering method for gene expression analysis, which use intrinsic dimension (fractal dimension) as metric for the clustering process. It relies on the intrinsic relationship among points in the clusters. Assessment carried out based on gene annotation. Zhiwen Yu [4] investigated the problem of mining uncertain data in a low dimensional subspace that is addressed for the first time. The major contribution is a fuzzy dimensional induced clustering (FDIC) algorithm that combines fuzzy distance function and fractal correlation dimension to find out the interesting clusters in the low dimensional manifolds in an uncertain database. Also fuzzy kth nearest neighbor algorithm (FkNN) was proposed to retrieve the kth nearest neighbor

efficiently. GuangHui Yan [5] proposed a Fractal-based Cluster Hierarchy Optimization (FCHO) algorithm that integrates the cluster similarity with the cluster shape and the cluster distribution to construct cluster hierarchy tree from the disjoint initial clusters. Robert S. H. Istepanian [6] presented a comparative analysis of two genomic signal processing methods for robust micro-array data clustering based on Fractal Dimension and Discrete Wavelet Decomposition with Vector Quantization are validated for standard data sets. Guoping Lin [7] proposed GFDStream algorithm based on the usage of the fractal dimension. It is insensitive to noise, capable of finding clusters of arbitrary shape in data streams and capable of dealing with points of high dimensionality and it is an effective and efficient method for clustering an evolving data stream. ZENG XIU [8] proposed a method K-line, which reflects the trend of stock and it is a fractal graphics with a stable fractal dimension where a fractal dimension is used as an important parameter in the research of stock cluster analysis. Christopher C. Yang [9] proposed the SDC algorithm for Web opinion analysis. The SDC algorithm overcomes the weakness of DBSCAN algorithm by grouping less number of less relevant clusters together when they are density-reachable. Experimental work carried out by utilizing both SDC and DBSCAN algorithms to cluster the major themes. Dhiah Al-Shammary [10] proposed an unsupervised auto class Fractal clustering technique for clustering SOAP messages into a dynamic number of clusters according to their Fractal similarities. The proposed Fractal clustering technique potentially reduces the required processing time in comparison with other standards. Guanghui Yan [11] proposed a new solution adopting the fractal dimensionality reduction, cluster evolution tracking and self-adaptive sampling technique with time decaying data window scheme. Here the cluster evolution detection algorithm was presented, which integrates the adaptive fractal dimensionality reduction and can discriminate the cluster evolution of stream data. Dhiah Al-Shammary [12] proposed Fractal clustering model to compute the Fractal clustering similarity of SOAP messages in order to cluster them and enable the aggregation of SOAP messages to significantly reduce the size of the aggregated SOAP messages. Thus two fast Fractal clustering models have been proposed that are aiming to reduce the required clustering time.

Review Finding

As it is observed in above defined literature findings that clustering technique is very useful technique to provide effectiveness in various business tasks and operations. Here user may require to implement clustering in any shape may be square, rectangle, triangle, circular and even fractal shape. Thus clustering technique can be very effective and mode capable if provides the facility of defining fractal shape clustering tasks explored in [1,2,3,5,6,7,8,10,11,12].

Usage of almost all the clustering technique and models requires that end- user must have deeper understanding of techniques to be applied and thorough understanding of algorithm to be used for clustering implementation [2,4,5,11,12]. End-user may or may not be able to understand such complicated algorithm and its know-how about how to use the techniques. Thus challenge is to provide a model which requires very less prerequisite knowledge to use the clustering technique and very easy process to be understood. Another problem is inflexibility where majority of the clustering techniques and models create the cluster automatically whenever user provides the input and that may not be according to the user's expectation and requirement. Thus decrease the effectiveness of clustering operations and efficiency also. Thus another challenge is

to provide facility where user can create the cluster according to his desire and requirement.

Proposed work with research application components introduction

First of all authors proposed the design of fractal shape clustering model and on the base of design the model was develop to provide the facility of fractal shape clustering and that tool is user customized.

