Visual Space Research and Application of the Data Mining in Soil Fertility Evaluation

Hang Chen, Guifen Chen

Visual space research and application of the data mining in soil fertility evaluation

Hang Chen,Guifen Chen*

(College of Information and Technology, Jilin Agricultural University, Changchun, 130118, China)

Author: Hang Chen .tel: 0431-84532775, Email:chenhang0811@163.com

Corresponding author: Guifen Chen. Tel:0431-84532775,Email: guifchen@163.com

Abstract. China is a populous agricultural country，the arable land is the most basic resources which is needed for agricultural production, the quality of arable land has a direct impact on the entire agricultural production and development ,so to carry out the Evaluation of Farmland is a very important foundation work which is advantage to know china's arable land resources and to improve the efficiency of use for arable land and promote the development of modern agricultural , and which is helpful to further identify china fertility status of cultivated soils, the distribution of nutrients, soil fertility and agricultural land degradation, etc., so as to achieve the realization of agricultural modernization and to improve the comprehensive capacity of agricultural production.This paper use visualization of spatial data mining technology which is used to research and apply on soil fertility evaluation, which is put the technology of spatial data mining and visualization technology, GIS, technology tightly integrated, that makes a research to the elements of soil fertility, yield and other data mining, and achieve the application of visualization of spatial data mining techniques in the evaluation of soil fertility, this application is more conducive to promotion and application of precision agriculture for variable rate fertilization, variable spraying, at the same time it provides real-time, accurate decision making basis for agricultural production

management department, to promote the transformation of traditional agriculture to modern agriculture, which has important theoretical significance and practical value.

Keywords：Soil fertility evaluation; GIS; spatial data mining;

1 Introduction

In resent years，with the development of visualization techniques and data mining technology, in order to solve the visible and interactive problem of the data mining process, under the combination of the two information visualization techniques and data mining technology to form a visual data mining ,Visual data mining technology make full use of the advantages of data mining and visualization technology to solve the problem of the "black box" data mining operation, and gradually being valued by the people,   which is an important aspect of data mining research and has a very high research value.to carry out the evaluation of cultivated land which is use spatial data mining technology and agricultural technology can reduce the manpower, material and financial resources of arable land evaluation work which is required，and it can improve the efficiency of the evaluation, but also can improve the accuracy of the evaluation results, and conducive to the evaluation results summary, management, updates, cartographic output and promote the application. So as to identify the production capacity of China's cultivated land base, soil fertility status, to achieve the purpose of a strong impetus to the modernization of agriculture and sustainable development.

2. Soil spatial data acquisition

In support of GIS and related technologies, we can quickly obtain a series of basic spatial information about soil and its influencing factors, such as topography, land use, crops, soil fertility, etc.. These information is the basis of the analysis and evaluation of soil space. This paper uses global positioning system (GPS) and3G wireless sensor technology for data acquisition.

2.1 Global Positioning System (GPS)

Global Positioning System (GPS) for information acquisition and implementation of accurate positioning. Widely used in order to improve the accuracy DGPS (Differential Global Positioning System) technology, the so-called "differential correction global satellite positioning technology." In order to improve the accuracy,, widely used DGPS (Differential Global Positioning System) technology, which is called" the differential correction of global satellite positioning technology". It is characterized by high positioning accuracy, the accuracy of the GPS system are free to choose according to different purposes.

2.2 3G wireless sensor

3G wireless sensor can be used to collect information on soil plots and be used to environmental monitoring on soil topography, crop of soil, and soil fertility, etc.. The user can monitor, process the data and show the data chart on soil environmental through INTERNET and 3G networks, to provide users with efficient and stable environment parameters, while addressing the defects of traditional long-distance wireless transmission barriers, and the true realization of remote soil environmental monitoring.

3. GIS spatial analysis

Spatial analysis is one of the important functions of the GIS system is based on the geographic location of the object and form of spatial data analysis techniques, with the aim of the extraction and transfer of spatial information. The object of spatial analysis is a series of spatial location data, which include two spatial coordinates and professional attributes.

3.1 Overlay analysis

Overlay analysis in GIS is to refer to the same region, the unified scale, two sets of data files or polygon feature more than two groups superimposed by the simulation of multiple properties in the region, to find and identify multiple attributes polygon or

statistical analysis of the property characteristics within the scope of the polygon. In this paper, which the district overlay analysis of the District to be about the level of thematic maps superimposed to produce a new data aspects of the operation, the results of a combination of the original two-tier or multi-layer elements with attributes.

