JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Prediction abstraction from agricultural data mining using real geographic plotted graph

Jay Poojara, Sweta Patel
Research Scholar (M.Tech), Assistant Professor
Computer Engineering,
RK University, Rajkot, India.

Abstract:

In agriculture, both play a significant role in generating maximum crop output at the lowest cost, and developing prediction outcomes using data mining methods is a research goal. The decision-making process is complicated by the fact that various elements influence the entire farming process. Retrieval of data set from previous business data warehouse bulky data about agricultural crops, soil conservation, productivity forecasting, resource cost forecasting, and so on. Also using basic inputs such as mean temperature, water prediction, land productiveness, location of farm, kind of land, and so on, it is possible to match crops to people. Ionic strength of soil, average weather needed, necessary water usage, temperature range, and so on. This method was designed to boost crop productivity by delivering basic information and a list of crops. Through use of on-line efficiency analysis and information mining enhances understanding and acquaintance by providing fresh farming data in the form of datasets and lists in a customer style with a simple graphical selector.

1. Introduction

A cluster is a collection of objects from the same class. To put it another way, cluster has many associated objects combined together. Clustering is kind of unsupervised learning technique for discovering natural data groupings and statistical distributions.

Data Stream Nonlinear Cluster is currently in the working window. Stream data is constantly flowing in and out of a system, with varied update rates. They are dynamic, enormous, possibly infinite, and limitless. Due to the limited memory and real-time query response requirements imposed by Data Stream features, stream clustering became difficult. Lot of data should be managed with a small amount of memory. We can't possibly scan such a large amount of data twice. Data streams may continue to evolve new notions over time. As more points are added, the features of the processed data help determine the vital statistics of new clusters. The proposed data stream mining technique should be able to successfully handle realtime networking and ensemble maintenance while also considering cluster potential. Because data stream clustering prioritises the most recent information, a data point affiliated with a clustered in one time window may become an anomaly in another. Likewise, in a cluster with no pieces of data in one time horizon, more elements in another timeframe horizon can be accommodated, making it more advanced. As a result, the suggested algorithm should strike a compromise between rejecting outliers immediately and waiting a period of time before removing them as outliers in order to locate clusters of any shape. Many clustering techniques for streaming data have been developed, in which the similarity of something like the elements is defined using some dissimilarity measure or target value. For identifying arbitrary shaped clusters and boosting clustering precision, the suggested approach employs correlation fractal dimension. Data points are regarded actual outliers frame. Using the idea of nonlinear model and a multi-layered grid, this research builds on previous work. Weights assigned to clusters using a damped window model, and trimming is done depending on cluster importance, which improves clustering outcomes.

Big Data is a term used to describe extraordinarily massive data collections that may be computationally analysed to uncover patterns, trends, and relationships, particularly in the context of human behaviour and relationships.

Problem Statement

Agriculture is such a large field in terms of varied agriculture practises, crop management, land types, global commodity market position, and so on that farmer can't forecast which commodity is the best for their land and region when planting a new crop. Even though most countries in the globe rely on rainwater for agriculture, the amount of rain that falls is very unpredictable depending on the time of year. The technical issue with the previous data set is that the end-user may not have been able to apply the clustering

technique effectively. As a result, the research conducted in this study focused by utilising a graphical selector. The user can utilise multiple data mining techniques on previous data sets with a large number of records.

