



**HAL**  
open science

## Application of DBSCAN Algorithm in Precision Fertilization Decision of Maize

Yang Li, Guowei Wang, Yu Chen, Yang Jiao, Haijiao Yu, Guogang Zhao

► **To cite this version:**

Yang Li, Guowei Wang, Yu Chen, Yang Jiao, Haijiao Yu, et al.. Application of DBSCAN Algorithm in Precision Fertilization Decision of Maize. 11th International Conference on Computer and Computing Technologies in Agriculture (CCTA), Aug 2017, Jilin, China. pp.453-459, 10.1007/978-3-030-06179-1\_45 . hal-02111561

**HAL Id: hal-02111561**

**<https://inria.hal.science/hal-02111561v1>**

Submitted on 26 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Application of DBSCAN Algorithm in Precision Fertilization Decision of Maize

Yang Li<sup>1,2</sup>, Guowei Wang<sup>1,2,3</sup>, Yu Chen<sup>1,2</sup>, Yang Jiao<sup>4</sup>, Haijiao Yu<sup>5</sup>, Guogang Zhao<sup>5(✉)</sup>

<sup>1</sup>School of Information Technology, Jilin Agricultural University, Changchun 130118,China;

<sup>2</sup>Jilin province research Center, Changchun 130118,China;  
547229667@qq.com, 123929697@qq.com

<sup>3</sup>School of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China;

41422306@qq.com

<sup>4</sup>Sixth Middle School in Changchun, Changchun, 130000, China;  
61516131@qq.com

<sup>5</sup>The City College of JLJU, Changchun, 130000, China;  
haihai951k1@163.com, zhaogg2010@foxmail.com;

**Abstract.** In the current era of big data, information technology is developing quite rapidly, the most important data mining technology in information technology is also widely used, and now it is applied to the field of agricultural production, what can solve many problems such as agricultural production, fertilization and so on. In this paper, data mining technology is applied to the process of corn fertilization, because in corn production, effective and reasonable amount of fertilizer can make corn grow better, however, if there is no specific fertilization according to the soil properties of the corn, it will lead to the soil which needs fertility can not be with enough fertility, and the soil without fertility will be added more and more. In view of this problem, the soil planted with corn was graded and treated with different levels of soil, so as to achieve the purpose of effective utilization of soil fertility. In this paper, the DBSCAN algorithm in clustering analysis is used to classify the soil, the DBSCAN algorithm to this field have not been reported so far. By applying the nutrient balance method, the amount of soil fertilizer was calculated at each level, and the fertilizer was targeted according to the amount of fertilizer. Through the pilot application in Nong'an County of Jilin province Chen hometown, compared with the traditional fertilization results, Fertilizer reduced by 25%, corn production increased by about 15%, effectively reducing the input of chemical fertilizer and increasing the output of crops.

**Keywords:** Data mining · Cluster analysis · Soil classification · DBSCAN algorithm

## 1 Introduction

Precision agriculture developed in the early 1980, and precision agriculture is a subject jumping. Precision agriculture is the main system of modern agricultural production operation according to the spatial location, timing, spatial variability and quantitative to achieve, it is mainly the information technology to support, most of the data mining technology<sup>[1]</sup>.

Precision agriculture is mainly composed of intelligent decision-making technology, field information collection technology, and intelligent equipment technology. Through the collection of farmland information for digital analysis and processing, intelligent decision-making, mechanization of farmland information and application of information, can increase crop production, increase efficiency and increase the income of farmers<sup>[2]</sup>. The main idea of precision agriculture is that agricultural production can be adjusted according to local conditions. It can be reasonably invested, not wasted, and managed scientifically. Precision agriculture has achieved orderly management in space and time, and improved the efficiency of agricultural production. Through the use of modern agricultural machinery operation scale, improve the agricultural production efficiency and the efficiency of agricultural production, greatly changed the peasant workers will always work hard in the fields only in traditional agriculture, backward production mode and the "inspiration" extensive industry management<sup>[3]</sup>.

