The Soil Nutrient Spatial Interpolation Algorithm Based on KNN and IDW

Xin Xu, Hua Yu, Guang Zheng, Hao Zhang, Lei Xi

The Soil Nutrient Spatial Interpolation Algorithm Based on KNN and IDW

Xin XU[1,3], Hua YU[1,3], guang ZHENG[1,3], Hao ZHANG[1,2,3], Lei XI[1,2,3,*]

[1]College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, China;

[2]Henan grain crops Collaborative Innovation Center, Zhengzhou 450002, China;

[3]Ministry of Agriculture Agricultural Information Technology Science Observation Station in Huanghuaihai district, Zhengzhou 450002, China

Abstract. For breaking the limitation of the GIS platform and realizing the soil nutrients spatial interpolation algorithm for any points in the monitoring area to transplant to the mobile platforms, this paper established the spatial index of the soil nutrient sampling points utilizing the K-D Tree as the space splitting algorithm of the soil nutrient sampling points. On this basis, the K nearest neighbor search of the soil nutrient sampling points was also implemented employing KNN algorithm. Finally, the soil nutrient spatial interpolation was realized based on KNN and IDW algorithm. Meanwhile, the accuracy of the algorithm and the influence to the different soil nutrient elements affected by the K value in KNN algorithm were also verified. The results show that the soil nutrient spatial interpolation algorithm was viable to predict the element contents of soil nutrient during the running time was less than 3s. To reach the best accuracy, the values of the proximal point K for predicting the PH, organic matter, rapid available phosphorus and rapid available potassium should be set as 85, 15, the largest sample space and 65 respectively. The optimal average absolute error of the pH, organic matter, rapid available phosphorus and rapid available potassium was 0.0405, 0.3870, 0.0015 respectively.

Keywords: IDW, KNN, KD-Tree, Soil Nutrient, Spatial Interpolation

1   Introduction

Due to the problems of the obtaining the sampling points data of the soil testing and formulated fertilization such as difficult sampling[1], heavy workload[2], high cost[3] and so on, the data points during the fertilizing decision stage in the process of testing soil for formulated fertilization was not enough to achieve the complete coverage for the field parcel in each region. Therefore, the sampling point soil nutrient data of the currently unknown spatial points were obtained on the basis of interpolation operation of the existing sampling point data frequently in real application. The utilizing of spatial interpolation technology was commonly used method. In the method, the statistical methods were applied to the nutrient data of some smaller density soil sample points to conduct interpolation operation to the data from the points that were not sampled to form the more dense point data distribution or the area data of

different areal unit. The soil nutrient values of the points which were not sampled were predicted scientifically. Then, the precise fertilization could be implemented to the whole area.

Following the rapid development of the mobile intelligent terminal and the distributed computing technology, the computing and data processing of more and more applications was transferred to client-side错误!未找到引用源。. Due to the frequent disconnection[5] and low reliability[6] of mobile GIS, most of space interpolation for mobile side had to rely on the service side in the using. It made the usage scenarios of the mobile-oriented space interpolation limited. Thus, making full advantage of modern computer technology could provide more configurable, extensible and customizable distributed applications and made traditional soil testing and formulated fertilization recommendation system running in embedded devices such as high performance and low cost mobile intelligent terminal possible错误!未找到引用源。 7,8]. However, most of current soil testing and formulated fertilization system错误!未找到引用源。9错误!未找到引用源。 and soil nutrient spatial interpolation technology错误!未找到引用源。 2-错误!未找到引用源。 were based on the GIS platform and hard to be transplanted to mobile platform. Hence, to realize the soil nutrient inquiry to any land parcel and the fertilization information recommendation by the users through intelligent mobile terminals and the personalized recommendation of the soil testing and fertilizer recommendation technology by fully integrating mobile technology and soil testing formula technology, the researching of a type of soil nutrient spatial interpolation technology which was divorced from the limitation of the GIS platform was significant for the application and popularization of soil testing formula fertilization technology.

2   Material and Methods

2.1   Source of Material

The data in this paper came from the 238 soil nutrient sampling points in 10 thousand mu precision agriculture model production field of Changge city, Henan provice in 2014. The sampling point data was consist of sampling number, numbering of land parcel, soil texture, preceding crop, PH values, organic matter, rapidly available phosphorus, rapidly available potassium, latitude and longitude information.