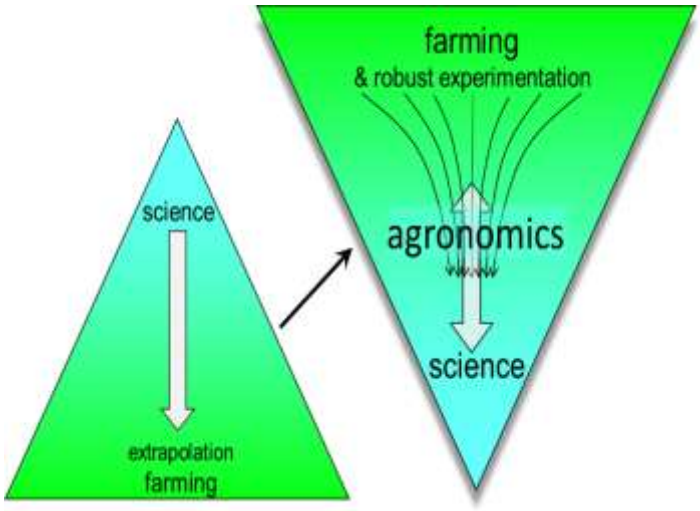


Figure 1: Role of Data Science and Agronomics

2. Related Work

Papers on fractal shape clustering were analysed and noticed during the literature review. Following the literature research. In [1], D. Shammary recommended a nonlinear model to compute the resemblance of SOAP messages in a novel method. Fractal is a technique for dynamically grouping SOAP messages that is offered as an unstructured clustering methods. In [2], L. Wang proposed a novel clustering approach for identifying proteins based on current geometry's inherent (fractal) dimension. The level of distinctiveness among the parts in the clusters is measured using the fractal dimension. Fractal clustering's primary principle is to arrange points within a cluster so that none of them change the cluster's inherent dimension. The calculations were carried out using the Multifractal Initiatives and the best and most accurate box-counting visualisation technique. In [3], L. Wang [3] presented a novel method utilizes internal dimension (nonlinear dimension) as the clustering metric. It is based on the inherent link between the elements in the clusters. Gene annotation was used to make the assessment. For the first time, Zhiwen Yu [4] looked into the challenge of extracting fuzzy numbers in a low-dimensional subspace. The primary strength is a fuzzy dimensionality influenced clustering (FDIC) method that uses a combination of fuzzy distance measure and fractal inference to locate interesting clusters in low-dimensional surfaces in an ambiguous database. To obtain the kth nearest neighbour effectively, the fuzzy kth nearest neighbour algorithm (FkNN) was also devised. GuangHui Yan [5] introduced a Fractal-based Cluster Hierarchy Optimization (FCHO) technique for constructing a cluster hierarchy tree from disjoint initial clusters that incorporates cluster similarity, cluster form, and cluster dispersion. In [6], For typical data sets, Robert compared two chromosomal signal processing techniques based on Nonlinear Scaling and Discrete Wavelet Reduction with Data Compression for accurate precision data grouping. In [7], G. Lin created the Stream technique, which itself is based on the use of fractal dimensions. It is speckle, capable of detecting random clusters in data feeds, and capable of dealing with high-dimensionality points, making it a good choice for clustering a dynamic data stream. In [8], Z. Xiu introduced the K-line method, which depicts stock trend and is a creative visuals with a steady fractal dimension, with the fractal dimension being an essential property in stock cluster analysis study. For Web opinion analysis, Christopher C. Yang [9] introduced the SDC method. When density-reachable clusters are grouped together, the SDC method solves the weakness of the DBSCAN algorithm by combining a smaller group of less essential clusters together. To cluster the primary themes, experimental work was done using both SDC and DBSCAN algorithms. For grouping SOAP packets into a flexible number of clusters based on their Fractal similarities, Dhiah Al-Shammary [10] suggested an unsupervised auto class Harmonic clustering algorithm. In compared to existing methods, the suggested Fractal clustering algorithm has the potential to cut processing time. Guanghui Yan [11] presented a novel approach that combined fractal dimensional reduction, cluster development tracking, and self-adaptive sampling with a time-decreasing data window scheme. The cluster history detection technique, which incorporates adaptable nonlinear dimension reduction and can identify the cluster development of stream data, was given here. Dhiah Al-Shammary [12] suggested a Spiral SOAP message aggregation. As a result, two rapid Nonlinear clustering models are presented, with the goal of reducing the clustering time.

