The digital system of archaeological

CAO Ligang[1], WANG Xuben[1], MIAO Fang[12], YI Fasheng[2]

1.Key Laboratory of Earth Exploration and Information Techniques（Chengdu University of Technology），Ministry of Education，China, 610059

2.Director of Key Lab of Pattern Recognition & Intelligent Information Processing of Sichuan Province, Chengdu, Sichuan, P.R. of China 610106

ABSTRACT: The digital system of Archaeological includes multi-scale non-destructive detection of archaeological methods, data mining technologies and the GIS of archaeological detect. Heritage preservation is not just for the protection of cultural relics have been excavated. Using non-destructive detection method to detect archaeological artifacts and clarify the situation of cultural relics buried underground. The detection data through the using of data mining algorithm make out the archaeological information. Finally, use GIS technology to achieve the detection of data management and integration of data mining methods. A whole system of digital archaeological built on the GIS platform, based on the using of data mining technology to achieve a detection method and archaeological information mapping, the system for the digital archaeology has provided a complete technical support.

Keywords: Archaeology explore, GIS, Data mining

1. PREFACE

The digital system of Archaeological includes multi-scale non-destructive detection system of archaeological methods, data mining system and the realization of such archaeological GIS. The non-destructive method to detect archaeological relics buried underground data; data mining technology to detect archaeological data information extraction, the archaeological GIS management and display the data mining results.

The geophysical methods for archaeological exploration from a single method to the multi-method, multi-scale. In 1982, the Archaeological Institute of Shanxi Province with magnetic, electric and geochemical exploration methods proved the scope of the Qin Shi Huang Mausoleum and the Tomb exist large quantities of mercury interior; in 1987, Shanxi Provincial Center for Remote Sensing and Aerial Survey cooperate with Zhao-ling Museum "Zhao-ling, Jian-Ling ancient burial sites and positioning of Remote Sensing Research"; 2004, Chengdu University of Technology and the archaeological institute in Chengdu and the archaeological institute in Sichuan research in Jinsha and Sanxingdui sites and the use of electric, magnetic, remote sensing and other geophysical methods to archaeological exploration.

The data mining method has been applied to the field of archaeological research in recent years. Archaeological data mining is currently mainly used in the traditional archaeological relic, according to clusting and classification disentomb historical relic. Such as in Shanxi Lintong Jiangzhai Site to conduct archaeological stratigraphic section of Spatial Data Mining research; The settlement of Jiangzhai archaeological application by clustering algorithm. This paper studies the digital system of archaeological based on GIS, specific the detection of data for data mining and archaeological information extraction, and ultimately the use of GIS platform for data management and display of results.

Received: **; Accepted:**

Foundation: Development Fund of Key Laboratory of Earth Exploration and Information Techniques（Chengdu University of Technology）

First author biography: CAO Ligang(1976—), male, Lecturer (Chengdu University of Technology,Earth Exploration and Information Techniques)
E-mail:ligang.cao@gmail.com

*ligang.cao@gmail.com; phone +86 028 84078828;