2.2 Research Method

2.2.1 Inverse Distance Weighted (IDW) Interpolation

At present, there were many kinds of methods for the soil nutrient interpolation. In which methods, the Inverse Distance Weighted interpolation (IDW) and Kriging interpolation错误!未找到引用源。 15错误!未找到引用源。 were considered to be the most widely used two common interpolation methods. The Kriging interpolation was often influenced

by many factors in actual application[错误!未找到引用源。18错误!未找到引用源。]. Considering the space complexity of Kriging interpolation algorithm, the Inverse Distance Weighted interpolation was adopted in this study to design the interpolation algorithm for the soil nutrient data.

The inverse distance weighted interpolation was based on the theory of "like similarity principle". Through the calculating of the inverse distance weighted average of the discrete points of nearby area, the value in the cell was also calculated. According to the principle that the point with closer distance had the greater weighting values, the value of the estimate point was fitted employing the linear weighting of some neighboring points around.

This algorithm was a simple and effective data interpolation method with relatively fast computing speed. However, it also had the obvious flaws. For example, the method only considered the space distance between the estimation points and the interpolation points. Moreover, the weight calculation used in the algorithm lacked the exact physical basis[错误!未找到引用源。 20].

2.2.2 K-Nearest Neighbor (KNN) Algorithm and Spatial Index of Soil Nutrient Sampling Points Based on K-D Tree

The soil nutrient content of the sampling point was not only affected by the change of the sample point distance, but also was affected by many other factors such as soil texture variation. Therefore, theK-Nearest Neighbor algorithm was also introduced in the process of the adjacent points selecting and searching during the process of the interpolation to the soil nutrient using inverse distance weighted interpolation to improve the IDW algorithm.
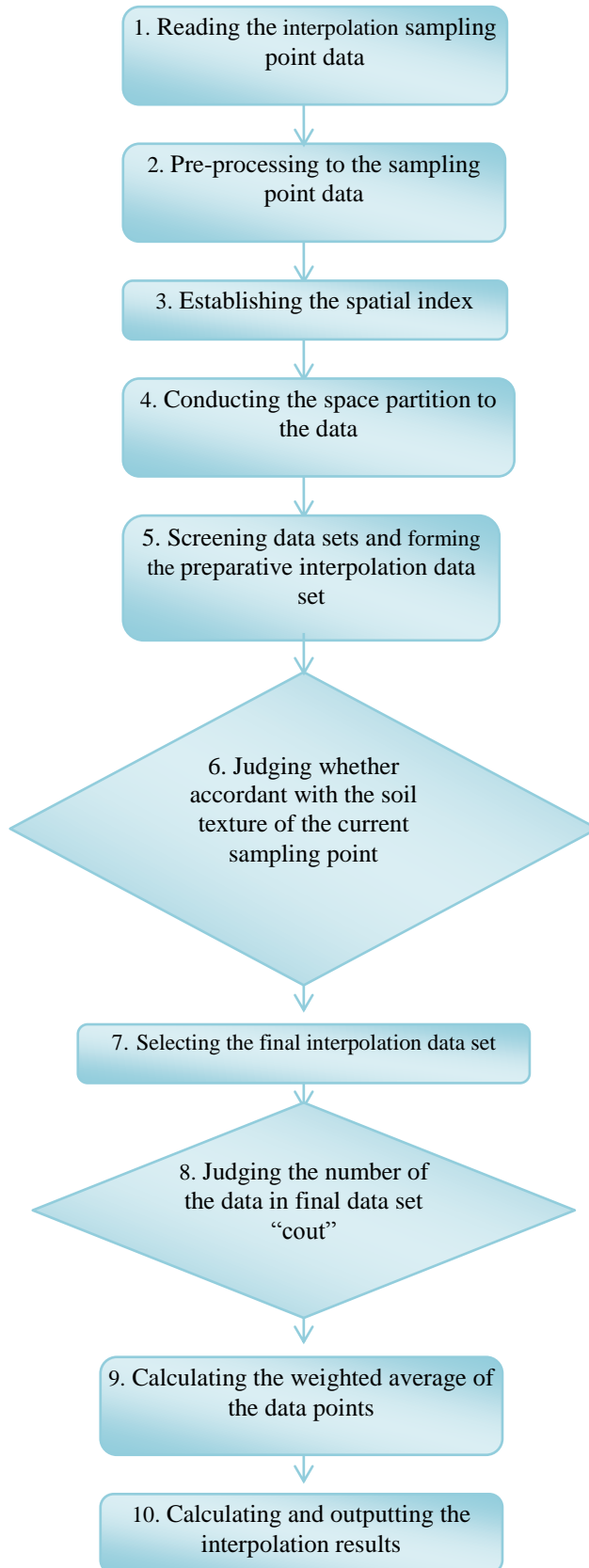
K-Nearest Neighbor (KNN) sorting algorithm was proposed by Cover and Hart in 1967[错误!未找到引用源。]. It was a simple algorithm based on analogy and had become one of the most mature machine learning algorithms in theory at present. The idea of the algorithm was: if most of the k most similar (i.e., the adjacent) samples of a sample in the characteristic space belonged to a category, then the sample also belonged to this category. If the query point and the positive integer K were given, the k closest data to the query point were also found from the data set.