Clustering approach is a very important technique for providing effectiveness in many business tasks and processes, as evidenced by the above-mentioned literature findings. In this case, the customer may need clustering in different shapes. As a result, if a clustering technique allows for the definition of fractal shape clustering tasks, it can be exceedingly effective and mode capable. Almost all clustering approaches and models necessitate a deeper comprehension of the techniques to be implemented and a complete understanding of the algorithm to be utilised for clustering implementation on the part of the end-user. The end-user may or not be able to comprehend such a complex method, as well as its knowledge of how to use the strategies. As a result, the goal is to create a model that requires very little prior knowledge to apply the clustering method and that is simple to comprehend. Another issue is inflexibility, as most clustering approaches and models form clusters automatically whenever the user provides input, which may or may not be in accordance with the user's expectations and requirements. As a result, the effectiveness and efficiency of

clustering activities are reduced. Another problem is to give a service that allows the user to design the cluster appropriate to his or her preferences.

3. Proposed Work

First, the authors suggested a design for a fractal shape clustering model, and based on that design, a model was developed to provide fractal shape clustering capability, with the tool being user customizable.

In terms of agronomy, GEO mapping is output in the form of a data table and a LIVE GEO MAP, with world level global commodity rates forecasts based on data mined patterns on previous data sets as shown in Fig. 2.

Figure 2: Real geo mapping with crop rate forecasts

4. Sample records from the agriculture data set will be stored in a Cluster Table

As a sample database, the researcher constructed a Cluster Database to contain data connected to clustering activities. Here, the researcher has established a cluster table from whom records are retrieved and plotted. In addition, all records picked by the graphical selector are compared to the cluster table's fields Lx and Ly before being put in the data grid. The structure of a cluster table is as follows.

Cluster Database

As a sample database, the researcher constructed a Cluster Database to contain data connected to clustering activities. Here, the researcher has established a cluster table through which records are retrieved and plotted. In addition, all records picked by the graphical selector are compared to the cluster table's fields Lx and Ly before being put in the data grid as shown in Table 1 and Table 2.

Table 1. **Database Table**

Number	Column Name	Data Type	Details
1	Field_No	Number	Field Number and Primary Key
2	Lx	Number	Value of X coordinate
3	Ly	Number	Value of Y Coordinate
4	Crop_Production	Number	Crop Produce happened at Lx and

Table 2. Cluster Database

Field_No	**Lx**	**Ly**	**Crop_Production (Quintal / Acre)**
1	40	40	5
2	80	40	10
3	120	40	15
4	160	40	20

5	200	40	30
6	240	40	40
7	280	40	25
8	40	80	15
9	80	80	20
10	120	80	5
11	160	80	5
12	200	80	10
13	240	80	15
14	280	80	20
15	40	120	30
16	80	120	5
17	120	120	10
18	160	120	15
19	200	120	20
20	240	120	30

Fractal Cluster

By using a Fractal graphical selector, the Harmonic Cluster, the main component of a Clustering Model, may extract displayed data in the form of a tabular dataset as shown in Fig. 3.

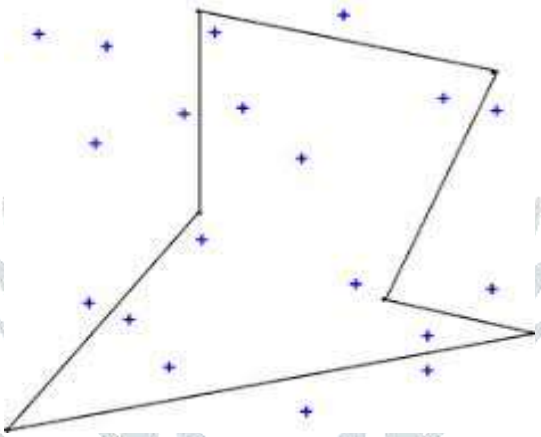


Figure 3: Fractal graphical selector

Using a Fractal graphical selection, extract plotted data from a real-time image of a farm field as shown in Fig. 4.

Figure 4: Real-time image of a farm field

Fractal Illustrator Nominator

The model's Graphical Selector component is used to choose records from of the plotted region. The researcher has supplied a method for users to generate fractal shape clusters. To do so, users must first click on the plotting region where they want to make the cluster. The Fractal shape cluster would be drawn when the user moves the mouse cursor about the plotting region. When the left mouse button is unpressed, a line would be drawn immediately between the beginning and ending points of cluster. The fractal shape will be generated as a result of this, and so this enclosed fractal shape will contain all of the needed records.