3.2 The interpolation of spatial data

Spatial data according to their own requirements for sampling point observations, such as land type, land nutrient sampling points, ground elevation, the distribution of these points are often irregular, discontinuous, it is impossible to fill the entire study area. Thus formed by the polygon internal changes can not be more accurate, more specific, but can only reach the average level expression, However, at some point, the user might like to know the exact value of the observation point certain characteristics of interest, which resulted in the spatial interpolation technology. The purpose of the interpolation is according to the known point attribute reasonable deduction and predict the position near the attribute values. The main methods: inverse distance difference, spline interpolation and kriging. In this paper, the Kriging interpolation method to establish the soil of nutrients, moisture and crop yields three-dimensional spatial variation diagram.The conditions of application of the Kriging interpolation method is regionalized variable spatial correlation.
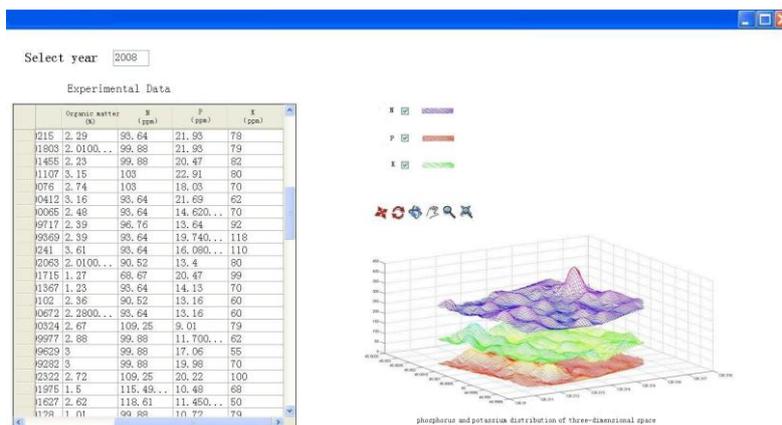

Fig. 1. phosphorus and potassium distribution of three-dimensional space

4. The Integrated application of GIS and technology of spatial Data Mining in agriculture

Spatial data mining is a process of extracting general knowledge, laws and rules from the spatial database or Spatial Data Warehouse. The knowledge, laws and rules are implied in them, unclear in advance, potentially useful and spatial or non-spatial which ultimately easy to understand[3]. The purpose of spatial data mining is that help people to improve the accuracy and reliability of the decision-making, and enable people to maximize the efficient use of data.

4.1 The classification of spatial data mining technology with the support of GIS

The spatial data mining technology can be divided to describe, explain and predict by function. Descriptive model makes the spatial distribution characterized, such as spatial clustering. Explanatory model deals with spatial relationships, such as the relationship between a space object and the factors affecting their spatial distribution.

4.2 The evaluation of soil fertility based on spatial weighted fuzzy clustering algorithm

4.2.1 The general situation of test area

Using GPS, GIS, RS, and sensor technology, access information of corn farmland from experimental ground, and then divided the ground into grid cell (specifications size is 40m*40m), A1~L10 are sampling points. To sample in this grid cell, the sampling depth is 25 cm and the sampling method is the five points plum sampling, which the soil samples of the four corners of the mesh and grid center were mixed as the grid soil samples. After sampling, select 63 sampling points as the sample data, record the attribute values of spatial coordinates in the soil, organic matter, available nitrogen, available phosphorus and available potassium.

4.2.2 Algorithm application

First of all, according to the analytical data of the soil nutrients from pilot area, soil nutrient spatial variability law analysis. Analysis results show that, the available phosphorus in soil spatial variability, the coefficient of variation was 31.12%,

available potassium spatial variation coefficient is 21.51%, organic matter spatial variation coefficient is 29.31%, available nitrogen spatial variability coefficient of 11.69%. Be evaluated in accordance with the results and the local characteristics of the soil and construct paired comparison matrix B.

$$B = \begin{bmatrix} 1 & 1.333 & 0.4 & 2 \\ 0.75 & 1 & 0.3 & 1.5 \\ 2.5 & 3.333 & 1 & 5 \\ 0.5 & 0.667 & 0.2 & 1 \end{bmatrix}$$

Thus obtain nutrients weight vector $A = \{0.2105, 0.1579, 0.5263, 0.1053\}$.

The fuzzy equivalent matrix $t(R)$ is calculated from original data matrix D using calculate weighted, and then adjust the value of λ by decreasing to establish the dynamic diagram of fuzzy clustering.

when values of λ are different, calculating the Quantity statistics of F-, Results such as shown in table 1.

Table 1. F-value calculations

Num.	λ	Class number	Value of F-
1	1	100	0.339
2	0.997	92	0.652
3	0.996	86	1.305
4	0.995	71	1.371
5	0.994	48	1.469
6	0.993	40	1.543
7	0.992	32	2.304
8	0.992	28	3.445
9	0.990	23	4.594
10	0.987	15	4.875
11	0.986	11	4.898
12	0.984	10	4.787
13	0.983	5	2.069
14	0.981	4	1.121

Compared F with $F_{0.05}(r-1, n-r)$, then we know when $F > F_{0.05}(r-1, n-r)$,the values of λ are 0.981, 0.983, 0.984, 0.986, 0.987, 0.990, 0.991, 0.992, 0.993, 0.994, 0.995，calculating $F - F_{0.05}(r-1, n-r)$,the bigger value of λ is 0.986, From the table also can be seen the class number is 11. Total divided into 11 classes. Shown in Table 2.