2. THE KEY TECHNOLOGY OF DIGITAL SYSTEM OF ARCHAEOLOGICAL

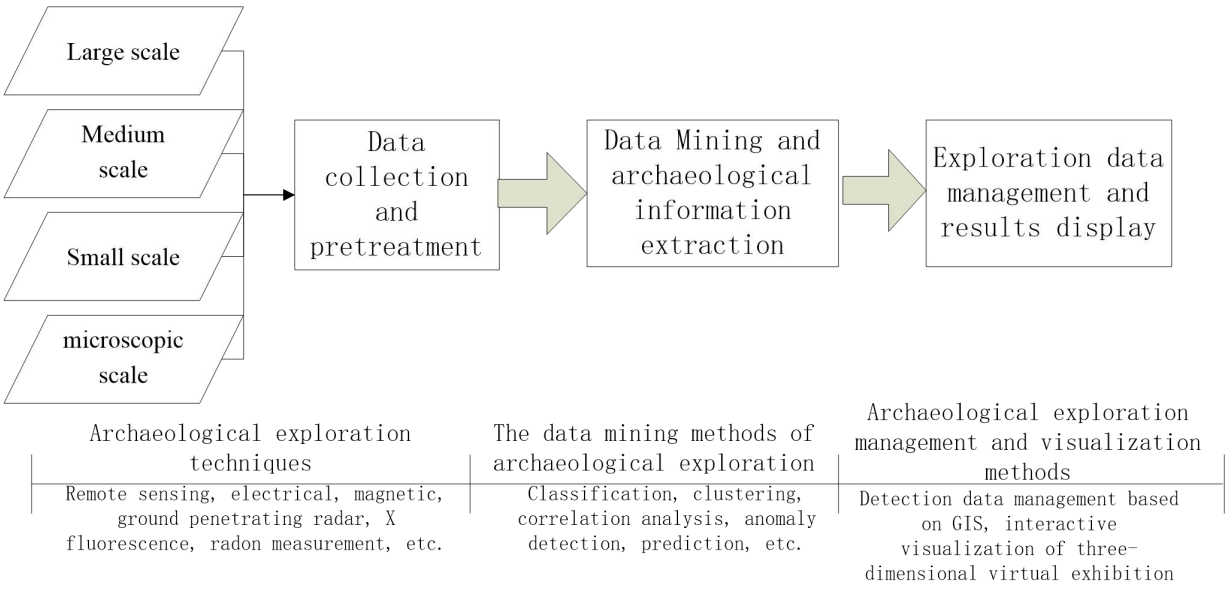

Figure.1. Archaeological exploration flow

The digital system of archaeological includes detection of archaeological techniques, data mining methods of archaeological exploration and the management of archaeological exploration and visualization methods, namely the realization of the archaeological exploration data acquisition, data analysis, and the visualization functions of archaeological result.

2.1 The technology of archaeological detection

Archaeological detection methods include the detecting the ground, underground, underwater relics of human history, tombs, artifacts, etc., which mainly consist of land-based geophysical survey and water archaeological geophysical survey. The land-based geophysical detect methods include remote sensing, electrical, magnetic, seismic, gravity measurements, and geochemical surveys. Under each method also include a number of geophysical methods, such as remote sensing, including aerial remote sensing, space remote sensing, microwave remote sensing, infrared remote sensing; electrical methods include electrical resistivity method, ground penetrating radar, the natural electric field method, direct-current induced polarization method and multi-frequency electromagnetic method, etc.; magnetic method is mainly high-precision magnetic measurement methods; gravimetry mainly microgravity measurement; geochemical survey, including mercury measurement, and gas measurements, X fluorescence measurements, isotopic measurements and C14 test methods. Water archaeological geophysical methods include magnetic, seismic and acoustic wave method.

2.2 The data mining methods of archaeological exploration

The data mining methods of Archaeological exploration include classification, clustering, correlation analysis, anomaly detection and prediction methods.

Classification must first have a clear definition of the class, but also has a series of classified examples, classification process is to build models, and then the model used to classify unknown data. The classification method include: decision tree classification, rule-based classification, neural networks, support vector machines, Bayesian classifier, genetic algorithm.

Clustering is to vary the individual divided into more sub-group similarity or cluster work. It is the biggest difference from classification is the clustering does not rely on a priori knowledge of pre-established categories. Clustering methods include: Based on the level of approach, based on the division of methods, density-based methods, model-based methods and grid-based methods.

Correlation analysis is used to determine which matters are interrelated and analysis should be considered together. Correlation analysis methods include: frequent pattern mining, sequential pattern mining and structural pattern mining and so on.

Anomaly detection is mainly used for found the different object from most audiences. In the clustering process, the anomalous data points as noise filtering, in some application scenarios, we need to look for these outliers. Anomaly detection methods include: statistical methods, anomaly detection based on the nearby degree, density-based anomaly detection and anomaly detection based on clustering and so on.

Prediction mainly refers to the future behavior prediction or estimate of future value, it is the historical data used to construct the model to explain the current observed data or behavior. Reuse this model applies to draw on the future behavior prediction. Prediction methods include: linear regression, nonlinear regression and other regression-based methods.

More of these methods are not completely separated, a variety of methods can be used in conjunction with each other. Reality, the issue of data mining can also be used to solve a number of ways.