Graphical Selector, also known as illustrator polygon, is a feature of the Clustering Model that uses a Fractal shape as a graphical selector to retrieve plotted data is in the form of a table as shown in Fig. 5.

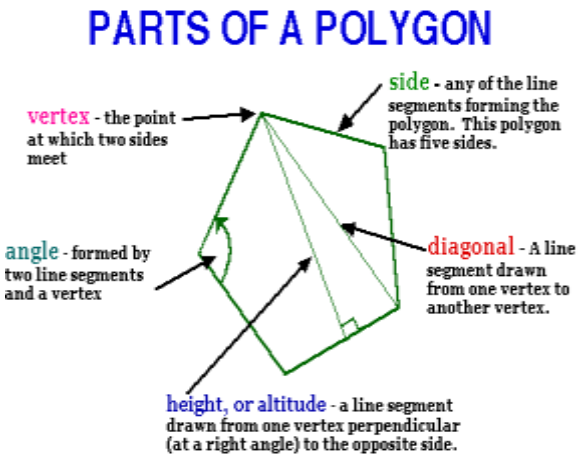


Figure 5: Illustrator Polygon

View of the plotting area and the graphical selection in further detail as shown in Fig. 6.

Figure 6: View of the plotting area

On a map viewer with precise longitude and latitude, a tabular data set was covered by utilising a graphical selection on a geo spatial database as shown in Fig.7.

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 7: Tabular Data Set

Populating Data

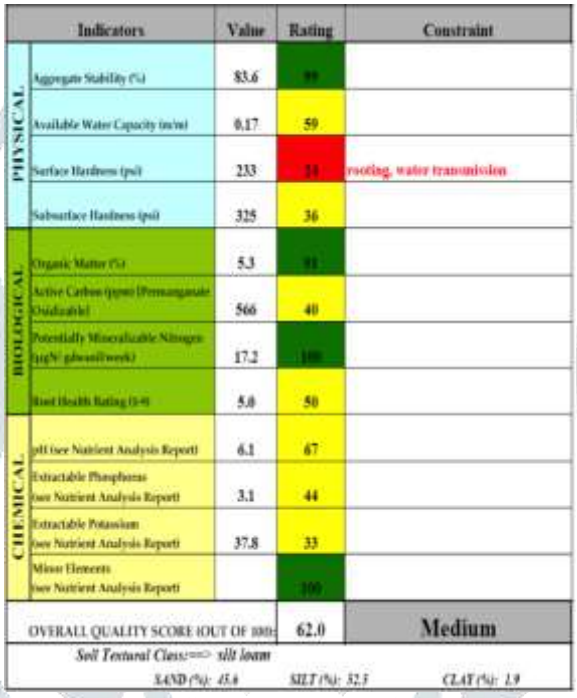
The Data Populator element of the model is responsible for displaying entries in a graphical way. The process of getting values either from the database and putting them on the charting creative expression whenever the user selects this component.

Data Populator is a Clustering Model component that uses a circle as a graphical selector to retrieve plotted data is in the form of a table as shown in Fig. 8.

Figure 8: Populating Data

Data Grid Analysis

The Data Grid is a model component that is responsible for storing and displaying the records chosen by the graphical selector as shown in Fig. 9. It displays the conclusive extraction data in tabular format using various data mining techniques such as association rule, regression, summarization, clustering, and so on, and provides knowledge extraction to the decision maker.



	Indicators	Value	Rating	Constraint
PHYSICAL	Aggregate Stability (%)	83.6	99	
	Available Water Capacity (in/in)	0.17	59	
	Surface Hardness (psi)	233	34	rooting, water transmission
	Subsurface Hardness (psi)	325	36	
BIOLOGICAL	Organic Matter (%)	5.3	91	
	Active Carbon (ppm) [Permanganate Oxidizable]	566	40	
	Potentially Mineralizable Nitrogen (ppN/ gds soil/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 9: Datagrid for knowledge extraction

Differential Forecasting using Time Series Analysis

A time series is a collection of very well data points collected at regular intervals over time. A time series is not formed by data acquired on an ad hoc or irregular basis. The application of statistical tools to analyse time series data to extract useful statistics and features about the data is known as time series analysis. It assists us in understanding through the application of appropriate models.