Table 2. Classification table

category of Classes	Number of sample points	higher of the soil Nutrient Content	Lower of the soil Nutrient Content	The soil nutrient tend to average	Note
category 1	43				The soil nutrient is average
category 2	7	Available N	Available P	organic matter, Available K	
category 3	1	Available K	Available N, Available P	organic matter	
category 4	1	Available K, Available N	Available P	organic matter	
category 5	4		Available P	Available K, organic matter Available N	
category 6	2	Available N	Available P	Available K, organic matter	
category 7	1			Available P, Available N,	

				Available K, organic matter	
category 8	1	Available N	Available P		below-average
category 9	1	organic matter			below-average
category 10	1	Available K	Available P		Tend to average
category 11	1				Artificial add noise

The table 2 shows that space fuzzy clustering results based on weighting than unweighted space fuzzy clustering results are obviously different. Category 1 , Category 2, Category 3, Category 4, Category 5 and Category 6 are divided into Category 1 in the space fuzzy clustering results of traditional. Available P tends to average in the soil, and available N, available K, organic matter are higher than average by analysis. The kinds of nutrient content in the soil tend to average what are divided into category 1 after weight, however the space fuzzy clustering results of traditional in the category 1 is divided into category 2 to category 6 under the influence of available P which has the Maximum Weight. As the same time, we find that adjacent grids are adjacent; it is mean that spatial continuity is in soil nutrient content, above characteristics benefit the precise task of fertilizer applicator and convenient to soil sampling and field management.

Calculation of the factor of each index weight is by AHP. And according to the reality, making   a comprehensive evaluation to organic matter in the soil , nutrition of available N, available P and available k and reflecting the fact is in soil nutrition.

5. Implementation of Visual spatial data mining in soil fertility evaluation

Under Visual spatial data mining support to achieve automation and quantitative evaluation arable land in fertility by mathematics method and mathematical model of systematic clustering method, analytic hierarchy process, fuzzy evaluation , to get the figure of Fertility of arable land and level distribution, to analysis limiting factor of

each level of cultivated land. fertility evaluation of the farmland、the cultivated land fertility dynamic analysis and farmland improvement use application of the feasibility and the scientific nature to provide feasible technology for accurate in the cultivated land fertility evaluation dynamic analysis and improvement, and it has the practical application value of scientific management and sustainable utilization of the cultivated land resource.

5.1 Comprehensive evaluation index system of the cultivated land fertility

5.1.1 The choice of evaluation index system
According to the national general 55 index frameworks, to select climate , terrain parts, Ground slope, soil thickness, into soil organic, section configuration, soil erosion, top layer texture, farmland infrastructure, the soil physical and chemical properties, traffic and economic development level and the level of production, etc. According to the selection rules of which has great influence on cultivated land, has a larger variation in YuShu city demonstration area and which is stable in time series, to select the profile character, the site conditions, physical and chemical characteristics, nutrient, soil management five types and 14 indexes as evaluation factors, which reflects objectively YuShu city   productivity grade difference and has the very strong objectivity.

5.1.2 Sample library established
Based on the Second soil survey results, Representative face of Average each point in the center of Northeast China is 333.33～666.67 hm2. Cultivated land area in Yushu city 331100 hm2, the glebe is about 250000 hm2, Paddy fields is about 80000 hum2, and other is Awning vegetables about 1100 hm2. In YuShu city, the amount of the test-bed for sample points is determined and is the control number, superposition of painted for   figure of land used status and   the compiling of soil map to confirm the final sample points for 574.

5.2   realized of visual spatial data mining on soil fertility evaluatio

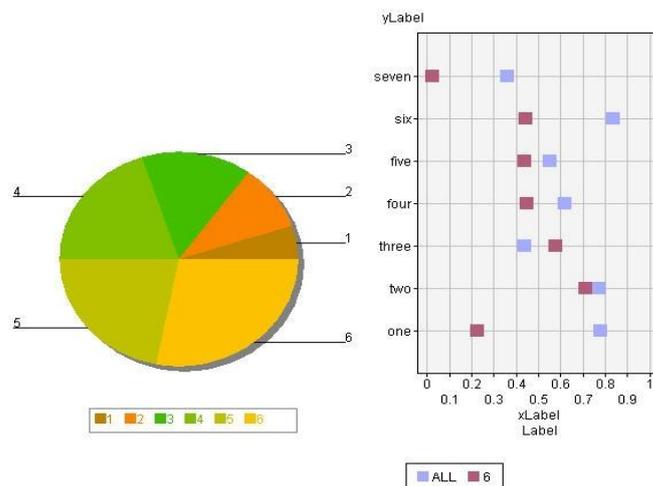| Soil classification	Integrated fertility index(IFI)	Yield(kg/667m2)	Cultivated land area(Million hm2)	Percentage to total area(%)
Grade 1	>0.85	>600	4.14	12.50
Grade 2	0.81-0.85	551-600	11.25	33.99
Grade 3	0.76-0.80	501-550	8.37	25.29
Grade 4	0.71-0.75	451-500	4.17	12.60
Grade 5	0.65-0.70	400-450	3.24	9.79
Grade 6	<0.65	<400	1.94	5.83