The realization of the KNN algorithm was based on the establishment of the spatial index for data. The commonly used algorithms for establishment of a spatial index in GIS interpolation algorithm has K-D tree, K-D-B tree, BSP tree, R tree series, quadtree, grid and many other spatial index[21]. K-D (K dimension search binary tree) was a main memory data structure to generalize binary tree to multidimensional data and a type of binary tree in K dimension space. It was also a type of dynamic index structure for the segmentation of data space and suited for the space point target indexing[22]. Therefore, this paper constructed the spatial index algorithm of the soil nutrient utilizing the K-D tree.

3 Design and Simulation of Algorithm

3.1 Workflow Design

The idea of the soil nutrient spatial interpolation algorithm based on KNN and IDW proposed in this paper was: Constructing the KD-Tree firstly to establish the index space of the soil nutrient data sampling points. After that, the sampling point data set was further screened. The specific method was: using the given soil nutrient training data set, the tree structure data index was constructed adopting the space partitioning tree KD tree to conduct the hierarchical division for the search space of soil sampling point. After that, the fast matching was conducted to the new input sampling point data to find the nearest neighboring K sampling points of the sampling point in the training data set. The K sampling point data would be used as the spatial interpolation data. The algorithm process was shown in Figure 1.

```
┌─────────────────────────────┐
│ 1. Reading the interpolation │
│       sampling              │
│       point data            │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ 2. Pre-processing to the     │
│       sampling              │
│       point data            │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ 3. Establishing the spatial  │
│       index                 │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ 4. Conducting the space      │
│       partition to          │
│       the data              │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ 5. Screening data sets and   │
│   forming the preparative    │
│   interpolation data set     │
└─────────────────────────────┘
              │
              ▼
        ◇ 6. Judging whether
        accordant with the soil
        texture of the current
          sampling point ◇
              │
              ▼
┌─────────────────────────────┐
│ 7. Selecting the final       │
│    interpolation data set    │
└─────────────────────────────┘
              │
              ▼
        ◇ 8. Judging the number
           of the data in final
             data set "cout" ◇
              │
              ▼
┌─────────────────────────────┐
│ 9. Calculating the weighted  │
│    average of the data points│
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ 10. Calculating and          │
│     outputting the           │
│     interpolation results    │
└─────────────────────────────┘
```

Fig. 1. The process design of algorithm

(1) Reading the current sampling point data set and ready for operation;

(2) Conducting the data pre-processing to the sampling point data and removing critical value such as the outliers and the value beyond administrative boundary and so on to reduce the measuring error of the interpolation results;

(3) Constructing the KD tree and establishing the spatial index of the sampling point data;

(4) Conducting the space partition to the soil nutrient sampling point data using the KD tree;

(5) Conducting the screening to current data set utilizing the K-Nearest Neighbor algorithm to seek the nearest neighbor sampling point data of the current interpolation sampling point data to form the preparative interpolation data set;

(6) Traversing the current sample point data set, judging the data whether was accordant with the soil texture of the current sampling point. If the data was inconsonant with the soil texture of the current sampling point, the data would be discarded.

(7) Selecting the final interpolation data set based on the step 6;

(8) Judging whether the number of the data in final data set "cout" was less than 2, if "cout" was less than 2, this time interpolation was fail and returned the "false" information. Otherwise, the process continued.

(9) Looping through the data set and calculating the weighted average of the current unknown sample points with the current data points.

(10) Calculating and outputting the interpolation results based on inverse distance weighted interpolation.

3.2 Algorithm Implementation

Based on the above algorithm design, the process of the algorithm implementation consisted of the following three steps: 1. KD tree Constructing, 2. K neighboring searching, 3. the inverse distance weighted interpolation calculation.