The most significant benefit of employing time series data is that it could be used to both understand and predict the past. Furthermore, forecasting model is based on historical displayed against time, which is common in most fields of research.

Finally, a traditional statistical graphical representation is developed based on multiple data-mining strategies, data gathering, and a region selected from a map prediction for crop yield as shown in Fig. 10.

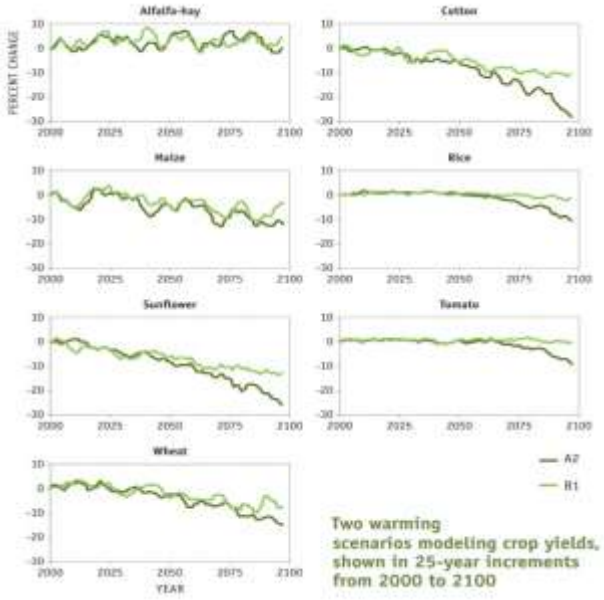


Figure 10: Traditional Statistical Graphical Representation

5. Conclusion

The fact that many factors influence the entire farming process complicates the decision-making process. Retrieval of bulky data about agricultural crops, soil conservation, productivity forecasting, resource cost forecasting, and other topics from a previous business data warehouse, this strategy was supposed to increase crop yield. The use of on-line efficient analysis and data mining enhances knowledge and understanding by providing fresh agricultural style with a simple graphical selector.

6. References

[1] D. Shammary et.al. "Fractal self-similarity measurements based clustering technique for SOAP Web messages", Elsevier, 2013.

[2] L. Wang et.al. "Fractal Clustering for Microarray Data Analysis", Computational Systems Bioinformatics Conference Workshops, IEEE, 2005

[3] L. Wang et.al. "Fractal Clustering and Knowledge-driven Validation Assessment for Gene Expression Profiling", IEEE, 2005

[4] Z. Yu et.al. "Mining Uncertain Data in Low-dimensional Subspace", International Conference on Pattern Recognition, IEEE, 2006

[5] G. Yan et.al. "Using Cluster Similarity to Detect Natural Cluster Hierarchies", 4th International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, 2007

[6] Robert et.al. "Fractal Dimension and Wavelet Decomposition for Robust Microarray Data Clustering", IEEE, 2008

[7] G. Lin et.al. "A Grid and Fractal Dimension-Based Data Stream Clustering Algorithm", IEEE, 2008

[8] Z. Xiu et.al. "Clustering In Stock Market Based On Fractal Theory", 8th International Conference on Machine Learning and Cybernetics, IEEE, 2009

[9] C. Yang et.al. "Analyzing and Visualizing Web Opinion Development and Social Interactions With Density-Based Clustering" , IEEE Transactions On Systems, Man, And Cybernetics, IEEE, 2011

[10] D. Shammary et.al., "Dynamic Fractal Clustering Technique for SOAP Web Messages", International Conference on Services Computing, IEEE, 2011

[11] G. Yan et.al. "Integrating Fractal Dimensionality Reduction with Cluster Evolution Tracking", International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, 2011

[12] D. Shammary et.al. "Clustering SOAP Web Services on Internet Computing Using Fast Fractals" IEEE, 2011.