Acta Technica 62 (2017), No. 5A, 345–356

Development and research of agricultural big data visualization system

Weijie Chen[2], Pingzeng Liu[2,3], Chao Zhang[2], Maoling Yan[2], Rui Zhao[2], Bangguo Li[2]

Abstract. In order to effectively handle and analyze the agricultural big data with explosive growth, the agricultural big data visualization system is designed and implemented. With the aid of mature third party controls, named ECharts, the design process of collecting data, handling data, matching graph, optimizing graph and inspecting test data visualization is proposed. Then based on the four kinds of farming data including meteorological, soil, water level and crop, the agricultural data visualization module based on space and time sequence was designed and implemented. This not only improves the utilization rate of agricultural big data, but also enables the interaction between data and farmers. It makes easier for farmers to see the changing regulation of multiple attributes that represent the data of agricultural objects or events. The visualization mechanism of agricultural big data provides a great supporting role in the progress of precision agriculture.

Key words. Agricultural big data, data visualization, precision agriculture, third party control.

1. Introduction

Agricultural big data is the application and extension of big data in the agricultural field [9]. It combines the characteristics of agriculture, regional, seasonal, diversity and periodicity. After that, its data set has the characteristics of extensive source, various types, complex structure, potential value, and difficult to apply the usual methods of processing and analysis [1,8].With the rapid development of agricultural information technology, agricultural data also showed explosive growth. It is of great practical significance to implement the efficient utilization of agricultural

[1]Acknowledgment - We acknowledge the financial supports of the Independent Innovation and Achievement Transformation of Shandong Province under Grant No. 2014ZZCX07106.

[2]Workshop 1 - College of Information Science and Engineering, Shandong Agricultural University,Tai'an Shandong, 271008, China

[3]Corresponding author: Pingzeng Liu; e-mail: `lpz8565@126.com`

big data.

Visualization of agricultural data is a kind of service mode that combines agricultural data and visualization technology. Visualization of agricultural data can not only provide basic information from large amount of data, but also apply complex analysis, improve the utilization rate of agricultural big data and facilitate the interaction between data and farmers. It can also make farmers more convenient to see the change rules of multiple attributes of agricultural objects or events, so as to provide convenience for farming operations [3]. The origins of data visualization can be traced back to the early days of computer graphics in the 1950s. In 1987, the National Science Foundation Report, called "Visualization in Scientific Computing", was written by Bruce, Thomas and Maksim, which had greatly promote and stimulate in data visualization. At present, in the field of research and development of agricultural big data processing analysis and visualization display, some visualization techniques and methods have emerged, which has improved the efficiency of visualization technology in the field of agriculture. In the field of simulation of plant growth, through the virtual reality technology, many kinds of plant growth models, such as iterative function system and particle system, have been established by scholars at home and abroad. A variety of software to simulate plant morphology has been developed. The purpose of visual observation of the morphological structure of the crops was realized on the computer [10,11].In the field of farmland environmental monitoring, researchers at home and abroad have developed a visualization system for farmland environmental monitoring data. Lin Lanfen et al., Wang Li et al. and Chen Yanqiu designed the multi-dimensional climate and environment data to be synchronized and visualized, and developed the agricultural microclimate environment video monitoring system based on GIS, WebGIS and Geo-WebServices which facilitates the agricultural participants to master the microclimate environment of farmland at any time and place. Wu Y H et al. evaluated the effects of interpolation on unsampled locations by the three spatial interpolation methods (inverse distance weighted, spline and kriging), 2D visualization and 3D visualization. It shows that visualization can provide information not visible from tables, such as exact/inexact interpolation or disintegration along borders [6,7]. Xu Jianning et al. have designed and developed a farmland environmental monitoring data visualization system combining DNDC model, which greatly improves the understandability of agricultural environmental monitoring data.