1. KD tree Constructing

The KD tree data structure was established firstly in the algorithm. After that, the KD tree was constructed based on KD tree data structure. The steps of the algorithm were as follows:

(1) If the sampling point data sets "Dataset" was empty, the empty KD tree was returned. Otherwise, the node generation program would be executed.

(2) Determine the maximum variance dimension the "Split" domain. The maximum variances in each space dimension of all the sampling point data which prepared for interpolation were counted. The maximal dimension in their variance was elected as the space division dimension. The quantitative value was used as the value of the "Split" domain.

(3) The root Node "Node" was confirmed. The data of the selected space partition dimension was sorted. The sampling point data located in the middle was used as the data of the "Node" node.

(4) The left subspace "leftNodeData" and the right subspace "rightNodeData" of the "Node" were determined. The data of the sampling points whose space dimension

was less than the "Node" node was used as the data of the left subspace of the "Node" node. The data of the sampling points whose space dimension was larger than the "Node" node was used as the data of the right subspace of the "Node" node.

(5) The above process was recursive to conduct the space division partition for the "leftNodeData" and the "rightNodeData" until the subspace had not data.

2. K-Nearest Neighbor Searching

The KD tree nearest neighbor search query algorithm was adopted in this study to construct the KD tree "Kd". Adopting the target point "target" as the input, employing the obtained K nearest neighboring "nearest" sets of the "target" as the output, the steps of the K search algorithm establishment were as follows:

(1) The leaf nodes containing the target point "target" were found from the KD tree. From the root node, the KD tree was recursively searched downward. If the coordinates of the current dimension of the target point "target" was less than the coordinate of the split points, the target point "target" was moved to the left child node space. Otherwise, it was moved to the right child node.

(2) The leaf node searched in step 1 was used as the current nearest point "currnearest".

(3) The other possible nearest points were backtracked recursively. If there was a node closer the target point "target" than the current nearest point "currnearest", the value of the current "currnearest" was updated.

(4) When the backtracking backed to the root node, the search ended. The last nearest point "currnearest" was the nearest neighboring point of the "target".

(5) The current nearest neighboring point was deposited in the "nearest" set and removed from the tree of Kd.

(6) The above process was repeated until the number of set elements in the "nearest" was equal to K. Stopped the search and returned back to the "nearest" set.

3. Inverse Distance Weighted Interpolation

Based on the KD tree generating algorithm, the spatial index tree of the current sample sampling points was constructed. On the basis of this, the K nearest-neighbor sampling points were searched out combining k-nearest neighbor algorithm to form the nearest-neighbor set "nearest". Employing the nearest-neighbor set "nearest" and the target point "target" as the input, the sampling points data of the current sampling points was conducted the spatial interpolation processing adopting the inverse distance weighted interpolation. The algorithm steps were as follows:

(1) Calculating the distance between the target point "target" and the sampling points in the nearest-neighbor set "nearest". Traversal the "nearest" set and calculating the distance from the "target" to each nearest-neighbor sampling point.
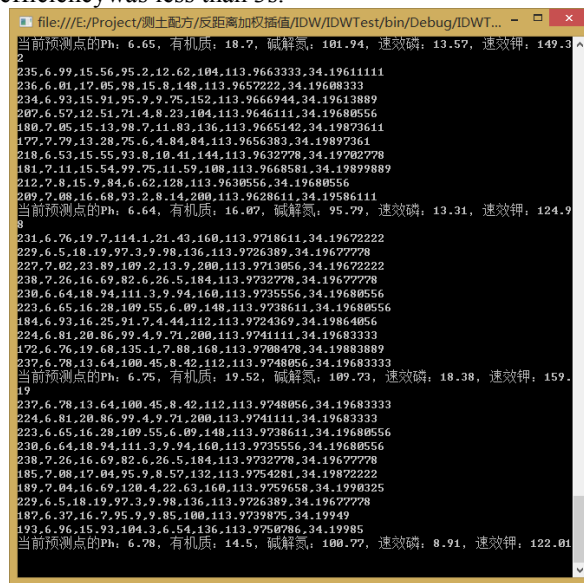
(2) Calculating the weight of each point. The weight of each sampling point in the nearest-neighbor set "nearest" was calculated to be used as the weight during the interpolation.

(3) Calculating the results of the weighted interpolation. Traversal the "nearest" neighboring set.   The results was worked out based on the weight of each point and returned.