The productivity of the soil is mainly composed of soil fertility, soil fertility mainly refers to organic matter, nutrient content, soil texture, soil thickness and other factors in the soil, these factors have different effects on soil fertility, and have different constraints on soil fertility.

For the study of soil fertility classification, there are several situations that appear below: Zhang Yulong et al<sup>[4]</sup>. applied the mathematical method of principal component analysis and discriminant function discriminant to classify soil fertility. In addition to the mathematical model approach, quantification of soil fertility has also been studied, So far, however, quantitative indicators have not appeared, so that the size of soil fertility cannot be calculated. Liu Changmei and Liu Longchao<sup>[5]</sup> proposed that the soil per unit area under general conditions should be used to measure the amount of soil formed by plants to measure soil fertility. Because of the different types of soil use, the annual crop yield of the farming soil; the annual growth of timber for the forest soil; the grassland soil is measured by the annual grass growth, from an ecological point of view, soils of all types of use can be measured by solar energy (Joules), which is fixed by plants in the unit area throughout the year.

## 2 DBSCAN Algorithm

### 2.1 DBSCAN algorithm concepts

DBSCAN algorithm is a density based clustering analysis method. The algorithm defines clusters as the largest set of points connected by density, and divides the regions with high density into clusters. The kernel idea of clustering is to measure the density of the space of the point with the neighbor number in the neighborhood of a point <sup>[6]</sup>. For example, the object p is the center, and epsilon is the radius of the region, that is, within the epsilon neighborhood of p, including at least one positive integer Minpts objects, and p is the core object, fields containing Minpts objects are clusters, otherwise p is on the boundary of a cluster and is called boundary point <sup>[7-8]</sup>. The following explains the relevant definitions:

### 2.1.1 Directly density-reachable

If the p is the core point, q in the  $\epsilon$ -neighborhood of p, the p direct density of up to q.

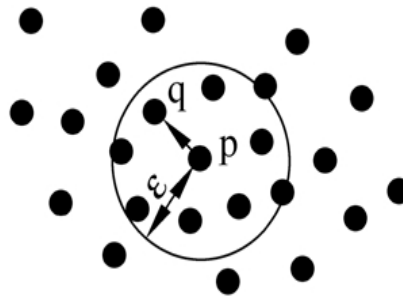


Figure 1. Directly density-reachable

### 2.1.2 Density-linked

If an object o exists in the object set D, which causes the object p and q to be reachable from the object o about epsilon and Minpts, then the object p and q are connected to epsilon and Minpts density <sup>[9]</sup>.

The definitions are parsed using Figure 2 below, In Figure 2, given the radius of the field epsilon, the minimum number is Minpts=3. According to the above definition, point m, o, p, and r core objects because there are at least 3 objects in their respective epsilon fields. In addition, it is also observed that m is directly accessible from p and vice versa. q is direct density reachable from m, but m is not directly reachable from q, because q is not the core object. The density from q to m is up to m, and the direct density of p from p can reach up to q so that the density is reachable. Similarly, p and s are density reachable from o, and o is density accessible from r.

As a result, o, r, and s are density-reachable between each other.

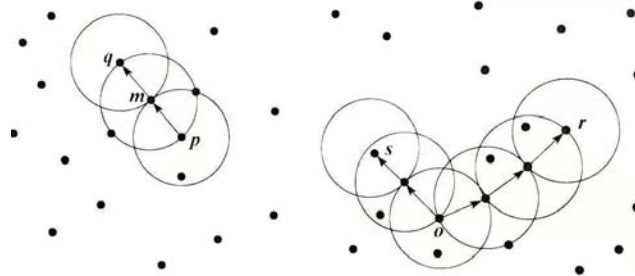


Figure 2. Density-reachable and density-linked

## 2.2 DBSCAN algorithm progress

Input:  $n$  data objects, radius epsilon, minimum number Minpts.

Output: all clusters that reach the density requirement.

Algorithm processing flow:

Step1 Extracts an unprocessed point from the data object;

Step2 IF points out is the core point, THEN finds all objects reachable from the point density, forming a cluster;

Step3 ELSE The point that you take is the edge point (non core object). Jump out of this loop and look for the next point;

Step4 Loop Step1 to Step3 until all points are processed<sup>[10]</sup>.