With the emergence of some visualization software, more and more visual controls and components based on third parties have been applied to various fields, which has improved the efficiency and promoted the visualization [5]. Dong Yadong et al. have developed specifically for the kernel-driven model by a user-friendly visualization tool, named MaKeMAT. The study shows that the visualization tool MaKeMAT can promote the widespread application of the kernel-driven model. Hong Luhe et al. designed and implemented data visualization system based on the DotNetBar control and the view component LoG JViews Diagrammer, which reflecting the interface design of office2007 more friendly and achieving the interaction between the screen object element and application[2].Yang Guangming used CiteSpace information visualization software to draw the scientific knowledge map of soybean research, and intuitively displayed and analyzed the knowledge base and frontier area of soybean research[12].

The third party control–ECharts has many advantages, such as visualized charts, easy operation, easy calling and multiple feedback and modification.It has obvious advantages in data visualization and analysis. With the ECharts, this paper uses four kinds of farming data as variables, according to the order of collecting data, handling data, matching graph, optimizing graph and inspecting test, designing and matching different graphical interfaces for data variables, which provides a great supporting role in precision agriculture.

2. Data material and methods

2.1. Related third party control

ECharts is a data visualization library based on Javascript and developed by Baidu's front-end technology department. It provides visual, vivid, interactive and customizable data visualization chart. It provides a large number of commonly used data visualization charts, and the bottom layer is based on ZRender (a new lightweight canvas Class Library), which creates coordinate systems, legends, hints, toolkits, and other basic components. On the basis of the above, line chart(regional map), histogram (bar chart), scatter diagram (bubble), pie chart (ring), K line graph, map, force directed layout diagram and chord diagram are constructed, while supporting the accumulation of arbirary dimension and multi mixed chart show.

2.2. Material selection

In this paper, the four types of farming data, which are perceived by the Internet of Things and artificial collection, are used as data materials. Data sets include meteorological data (air temperature, air humidity, illumination intensity, carbon dioxide concentration, wind speed, wind direction and rainfall), soil data (soil temperature, soil moisture, soil salinity, soil pH value, soil conductivity and organic matter), groundwater data (water level) and crop growth data (the number of secondary roots per plant, plant height, leaf area index, plant fresh weight, nitrogen content, phosphorus content and potassium content). Data materials from the "Bohai granary science and technology demonstration project".

2.3. Test area overview

"Bohai Granary" is essentially a short for "A granary on the low plain around Bohai". The low plain around Bohai refers to the low plain area below 20 m in the western part of Bohai. It lies in the eastern part of the North China Plain and belongs to the part of the Huang Huai Hai plain. It consists of alluvial the Yellow River, Haihe River, Luanhe River, Liaohe River and other rivers and mainly related to three provinces and one city, including Shandong provinces, Hebei provinces, Liaoning provinces and Tianjin city [4]. "Bohai Granary" demonstration base in

Shandong province covers Dezhou, Binzhou and Dongying three cities. It covers 26333.33 hectares of saline-alkaline land. And a total of more than 140 sets of Internet of Things equipment are installed in the region.

3. Design of agricultural data visualization

The design flow of agricultural data visualization is a cyclic process, each of which has its specific mission indicators, and developers need to move to the next level with the completion of the previous task. The design of visualization mainly consists of five parts: collect data, analysis data, match graph, optimize graph and inspection test, as shown in figure 1.

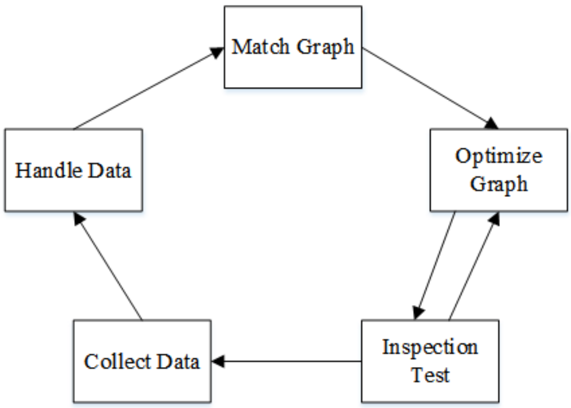

Fig. 1. Design flow chart of visualization system

3.1. Collect data

The data acquisition consists of two parts, the first part of which is the digitization of the crop growth. By analyzing the factors affecting the growth of wheat, maize and rice, combined with the characteristics of the saline-alkali landforms and climatic characteristics of the area, the types and quantities of information needed for crop production information management were determined. The use of Internet of Things technology and human resources to achieve real-time field information collection and crop growth information, to achieve the digital management of crop growth. The second part is to digitize the information program. Substantially, to capture the data of the Internet of Things perceived and artificially collected from the database through related programs and temporarily stored in the array of variables. The purpose of this layer is to provide data preparation for the next layer.