4 Design and Analysis of Experiments

Adopting the data from 238 soil nutrient sampling points in 10 thousand mu precision agriculture model production field of Changge city, Henan provice in 2014 as the data source, the ARM CPU architecture as mobile hardware environment(1.6 GHz, 4 core), the above soil nutrient interpolation calculation method was realized based on the Android 4.4 mobile platform. As shown in figure 2, the pH, organic material, rapid available phosphorus and rapid available potassium in the sample data were conducted the spatial interpolation processing and effect analysis during the algorithm implementation efficiencywas less than 3s.

Fig. 2. Algorithm running results

The Figure 2 showed the results page after interpolation to the 10 nearest-neighbor sampling point data. By the program, the nutrient content of the current point was predicted through the interpolation to the pH, organic matter, rapid available phosphorus and rapid available potassium of the current point and contrasted with the actual value.

The effect analysis of algorithm adopting the cross validation method: a small number of samples of the sampling data of nutrients were screened and reserved to be not involved in the interpolation. These reserved samples were conducted comparative evaluation with the interpolation results. That meant that the model established by the most of the sample was used to conduct the small sample prediction. Meanwhile, the error of prediction was also obtained. The mean absolute error, average relative error and root-mean-square error was used as the standard of measuring the interpolation precision错误!未找到引用源。错误!未找到引用源。.

In this experiment, the randomly selected 50 sample points from the total samples were used as the data waiting for interpolation and need to be predicted. The rest of the sample points were used as the interpolation data. After many times of experiment,

the influence of the different number of the nearest-neighbor to the result data of the sampling point was tested under the same soil texture. The interpolation results analysis were shown in the tables below:

Table 1. Results Statistic of pH Values

Number of adjacent points	Absolute error			Relative error			Root-mean-square error
	Maximum value	Minimum value	Mean value	Maximum value	Minimum value	Mean value	
5	1.2	-1.75	0.0365	0.203725	0.004342	0.0688	0.4681
10	1.17	-1.72	0.0405	0.200233	0.004342	0.06689	0.4569
15	1.15	-1.7	0.0390	0.197905	0.004342	0.065512	0.4484
20	1.14	-1.7	0.0405	0.197905	0.004292	0.065741	0.4478
25	1.13	-1.7	0.0395	0.197905	0.002894	0.06531	0.4448
30	1.12	-1.7	0.0390	0.197905	0.002894	0.065089	0.4431
35	1.12	-1.7	0.0390	0.197905	0.002894	0.065234	0.4427
40	1.12	-1.7	0.0385	0.197905	0.002894	0.065159	0.4422
45	1.12	-1.7	0.0380	0.197905	0.002894	0.06495	0.4415
50	1.11	-1.69	0.0395	0.19674	0.002861	0.064609	0.4388
55	1.11	-1.69	0.0385	0.19674	0.002861	0.064465	0.4383
60	1.11	-1.69	0.0390	0.19674	0.002861	0.064537	0.4386
65	1.11	-1.69	0.0385	0.19674	0.002861	0.064317	0.4380
70	1.11	-1.69	0.0385	0.19674	0.002861	0.064317	0.4380
75	1.11	-1.68	0.0400	0.195576	0.002861	0.064119	0.4366
80	1.1	-1.68	0.0405	0.195576	0.002861	0.064052	0.4362
85	1.1	-1.68	0.0405	0.195576	0.002861	0.064052	0.4360
90	1.1	-1.68	0.0425	0.195576	0.002861	0.064205	0.4363
95	1.1	-1.68	0.0435	0.195576	0.001431	0.064063	0.4364
100	1.1	-1.68	0.0440	0.195576	0.001431	0.063994	0.4365

Table 2. Results Statistic of Organic Material

Number of adjacent	Absolute error			Relative error			Root-mean-square error
	Maximum	Minimum	Mean value	Maximum value	Minimum value	Mean value	