GEO mapping out put in the form data table and LIVE GEO MAP in terms of agronomy with world level commodity market rates forecasting based on data mining patterns on past data set is looks like,

Figure a'': live geo mapping with commodity rates forecasting looks like.

Cluster Table to store sample records of agriculture data set,

Researcher created a Cluster Database as sample database to store the data related to clustering activities. Here researcher has created cluster table from which records are fetched and displayed on the Plotting area. Also all the records selected by graphical selector are first compared with the field lx and ly of cluster table and then stored into the data grid. Cluster table is having following structure.

(a) Cluster Dataset:

Researcher created a Cluster Database as sample database to store the data related to clustering activities. Here researcher has created cluster table from which records are fetched and displayed on the Plotting area. Also all the records selected by graphical selector are first compared with the field lx and ly of cluster table and then stored into the data grid.

Structure of Cluster Table			
No.	**Field Name**	**Data Type**	**Description**
1	F_No	Number	Field N.+ Primary Key
2	Lx	Number	X coordinates Value
3	Ly	Number	Y Coordinates Value
4	Crop_Approx_Production	Number	Crop Produce Taken Place at Lx and Ly Coordinates

F_No.	lx	ly	Crop_Approx_Production (Quintal / Acre)
1	80	80	10
2	160	80	15
3	240	80	20
4	320	80	25
5	400	80	35
6	480	80	45
7	560	80	30
8	80	160	20
9	160	160	25
10	240	160	10
11	320	160	10
12	400	160	15
13	480	160	20
14	560	160	25
15	80	240	35
16	160	240	10
17	240	240	15
18	320	240	20
19	400	240	25
20	480	240	35

(a) Fractal Cluster

It is a main component of the model on which all the other components like plotting area, graphical selector, Data Populator, Data Grid, Total crop produce and Average Crop Production are placed.

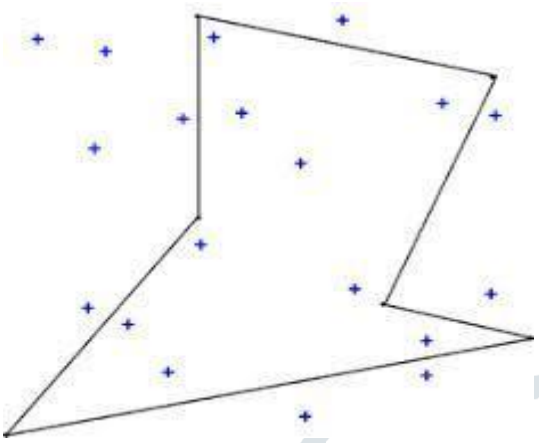

Figure a, Fractal Cluster, the main component of the Clustering Model to extract plotted data in the form of tabular dataset by taking a Fractal graphical selector.

Figure a', Real time image of farm field extract plotted data by taking a Fractal graphical selector.

(b) Fractal Graphical Selector

Graphical Selector is the component of the model that is used to select records from the plotted area. Researcher has provided the way where user can create a fractal shape cluster, for this End-user has to click on the plotting area where he/she wants to create cluster. The Fractal shape cluster will be drawn as user moves mouse pointer to select and bound records within the plotting area. Once user release the left mouse button a line will be automatically drawn between the starting point and the ending points of the cluster. This will create the fractal shape and all the desired records will be bounded by this closed fractal shape.

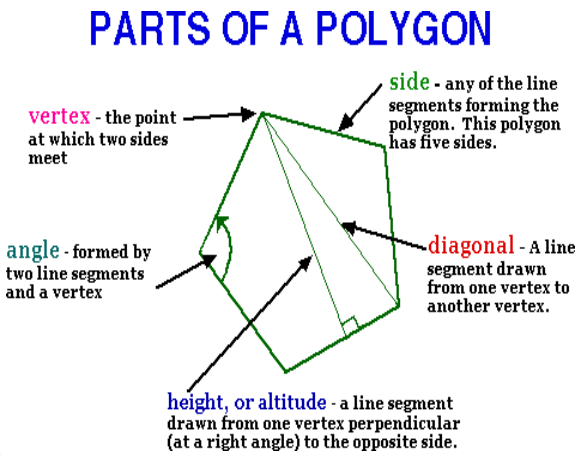

Figure b', Graphical Selector also called illustrator polygon, the component of the Clustering Model to extract plotted data in the form of table by taking a Fractal shape as graphical selector.