## 3 Maize precision fertilization

### 3.1 Data Acquisition

In Nong'an County of Jilin province Chen hometown pilot application, access to land information using the 3S technology, grid according to the distance of 40m\*40m. Soil sampling was carried out in the divided mesh, and 152 sampling points were obtained, which were collected for soil nutrient content, including four values of organic matter, available phosphorus, available nitrogen and available potassium.

### 3.2 Data Standardization

Since different data has different dimensions, it needs a unified dimension to compare the data, that is to standardize the data. normalization formula is:

$$g = \frac{G - \min}{\max - \min} \quad (1)$$

### 3.3 Soil Fertility Grading

The DBSCAN algorithm is used to cluster the processed data, that is, to classify the soil fertility, and each soil grade is called the cluster. In this paper, WAKA data mining tools are used to download DBSCAN data management package, set (field radius) 0.5, Minpts (epsilon field minimum point) is 10 of the cluster, the soil is divided into six cluster classes, that is, six levels.

According to the number of soil plots at each level, the four attributes of all soils in different levels are calculated, the average value is summed, and sorted according to the size of the sum, and the bigger one is ranked, and so on.

According to the sum of the average values of the attributes in the six clusters, the soil classification results are shown in table 1.

**Table 1.** Soil classification result

Soil level	Included sample points
Class A	a140 a83 a80 a3 a22 a54 a21 a131 a51 a90 a19 a20 a57
Class B	a40 a31 a11 a6 a149 a28 a314 a119 a64 a13 a151 a16 a70 a142 a100 a148 a109 a121 a47 a95 a96 a62 a116 a44
Class C	a39 a146 a10 a33 a38 a88 a7 a9 a147 a86 a130 a138 a14 a8 a139 a108 a144 a45
Class D	a35 a17 a145 a120 a23 a97 a71 a41 a143 a123
Class E	a1 a63 a133 a18 a89 a69 a84 a15 a50 a106 a60 a4
Class F	a55 a53 a81 a43 a59 a103 a105 a120 a78 a124 a75

### 3.4 Calculation of Fertilizer Application

A fertilizer calculation model was established, and the nutrient balance method was used to calculate the amount of fertilizer needed for the nutrient content in different soil levels. We calculate the average of four attributes in different soil levels and substitute them into the formula of nutrient balance method, the required values here are the raw values before data processing, and the data values are shown in table 2.

**Table 2.** Average properties of different soil classification

Soil level	Organic mean	Available phosphorus mean	available nitrogen mean	K mean
Class A	32.295	19.042	132.538	314.846
Class B	30.785	23.004	116.083	241.583
Class C	27.912	18.092	115.500	157.333
Class D	28.765	16.480	111.700	142.400

Class E	27.577	15.304	98.083	138.750
Class F	26.289	13.850	98.182	137.000

Calculation of nutrient balance fertilization model by fertilizer application rate (2):

$$sf = \frac{cl \times xs - cd \times ys}{hl \times ly} \quad (2)$$

sf: Fertilization; cl: Corn yield target;

xs: Grain corn nutrient absorption amount per 100kg;

cd: Soil nutrient determination;

ys: Soil available nutrient conversion factor;

hl: Fertilizer nutrient content;

ly: Fertilizer utilization season

Taking the mathematical model of fertilizer application rate as an example (The content of P<sub>2</sub>O<sub>5</sub> was 46% in the application of diammonium phosphate) :

$$\text{Phosphat fertilizer} = \frac{\text{Corn target amount} * 0.07 - 0.03 * \text{Soil nutrient content} * \text{Soil available nutrient conversion factor}}{0.46 * \text{Fertilizer utilization rate}}$$

$$\text{Soil available nutrient conversion factor} = \begin{cases} \frac{1578.8 * \text{Soil nutrient content}^{0.98}}{100} \\ \frac{1068 * \text{Soil nutrient content}^{-0.832}}{100} \\ \frac{732 * \text{Soil nutrient content}^{0.749}}{100} \end{cases}$$

$$\text{Blank area yield} = \frac{0.3 * \text{Soil nutrient content} * \text{Soil available nutrient conversion factor}}{0.022}$$

$$\text{Fertilizer utilization rate} = \begin{cases} \frac{(43.4 - 0.024) * \text{Blank area yield}}{100} \\ \frac{(36.6 - 0.025) * \text{Blank area yield}}{100} \\ \frac{(4.6 - 0.035) * \text{Blank area yield}}{100} \end{cases}$$