t points	value	value					
5	11.76	-5.08	0.4455	1.4067	0.0089	0.1507	2.5395
10	11.17	-5.21	0.4070	1.3361	0.0013	0.1415	2.4262
15	11	-5.09	0.3870	1.3158	0.0009	0.1417	2.4147
20	10.96	-4.98	0.4205	1.3110	0.0012	0.1389	2.4227
25	10.92	-4.97	0.4415	1.3062	0.0029	0.1381	2.4306
30	10.9	-4.96	0.4500	1.3038	0.0053	0.1380	2.4306
35	10.89	-5	0.4565	1.3026	0.0059	0.1390	2.4506
40	10.88	-5	0.4540	1.3014	0.0076	0.1389	2.4506
45	10.87	-5	0.4525	1.3002	0.0082	0.1390	2.4489
50	10.86	-5.02	0.4490	1.2990	0.0094	0.1395	2.4512
55	10.88	-5.06	0.4440	1.3014	0.0099	0.1405	2.4623
60	10.88	-5.09	0.4460	1.3014	0.0099	0.1409	2.4648
65	10.89	-5.13	0.4495	1.3026	0.0099	0.1413	2.4676
70	10.88	-5.15	0.4635	1.3014	0.0099	0.1416	2.4678
75	10.89	-5.19	0.4640	1.3026	0.0099	0.1425	2.4734
80	10.9	-5.21	0.4720	1.3038	0.0099	0.1429	2.4785
85	10.91	-5.22	0.4780	1.3050	0.0094	0.1436	2.4840
90	10.91	-5.21	0.4830	1.3050	0.0099	0.1442	2.4881
95	10.91	-5.21	0.4835	1.3050	0.0103	0.1446	2.4909
100	10.92	-5.2	0.4790	1.3062	0.0108	0.1448	2.4939

Table 3 Results Statistic of Rapidly Available Phosphorus

Number of adjacent points	Absolute error			Relative error			Root-mean-square error
	Maximum value	Minimum value	Mean value	Maximum value	Minimum value	Mean value	
5	8.16	-18.28	-1.7905	1.1209	0.0019	0.3219	6.9795
10	8.15	-17.75	-1.7320	1.1195	0.0102	0.3178	6.6646

15	8.06	-17.84	-1.7665	1.1071	0.0015	0.3166	6.5797
20	8.01	-17.76	-1.7455	1.1003	0.0029	0.3162	6.5122
25	7.99	-17.76	-1.7300	1.0975	0.0031	0.3169	6.4578
30	7.97	-17.79	-1.7420	1.0948	0.0051	0.3171	6.4259
35	7.95	-17.8	-1.7550	1.0920	0.0072	0.3177	6.4065
40	7.94	-17.77	-1.7735	1.0907	0.0072	0.3170	6.3900
45	7.93	-17.74	-1.7910	1.0893	0.0093	0.3179	6.3834
50	7.92	-17.7	-1.7915	1.0879	0.0103	0.3185	6.3733
55	7.92	-17.68	-1.7970	1.0879	0.0113	0.3181	6.3628
60	7.9	-17.68	-1.7950	1.0852	0.0144	0.3186	6.3545
65	7.9	-17.68	-1.7955	1.0852	0.0175	0.3184	6.3513
70	7.9	-17.68	-1.8010	1.0852	0.0175	0.3182	6.3490
75	7.9	-17.67	-1.8060	1.0852	0.0175	0.3179	6.3439
80	7.9	-17.67	-1.8145	1.0852	0.0163	0.3180	6.3433
85	7.89	-17.66	-1.8160	1.0838	0.0116	0.3181	6.3410
90	7.89	-17.67	-1.8160	1.0838	0.0088	0.3184	6.3395
95	7.89	-17.67	-1.8205	1.0838	0.0075	0.3185	6.3384
100	7.89	-17.67	-1.8285	1.0838	0.0034	0.3183	6.3368

Table 4 Results Statistic of Rapidly Available Potassium

Number of adjacent points	Absolute error			Relative error			Root-mean-square error
	Maximum value	Minimum value	Mean value	Maximum value	Minimum value	Mean value	
5	57.71	-47.5	0.0200	0.4809	0.0110	0.1964	24.2538
10	57.16	-48.44	-0.4010	0.5057	0.0180	0.1958	24.3808
15	56.74	-49.41	-0.2720	0.5072	0.0124	0.1942	24.3740
20	56.52	-50.03	-0.3100	0.4966	0.0090	0.1923	24.2862
25	56.25	-50.09	-0.2755	0.4954	0.0084	0.1908	24.1878
30	55.95	-50.21	-0.2020	0.4999	0.0037	0.1899	24.1507
35	55.73	-50.29	-0.0700	0.5020	0.0015	0.1897	24.1325
40	55.57	-50.35	-0.0160	0.5028	0.0012	0.1893	24.0965
45	55.41	-50.49	-0.0160	0.5035	0.0011	0.1887	24.0536
50	55.28	-50.56	0.0645	0.5024	0.0023	0.1891	24.0341