3.2. Handle data

First, the array of variables captured on the previous layer is treated as a dimension. Standardize the data according to the the following formula to the [0, 1] interval, as shown in table 1.

$$y = \frac{(x - MinValue)}{(MaxValue - MinValue)}$$

Subsequently, the dimensional variables are classified, segmented and transformed through algorithms of the data filtering and data conversion to convert an array of primitive variables into a form that can be used. Data filtering mainly deals with the problem of dealing with vacancy value and outliers. Since nearest neighbor objects have similar data values, the distance weighted nearest neighbor method is used for data filtering. The basic idea is to determine K samples which are correlated with missing or outlier data samples according to Euclidean distance or correlation analysis, using the K samples. The weighted average of the K values is used to estimate the missing or outliers' data values of the sample, such as the following formula,

$$f(S_i) = \frac{1}{k} \sum_{m=1}^{n} \sum_{j=1}^{k} W_j \partial(V_m, f(S_j))$$

In that formula , $\mathrm{w}_j$represents the distance weight of nearest neighbors' s_j. Data conversion is mainly used to remove noise and achieve the purpose of handling repetitive data and noisy data. The wavelet transform algorithm is adopted [10], such as the following formula,

$$WT(\partial, \tau) = \frac{1}{)a} \int_{-\infty}^{\infty} f(t) * \varphi(\frac{t-\tau}{a})dt$$

In that formula, the a (scale) controls the scaling of wavelet functions. The t (translation quantity) controls wavelet function evaluation. Scale correspondence frequency (inverse ratio), translation quantity corresponding time. Data preprocessing can not only improve the system efficiency, but also meet the user's diversity needs.

3.3. Match graph

Matching graph is an association display of data and graphical interface of the third party control. According to the data characteristics, the appropriate visual display form is selected, as shown in table 2. Among them, the meteorological data, soil data and water level data have a strong timeliness. They can be divided into short-term data and long-term data according to the time period, and can be divided into real-time dynamic data and historical data according to data usage. The effects of different environmental conditions on the development of crop internal contradiction are different, so the crop data has a strong regional character.

The specific steps are as follows:

1. To introduce the configuration file parameters for this framework.

<script type="text/JavaScript" src="js/echarts.js"></script>

2. To prepare containers for graphics.

<div id="chartmain" style="width:600px"; height: 400px; "></div>

3. To specifie the configuration and data of the icon.

var option={};

4. Initializing the echarts instance.

var myChart = eharts.init(document.getElementById('chartmain'))

5. To use the established configuration items and data display charts.

MyChart.setOption (option)

This can make users better understand data and analyze data. To analyze the original data set and data mining result set from different emphases, analyst can also select multiple graphics to implement linkage display.

Table 1. Dimensional Processing of Variables Array

melogical data	variable	Air temperature	Air humidity	Illumination intensity	Carbon dioxide concentration	Wind speed
	unit	[ ]	[%]	[klux]	[ppm]	[m/s]
soil data	variable	soil temperature	soil moisture	soil salinity	soil pH value	organic matter
	unit	[ ]	[%]	[mol/L]	–	[%]
Groundwater data	variable	water level				
	unit	[m]				
Crop growth data	variable	number of secondary roots per plant	plant height	leaf area index	fresh plant weight	nitrogen content
	unit	[strip]	[cm]	–	[kg]	[%]

Table 2.Selection of graphical interfaces for agricultural data visualization systems

	ECharts Graphical	The type of a variable array	Corresponding data types
Data visualization based on time sequence	Line chart	Continuous ,two-dimensional data	Meteorological data, soil data, water level data
	Histogram	Continuous , two-dimensional data	
	Scatter plot	Continuous , two-dimensional data	
	Heatmap of the Descartes coordinate system	Continuous , two-dimensional data	
Data visualization based on space	Pie chart Histogram	Multidimensional, continuous, discontinuous data	Crop data soil data
	Map	Multidimensional, regional, continuous , discontinuous data	

3.4. Optimize graph and inspection test

The matching graphics are displayed on the interface of the system. Do detailed optimization and continuous testing for the displayed graphical interface according to your own requirements. If there is an inappropriate place, you need to go back to the upper layer and rematch the new graphics.