Fig.2. results visualization of data mining

From the figure we can see that the soil of 1~2 level is mainly distributed in the ridge of hilly platform, low terraces of the valley plain and high flood area and placer, diluvia plain, which the terrain is flat and basically no erosion or mild erosion; The land of 3~4 level most located in undulating terraces or the slope of alluvial flood plain, in undulating terraces or alluvial flood plain slope, in which the terrain is gentle, mild and moderate erosion; The 5~6 level most in low mountains and hills, the top of Undulating tableland and low river floodplain, which terrain is hilly or steep slope greater than 7°. Soil erosion is generally moderate to severe or the low-lying, easy to form a flood.

6. Conclusion

In this paper，based on spatial data mining, site conditions, soil management, soil nutrient ,profile traits and physical and chemical properties, the five factors as the sample input. Using AHP hierarchical analysis the cultivated land fertility soil comprehensive evaluation, with a strong ability of the evaluation and expansion ability. The research results of the cultivated land fertility soil comprehensive evaluation and soil the cultivated land fertility management, provides the theoretical basis and technology. However, the analytic hierarchy process applied to the soil cultivated land fertility is still a new field, its theoretical foundation to be further improved, such as the classification error of the sample data, and the type of kernel function and parameter selection, how to simplify the discriminate formula to reduce the computational volume and improve the discrimination efficiency. The main conclusions are as follows:

①According to the analysis of the spatial clustering of the actual needs of the status and soil fertility decisions, the use of fuzzy clustering algorithm based on the weighted space of soil fertility evaluation, propose a new method for the accurate evaluation of soil fertility.

②Use of spatial analysis methods and GIS in the statistical analysis methods and data mining visualization combined. Conducted in-depth analysis of the various factors that affect soil fertility, and evaluated, to establish an evaluation system to improve the administrative departments for agriculture take full advantage of the ability of the various properties and spatial information data to make decisions.

③Spatial data mining technology has penetrated the field of agriculture, but because of the complexity and variability of the agriculture itself, making the spatial data mining techniques in dealing with these large, multi-dimensional, dynamic, incomplete, uncertainties and other characteristics of data, data processing and mining tools need to further research and exploration.

So, the diversity of the data entry forms based on the uncertainty and ambiguity of spatial data mining, data warehousing and spatio-temporal data mining, knowledge of the visual representation of user participation and intelligent data mining and other future agricultural areas of application of data mining technology development trends and hot spots.

References

1. Zhang S H. Based on GPS, GIS for precision agriculture variable rate fertilization theory and technology research[D]. Jilin: Jilin University,2003: 1-3. (in Chinese)

2. Miao X K, Xia K J, Wang X. Precision farming variable rate fertilization intelligent decision support system research. Computer Applications, 2004, 24(11): 153-155. (in Chinese)

3. Li Y, Shi Z, Wu C F, Li F, Cheng J L. Definition of management zones based on fuzzy clustering analysis in coastal saline land. Scientia Agricultura Sinica, 2007, 40(1): 114-122. (in Chinese)

4. Chen G F, Cao L Y, Wang G W, Wan B C, Liu D Y, Wang S S .Application of spatial fuzzy clustering algorithm in precision fertilization .New Zealand Journal of Agricultural Research, 2007, 50: 1249-1254.

5. Zhang S H, Ma C L, Li W, Xu Y. Experimental study on the influence of variable rate fertilization on maize yield and soil nutrients. Transactions of the Chinese Society of Agricultural Engineering, 2006,22(8): 64-67. (in Chinese)

6. Chen L P. Precision agriculture variable rate fertilization theoretical and experimental research[D]. Beijing: China Agricultural University,2003: 98-104. (in Chinese)

7. Graham Sparling .Louis Schipper .Soil quality monitoring in New Zealand: trends and issues arising from a broad-scale survey, Agriculture, Ecosystem andEnvironment,2004,104:545~552

8. Turner BL, Meyer WB. Land use and land cover in global environmental change: considerations for study [J]. Int.SoiSci.,J.,1991,130:669-680.