Figure b'', Enlarged view of plotting area and graphical selector of figure b'.

zip	latitude	longitude	city	state	county
501	40.92233	-72.6371	Holtsville	NY	Suffolk
544	40.92233	-72.6371	Holtsville	NY	Suffolk
601	18.16527	-66.7226	Adjuntas	PR	Adjuntas
602	18.3931	-67.181	Aguada	PR	Aguada
603	18.45591	-67.1458	Aguadilla	PR	Aguadilla
604	18.49352	-67.1359	Aguadilla	PR	Aguadilla
605	18.46516	-67.1415	Aguadilla	PR	Aguadilla
606	18.17295	-66.9441	Maricao	PR	Maricao
610	18.28869	-67.1397	Anasco	PR	Anasco

Figure b''', Tabular data set covered by using graphical selector on geo spatial database on live map with exact longitude and latitude

(c) Data Populator

Data Populator is the component of the model that performs the process to display records on the plotting area. When user clicks on this component, it starts the process of fetching records from the table and displays the records on the plotting area.

Figure c, Data Populator, the component of the Clustering Model to extract plotted data in the form of table by taking a circle as graphical selector

(d) Analyzed Data Grid

Data Grid is the component of the model that is responsible to store & display the records selected by graphical selector it shows the conclusive extraction data in tabular format by applying different data mining techniques like association rule, regression, summarization, clustering etc. and provide the knowledge extraction to the decision maker.

	Indicators	Value	Rating	Constraint
PHYSICAL	Aggregate Stability (%)	83.6	99	
	Available Water Capacity (m/m)	0.17	59	
	Surface Hardness (psi)	233	24	rooting, water transmission
	Subsurface Hardness (psi)	325	36	
BIOLOGICAL	Organic Matter (%)	5.3	91	
	Active Carbon (ppm) [Permanganate Oxidizable]	566	40	
	Potentially Mineralizable Nitrogen (µgN/ gdwsoil/week)	17.2	100	
	Root Health Rating (1-9)	5.0	50	
CHEMICAL	pH (see Nutrient Analysis Report)	6.1	67	
	Extractable Phosphorus (see Nutrient Analysis Report)	3.1	44	
	Extractable Potassium (see Nutrient Analysis Report)	37.8	33	
	Minor Elements (see Nutrient Analysis Report)		100	
OVERALL QUALITY SCORE (OUT OF 100):			62.0	Medium
Soil Textural Class:==> silt loam				
SAND (%): 45.6		*SILT (%): 52.5*		*CLAY (%): 1.9*

Figure d, Data Grid, with land health card.

(e) Time Series Analysis with diff. Forecasting:

Time Series is a sequence of well- defined data points measured at consistent time intervals over a period of time. Data collected on an ad-hoc basis or irregularly does not form a time series. Time series analysis is the use of statistical methods to analyze time series data and extract meaningful statistics and characteristics about the data. It helps us understand what are the underlying forces leading to a particular trend in the time series data points and helps us in forecasting and monitoring the data points by fitting appropriate models to it.

The biggest advantage of using time series analysis is that it can be used to understand the past as well as predict the future. Further, time series analysis is based on past data plotted against time which is rather readily available in most areas of study.