The following four types of farming data collected at 001 site of Leiji Town, Xiajin County, Dezhou City, as the data set, according to collect data, analyze data, match graphics, optimize graphics and inspection test, to achieve visualization module based on space and time sequence.

4. Implementation and analysis of visualization module

4.1. Implementation and analysis of visualization module based on space

The design flow of the visualization module based on space is shown in Figure 2. Data sets based on space, in general, including geographic location information and crop information. At first, the geographic information data stored in the database is captured and temporarily stored in a variable array. Then, the graph is matched according to the data characteristics from the ECharts and displayed on the interface. Finally, to determine the feasibility of the graph in the process of inspecting test. If the visual effects are not qualified, it is necessary to rematch new graphics.

Data points with latitude and longitude are deployed on a map of a unit level. It represents the specific distribution of a part in a whole, mainly by points or areas scattered in the same or different geographical locations of the map. Figure

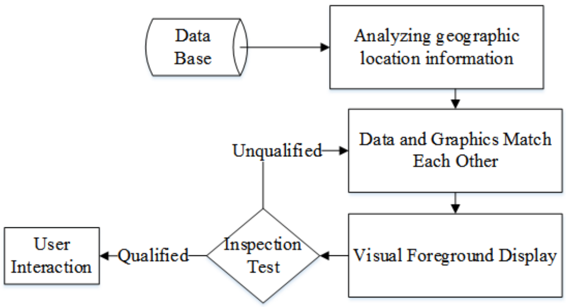

Fig. 2. Design flow of visualization module based on space

3 shows the detailed distribution of the Internet of Things equipment in the "Bohai Granary" demonstration base in Shandong province. This has a more profound impact on people.

Fig. 3. Spatial distribution of internet of things collection sites in shandong project demonstration base

Histogram is a statistical report chart that expresses the graph with rectangle length as variable. It represents data distribution by a series of highly striped longitudinal stripes, which is used to compare two or more values. Figure 4 shows the different regional wheat analysis chart. In the figure, the characteristic parameters (including the number of stems, number of single plant tillers, more than three big tiller number and number of secondary roots per plant) using Delphi method to determine the index weight characteristic parameters, on the basis of this, given the growth grade. This to the user more intuitive display in various regions of wheat, and inter regional comparative growth.

Pie chart is one of the most widely used statistical graphs. It mainly deals with the percentage relation between parts and whole. The combination of soil nutrients and pie chart in crop growth information will help users to clarify the changes in soil nutrient content during crop growth and provide scientific and reliable data support for agricultural production. Figure 5 is the proportion of soil nutrients in 0-20cm and 20-40cm at 001 site of Leiji town Xiajin Country Dezhou City. Users can dynamically add the number of elements in the pie chart, and can also compare multidimensional statistical data. Different colors can represent different data elements in pie charts.

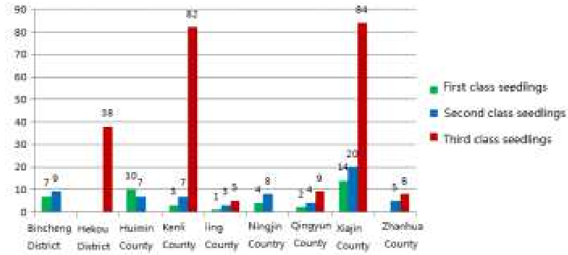

Fig. 4. Analysis of the growth of wheat seedlings

At the same time, it can clearly show the proportion relation between the part and the whole, so that the influence is more intuitive.