Yield=10000 kg/hm<sup>2</sup>

According to the nutrient balance method, the data of Table 2 are replaced by (2), and the specific amount of soil fertility at each level of soil as shown in Table 3 can be obtained.

**Table 3.** Soil nutrient fertilizer rate

Soil level	N amount of fertilizer(kg/hm <sup>2</sup> )	P amount of fertilizer(kg/hm <sup>2</sup> )	K amount of fertilizer(kg/hm <sup>2</sup> )
Class A	440.063	565.009	327.032
Class B	440.158	564.844	327.171
Class C	440.161	565.053	327.398
Class D	440.185	565.135	327.451
Class E	441.160	565.200	327.465
Class F	440.276	565.287	327.471

## 4 Conclusions

In this paper, the DBSCAN algorithm is used to classify the soil, and the classification results are applied to the corn precision fertilization decision. The experimental results of demonstration and application in Nong'an County of Jilin province Chen hometown. The average amount of fertilizer is 560 kg/hm<sup>2</sup> (the average amount of fertilizer is calculated by the total amount of fertilizer divided by the total area), compared with the traditional fertilizer 608.42kg/hm<sup>2</sup>, saving fertilizer 48.42kg/hm<sup>2</sup>; The average yield of the experiment is 8313kg/hm<sup>2</sup>, which is 930kg/hm<sup>2</sup> higher than that of the traditional output (the traditional values are derived from the statistical yearbook of China). Indeed, the purpose of reducing chemical fertilizer input, improving soil environment, increasing production and increasing income has been achieved.



## Acknowledgment

The study was conducted by 2016 jilin province rural special project supported by the modern agricultural development 《Demonstration and Application of Traceable System of Quinoa Products Based on Internet of Things and 3S Technology》,Jilin science and technology development plan project, 《Research on precision control and control technology for high quality and high efficiency production of major grain crops》 (20170204020NY).

## References

1. Miao Yubin. Research and application of intelligent agricultural production measurement system[D]. Shang Hai, Shanghai Jiao Tong University(2003)
2. Chen Guifen, Yu Helong, Cao LiyingMa Li. Data mining and precision agriculture intelligent decision system[M].Beijing, Science Press(2011)
3. National Research Council. Precision Agriculture in the 21st Century, Geospatial and Information Technologies in Crop Management. Washington : National Academic Press(1997)
4. Zhang Yulong. Preliminary study on soil fertility classification using numerical analysis method[J].Journal of Shenyang Agricultural University,17(3), 49-55(1986)
5. Liu Changmei, Liu Longchao. Study on quantitative classification of soil fertility[J]. guizhou agricultural sciences, 27(4), 20-22(1999)
6. Zeng Yiling, Xu Hongbo, Bai Shuo. Improved OPTICS algorithm and its application in text clustering[J]. Chinese Journal of information,22(1), 22-24(2008)
7. Wu Xiujie, Zhang Quanchao, Li Haijun. The value of clustering analysis and principal component analysis in the study of Anthropology[J].Journal of Anthropology, 29(4), 35-37(2007)
8. Wu Shanjie. Further thinking on classification method of fuzzy clustering[J].Journal of North China Institute of Science and Technology, 5(1), 108-110(2008)
9. Ester M., Kriegel H.P., Sander J., Xu X.A density-based algorithm for discovering clusters in large spatial database[C]. Knowledge Discovery and Data Mining(KDD'96), 226-231(1996)
10. Ester, Martin, Hans Peter Kriegel, et al. A Density Based Algorithm for Discovering Clusters in Large Spatial Data-bases with Noise[C].Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining(KDD-96). Ortland. Oregon(1996)