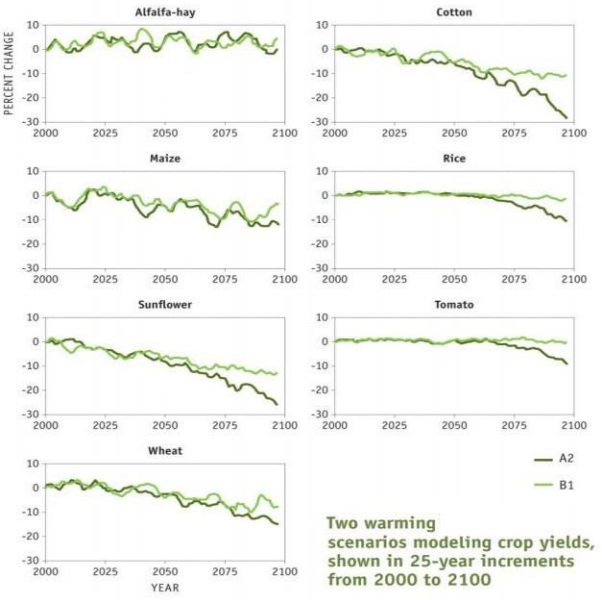

Figure e, Finally based on different data- mining patterns, data gathering and region selected from map forecast for crop yield is generated with time series analysis graphical representation.

Reference

[1] Fractal self-similarity measurements based clustering technique for SOAP Web messages by Dhiah Al-Shammary, Ibrahim Khalil, Zahir Tari, AlbertY.Zomaya in 0743-7315©2013 Elsevier Inc.

[2] Fractal Clustering for Microarray Data Analysis by Lu-yong Wang, Ammaiappan Balasubramanian, Amit Chakraborty, Dorin Comaniciu in Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference Workshops (CSBW'05) 0-7695-2442-7/05 © 2005 IEEE

[3] Fractal Clustering and Knowledge-driven Validation Assessment for Gene Expression Profiling by Lu-yong Wang, Ammaiappan Balasubramanian, Amit Chakraborty, Dorin Comaniciu in 0-7803-8740-6/05©2005 IEEE.

[4] Mining Uncertain Data in Low-dimensional Subspace by Zhiwen Yu, Hau-San Wong in The 18th International Conference on Pattern Recognition (ICPR'06)0-7695-2521-0/06©2006 IEEE

[5] Using Cluster Similarity to Detect Natural Cluster Hierarchies by GuangHui Yan ,ZhanHuai Li in Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007) 0-7695-2874-0/07 © 2007 IEEE

[6] Fractal Dimension and Wavelet Decomposition for Robust Microarray Data Clustering by Robert S. H. Istepanian, Ala Sungoor, and Jean-Christophe Nebel in 978-1-4244-1815-2/08 ©2008 IEEE.

[7] A Grid and Fractal Dimension-Based Data Stream Clustering Algorithm by Guoping Lin, Leisong Chen in 978-0-7695-3494-7/08© 2008 IEEE

[8] CLUSTERING IN STOCK MARKET BASED ON FRACTAL THEORY by ZENG XIU, PENG HONG, ZENG ZHEN in 978-1-4244-3703-0/09©2009 IEEE, Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009

[9] Analyzing and Visualizing Web Opinion Development and Social Interactions With Density-Based Clustering by Christopher C. Yang and Tobun Dorbin Ng in IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART A: SYSTEMS AND HUMANS, VOL. 41, NO. 6, NOVEMBER 2011,1083-4427© 2011 IEEE [10] Dynamic Fractal Clustering Technique for SOAP Web Messages by Dhiah Al-Shammary and Ibrahim Khalil in 2011 IEEE International Conference on Services Computing, 978-0-7695-4462-5/11 © 2011 IEEE

[11] Integrating Fractal Dimensionality Reduction with Cluster Evolution Tracking by Guanghui Yan and Yu Xu, Xin Shu and Xiang Li, Minghao Ai and Zhicheng Ma in 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 978-1-61284-181-6/11 ©2011 IEEE [12] Clustering SOAP Web Services on Internet Computing Using Fast Fractals by Dhiah Al-Shammary, Ibrahim Khalil and Loay E. George in 2011 IEEE