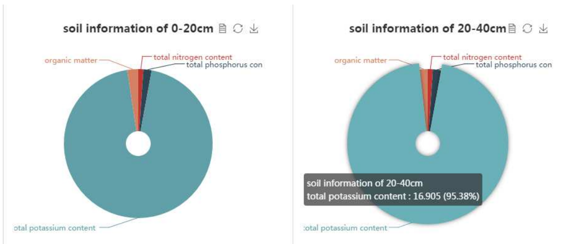

Fig. 5. Proportion map of soil nutrient content

4.2. *Implementation and analysis of visualization module based on time sequence*

The design flow of the visual module based on time sequence is shown in Figure 6. First, variables based on the time sequence are captured from the database and temporarily stored as arrays. Subsequently, the arranged in time order arrays are classified and processed according to the attributes of the data. Finally, the graph is matched according to the characteristics of the data after preprocessing from the ECharts, and displayed at the interface. Finally, to determine the feasibility of the graph in the process of inspecting test. If the visual effects are not qualified, it is necessary to rematch the new graphics.

A certain number of mapping in a certain unit length of the curve, according to the mapping relationship to find the corresponding point, these points are connected in order to form the graphics is the line chart. The law of increasing and decreasing of continuous data with time,the variation trend or the magnitude and rate of increase or decrease, and the peak value of the graph can be expressed by the line chart. The air temperature at site 001 of Leiji Town, Xiajin Country, Dezhou City is visualized by a line chart, as shown in Figure 7. The time data items are evenly distributed on the X horizontal axis, and the values of the air temperature are uniformly distributed along the vertical axis of the Y. The line chart shown in Figure 7 shows not only the real-time data, but also the maximum, minimum and average values of the data.

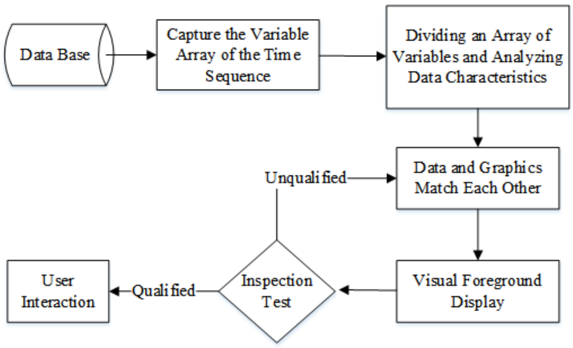

Fig. 6. Design flow of visualization module based on time sequence

Users can operate according to their own needs, so as to get more knowledge.

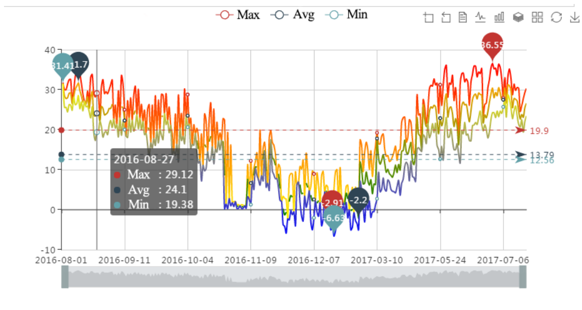

Fig. 7. Trend analysis of data collected by internet of things in real time

The data is displayed in the Descartes coordinates in a specially highlighted format, providing key information about the location and dispersion of the data, which is called heatmap of the Descartes coordinate system. The air temperature at site 001 of Leiji Town, Xiajin Country, Dezhou City is visualized by a heatmap, as shown in Figure 8. The X axis is uniformly distributed on the date, and the Y axis is evenly distributed at the time point of 0-24. The X axis and the Y axis intersect with a bright-dark color to indicate the temperature of the soil at some point in the day, which not only shows the short-term change of soil temperature (midday temperature is high, night temperature is low), but also shows long-term variation (winter temperature is low and spring or autumn temperature is higher).