55	55.15	-50.65	0.0175	0.5003	0.0019	0.1888	23.9893
60	55.03	-50.69	0.0065	0.4993	0.0023	0.1888	23.9824
65	55	-50.74	0.0015	0.4980	0.0031	0.1885	23.9818
70	54.95	-50.84	-0.0190	0.4979	0.0035	0.1885	23.9995
75	54.88	-50.94	-0.0210	0.4967	0.0042	0.1886	24.0169
80	54.84	-50.97	-0.0220	0.4962	0.0043	0.1885	24.0312
85	54.79	-51.04	0.0215	0.4958	0.0044	0.1885	24.0372
90	54.75	-51.04	0.0115	0.4952	0.0044	0.1882	24.0579
95	54.71	-51.1	-0.0100	0.4942	0.0042	0.1882	24.0725
100	54.68	-51.18	-0.0450	0.4933	0.0040	0.1880	24.0808

Because the algorithms was designed based on mobile computing, it could not only guarantee the accuracy but also provide high efficiency. Through many times of experiment contrasting, when the execution efficiency was less than 3s, the variation situation of the soil nutrients interpolation error in the algorithm was examined in 100 neighboring set. The comparing results as follows:

(1) The prediction error of Ph decreased gradually with the increase of the neighbor point firstly. The prediction error reach minimum when the number of the neighbor point increased to 85. The mean absolute error, average relative error and root-mean-square error was 0.0405, 0.064052 and 0.4360 respectively. Then, the error increased gradually again.

(2) The prediction error of organic material decreased gradually with the increase of the neighbor point firstly. The prediction error reach minimum when the number of the neighbor point increased to 15. The mean absolute error, average relative error and root-mean-square error was 0.3870, 0.1417 and 2.4147 respectively. Then, the error increased gradually again.

(3) The prediction error of rapid available phosphorus decreased gradually with the increase of the neighbor point.

(4) The prediction error of rapid available potassium decreased gradually with the increase of the neighbor point. The prediction error reach minimum when the number of the neighbor point increased to 65. The mean absolute error, average relative error and root-mean-square error was 0.0015, 0.1885 and 23.9818 respectively. Then, the error increased gradually again.

5 Conclusions

The K-D Tree was used as the space division algorithm for the soil nutrient sampling point in this study. The spatial index of the soil nutrient sampling point was also

established. On this basis, the K-nearest neighbor search of the soil nutrient sampling point was realized using KNN algorithm. Furthermore, combined with KNN and IDW algorithm, the soil nutrients spatial interpolation algorithm was implemented and breaked away from the limitation of GIS platform.

(1) The experiment showed that the soil nutrient spatial interpolation algorithm based on KNN and IDW was effective and feasible for predicting the content values of the Ph, organic matter, rapid available phosphorus and rapidly available potassium in soil.

(2) This paper tested the influence to different soil nutrient elements affected by the K value in KNN algorithm. The result showed that the neighboring points for the prediction of Ph, organic matter, rapid available phosphorus and rapid available potassium of respectively take 85, 15, maximal sample space, 65 could achieve the best accuracy. Of which, the optimal average absolute errors of Ph, organic matter, rapidly available potassium were 0.0405, 0.3870, and 0.0015 respectively.

(3) Through the analysis of the error size, it was found that the interpolation precision to pH and organic matter of the algorithm is higher than the interpolation precision of the rapid available phosphorus and rapid available potassium. The error of the rapid available phosphorus and rapid available potassium was larger. This was largely relevant with their large spatial variability[24].

Following the rapid development of the mobile intelligent terminal and the distributed computing technology, more and more applications transferred the computing and the data processing to thin client. The soil nutrient spatial interpolation algorithm based on KNN and IDW enabled the traditional soil testing and formulated fertilization recommended system to run on embedded devices such as high performance and low cost mobile intelligent terminal. The advantages of the mobile technology and the soil testing formula technology could thus fully integrate to achieve the positioning and query to the soil nutrient and recommended fertilization information of any plot for the users through the intelligent mobile platform in the fields. It had an important significance for the accurate utilizing of the soil testing and formulated fertilization technology and the resolving of the problem of "the last kilometer" for the scientific fertilization information promotion.

Acknowledgment

References

1. Wu Jing-san, Cheng Gui-cheng. The Existing Problems and Countermeasures of Unit Soil Samples Collecting in Soil Testing and Formulated Fertilization[J]. Modern Agricultural Science and Technology, 2010, (1):274.

2. Tong Yue, Zhang Xiao-xu, Wang Ce et al. The Existing Problems and Countermeasures in Soil Testing and Formulated Fertilization Project[J]．Modern Agricultural Science and Technology, 2014, (13):262.