5. Discussion and Conclusion

(1) The data visualization and visualization analysis rely on data. In the background of big data age, there are five characteristics of data, namely, large amount of data, many types, low value density, high speed and data in the linear. And agricultural data has the above five major functions of the premise, its data source with heterogeneous characteristics. And the transmission data interface and sensors

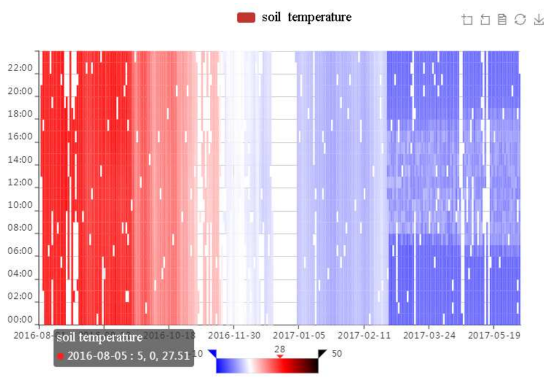

Fig. 8. A heatmap of the soil temperature in the descartes coordinate system

and other hardware devices make the data inherent error. The appropriate data processing algorithm is needed to analyze and deal with them, and the outlier data is excluded so as to make up for the data problems caused by hardware or software, and lay a solid foundation for data visualization.

(2)In order to highlight data features and display data content more fully, we can combine time sequence and space to do linkage display. The two-dimensional data representing the time is connected with the three-dimensional data representing the area by the linkage between the map and the line chart, the line chart and the pie chart, the pie chart and the scatter plot and so on, which can make user more intuitive observation of different regions based on the timing of the data changes.

(3)In summary, in order to effectively deal with the agricultural big data with explosive growth, the agricultural big data visualization system based on the third party controls is developed and implemented. This not only improves the utilization rate of agricultural big data, but also enables the interaction between data and farmers and also makes it easier for farmers to see the changing laws of multiple attributes that represent the data of agricultural objects or events. The visualization mechanism of agricultural big data provides a great supporting role in the progress of precision agriculture.

References

[1] F. Guan, H. Liu, W. Cao: *2D&3D interactive technology in disaster visualization system.* Science of surveying and mapping *37* (2012), No. 3, 97–99.

[2] L. H. Hong, J. L. Cai, S. X. Wu: *Design and implementation of data visualization system based on third-party control.* Computer engineering and design *31* (2010), No. 13, 3096–3099.

[3] D. Keim, H. Qu, K. L. Ma: *Big-data visualization.* IEEE computer graphics & applications *33* (2013), No. 4, 20–21.

[4] M. X. Gao, L. I. Ran, T. F. Gong: *Analysis on the promoting strategy of the "Bohai Granary" construction.* Research of agricultural modernization *36* (2015), No. 2, 245–251.

[5] M. Peng, H. Fu, J. M. Huang: *High quality microblog extraction based on kernel principal component analysis and wavelet transformation.* Computer engineering *42* (2016), No. 1, 180–186.

[6] R. P. Li, X. H. Lu, C. X. Ma: *Three-dimensional visualization of farmland consolidation planning based on GIS.* Transactions of the CSAE *26* (2010) 302–305.

[7] Y. H. Wu, M. C. Hung, J. Patton: *Assessment and visualization of spatial interpolation of soil pH values in farmland.* Precision agriculture *14* (2013), No. 6, 565–585.

[8] Y. Wan, F. L. Bian: *Design and implementation of interactive 2D&3D GIS architecture.* World of geographic information *6* (2008), No. 2, 48–52.

[9] Y. G. Ren, G. Yu: *Research and development of the data visualization techniques.* Computer science *31* (2004), No. 12, 92–96.

[10] F. S. Kong, Y. L. Wu, H. Y. Jiang: *Visual modeling of lettuce from based on image feature extraction.* Agricultural science in Anhui *43*, (2015), No. 24, 265–268.

[11] W. T. Reeves, R. Blau: *Approximate and probabilististic algorithms forshading and rendingstructured particle system.* Computer graphics *19* (1985), No. 3, 313–322.

[12] G. M. Yang, S. P. Li, T. F. Han: *Global soybean research dynamic analysis based on SCI-EXPANDED database.* Soybean science *33* (2014), No. 2, 232–248.

Received November 16, 2017