3. Feng Donghai. The Existing Problems and Countermeasures in Soil Testing and Formulated Fertilization [J]．Modern Agricultural Science and Technology, 2011, (18):298-299.

4. Wang Yue,Yue Wei-ning, Wang Heng, et al. Multi-Modal Iinteraction in Handheld Mobile Computing. Journal of Software, 2005,16(1): 29-30.

5. Yu Yisheng , Hu Fenggen . Processing model of mobile data in disconnecting computing environments[J].Computer Engineering and Design, 2006,27(14):2686-2687.

6. Yu Tao, Yu Lizhong, Wang Zheng.The Development and Application of GIS under Mobile Computing Environment[J].Bulletin of Surveying and Mapping, ,2002,(2):40-41.

7. Zhang Li, Tian Yin-lei. PDA-based mobile query system for testing soil for wheat formulated fertilization[J]. Journal of Henan Agricultural University,2010, 44(3),341-342.

8. Zhang Guo-feng, HE Li-yuan, Han Song, et al. Research and Development of Recommended Fertilization System Based on Mobile GIS[J]. Journal of Huazhong Agricultural University,2011,30(4):484-486.

9. Xu Xin, Zhang Hao, Xi Lei, et al. Decision-making System for Wheat Precision Fertilization Based on WebGIS [J]. Transactions of The Chinese Society of Agricultural Engineering, 2011, 27(14): 95-97.. (in Chinese with English abstract)

10. Jiang Hai-yan, Mao Jin-hui, Xu Xiao-ming. Support System for Wheat Production Management Based on Service-oriented Architecture and WebGIS[J]. Transactions of The Chinese Society of Agricultural Engineering, 2012,28(8):160-165.

11. Zhao Qing-song, Xu Tao, Jiang Hai-yan, et al. Cropping System Design Service System Based on SOA and WebGIS[J]. Computers and Applied Chemistry, 2014,31(6):757-759.
12. Zhu Qiuan, Zhang Wanchang, Yu Junhui. The Spatial Interpolations in GIS[J]. Journal of Jiangxi Normal University(Natural Science), 2004,28(2):184-186.

13. Wang Xiu, Miao Xiao-ke. Meng Zhi-jun et al.Effect on Soil Nutrition Interpolation Result of Different Interpolation Styles[J].Chinese Journal of Soil Science, 2005,36(6):827-829. 14. Zu Juan. Researched and Developed a Soil Testing and Formulated Fertilization Application System Based on PDA-GIS[D]. Anhui Agricultural University,2012.

15. Xu Li-Hua. Comparisons among different Prediction Method on the Soil Nutrient-Taking Wangjiagou small Watershed of Three Gorges Reservoir Area as Researching Zone. Southwest University,2012.

16. Li Zeng-bing, Zhao Geng-xing, Zhao Qian-qian.Comparison of Spatital Interpolation Methods for Soil Nutrients in Cultivated Land Fertility Evaluation[J].Chinese Agricultural Science Bulletin,2012,28(20):231-236.

17. Jin Guo-dong, Liu Yan-cong, Niu Wen-jie.Comparison between Inverse Distance Weighting Method and Kriging[J].Journal of Changchun University of Technology, 2003,24(3):54-56.

18. Chen Guang,He Li-yuan,Zhan Xiang-wen. Comparison of Spatial Interpolation Technique of Soil Nutrient and Reasonable Sampling Density [J]. Chinese Journal of Soil Science, 2008, 39(5): 1008－1011.

19. Dung Minl.Study on Soil Property with GA-RBF-Neural-Network-Based SpatialInterpolation Method [D]. Sichuan Agricultural University, 2009.

20. Cover T M，Hart P E．Nearest neighbor pattern classification[J]．IEEE Transactions on Information Theory，1967，13(1):21－27.

21. Zhang Zebao. Research of Spatial Indexes Optimization and Implementation[D]. Harbin Engineering University, 2005.
22. Yan Chaode,Zhao Xuesheng.The Review of Spatial Indexes in GIS[J]. Geograph and Geo - Information Science, 2004,20(4):23-26.
23. Shi Xiao-hua.Yang   lian-an,Zang   Lei. Comparison of Spatial Interpol ation Methods for Soil Avail abl e Kalium[J]. Journal of Soll and Water Conservation,20(2):70.
24. Wang Ke,Shen Zhangquan,John S.Bailey, et al. Spatital Variants and Sampling Strategies of Soil Properties for Precision Agriculture[J]. Transactions of The Chinese Society of Agricultural Engineering, 2001, (2):34－35.