2.3 Archaeological exploration management and visualization methods

There are a large amounts of space data from the multi-scale archaeological detect methods in the ancient ruins, which data necessary for geographic information systems technology management and analysis. After the data mining, the archaeological information used the virtual reality technology, embedded into the GIS platform to display the results. GIS platform is a key of digital system of archaeological, the data of archaeological exploration and data mining algorithm modules are integrated into the platform to achieve integrated management and analysis.

3. THE DATA MINING METHODS OF ARCHAEOLOGICAL EXPLORATION

"We are inundated by information, but lack of knowledge" (Rutherford D. Roger, 1985). Data Mining from the large amounts of data are extracted, or "mining" knowledge. Data Mining Technology combines cutting-edge of multi-disciplinary knowledge, including the database, artificial intelligence technology, knowledge of statistics, machine learning technology, high-performance computing and data visualization technology. We focused on non-destructive archaeological detect data for data mining algorithms and applied research.

The main purpose of the archaeological exploration is to answer two questions: Where is Heritage? What is heritage?

3.1 Positioning issues:

3.1.1 Ancient ruins partition problem:

In order to understand the whole distribution in archaeological sites, can be used remote sensing or ground the electrical, magnetic and other geophysical methods Act. These methods of data mining mainly uses the cluster, such as ancient river border extraction, the direction of the ancient city walls, ancient lakes and other border issues. In the clustering algorithm, we can consider the division based on the methods and density-based methods.

3.1.2 ancient palace zoning issues:

In order to understand the internal zoning issues, such as the sacrificial area, the main places of human activity - residential areas, work areas, the division of imperial tombs and other archaeological distribution problems can be used electrical, magnetic, etc. geophysical methods. The data mining algorithms of these detect data mainly use the classification, or a combination of clustering technology, the first cluster and re-classification, improve classification accuracy. In the use of classification techniques, the first according to the actual situation, collecting training samples, and then extended to the entire classification region. Classification algorithms that could be considered in K nearest neighbor, Bayesian classifier, neural networks, support vector machines and other classification methods.

3.1.3 Location specific relic issues:

In order to understand the distribution of ancient artifacts such as underground distribution, the distribution of ancient tombs and other archaeological issues, can be used magnetic, ground penetrating radar, microgravity and other geophysical methods to detect. These methods of data mining mainly uses clustering, anomaly detection and other methods. Clustering algorithm can be used to consider K-means, K-mode, as well as methods such as edge detection.

3.2 Differentiate questions:

3.2.1 Differentiate antiquarian questions:

In order to determine the types of cultural relics such as pottery, animal remains, ivory, the burning pit, can be used geochemistry, nuclear radiation detection, direct-current induced polarization method (detect metal objects). The data mining algorithms of these detect data use the classification, clustering, anomaly detection and other methods. Classification algorithm can be used to consider the decision tree classification, Bayesian classification methods, anomaly detection algorithms can be considered statistically-based methods and density-based detection methods.

Of course, use a variety of geophysical methods in the key areas can be considered relational analysis algorithm.

The above-mentioned data mining methods are commonly methods. In the data mining of archaeological exploration research, we can combine with some of the latest data mining algorithms to information retrieval and knowledge discovery, such as genetic algorithms, evolutionary computation, ant colony algorithm, and other advanced data mining algorithms.

In some archaeological site we used data mining methods to obtain some good results.

4. DIGITIZED ARCHAEOLOGICAL EXPERIMENT

4.1 Paleochannel recognition

Our archaeological exploration missions are mainly concentrated in the places of worship of the ancient ruins, based on previous archaeological experience, our ancient ancestors often held a ritual in the river, and we hope to be able to find Paleochannel flows to determine the distribution of the sacred area, guiding archaeological excavation or site

protection. So, in a non-destructive detection using high density resistivity method to track the flow of the Paleochannel, the riparian of which maybe the approximate range of the sacred area. The main study is using clustering algorithm for high-density electrical method data in recognition of the Paleochannel flow and change, including using K-means algorithm to cluster analysis in electrical method data for different depths in order to identify the Paleochannel.

4.1.1 Processing flow

The result of Paleochannel toward by using K-means algorithm cluster analysis in the data of the archaeological sites of high-density resistivity method, is consistent with interpret by the experts. Which prove the K-means algorithm can get the information from the data of high density resistivity and the electric method can be promote other ruins. The data-mining process of high density resistivity method is shown in Figure 1-1:

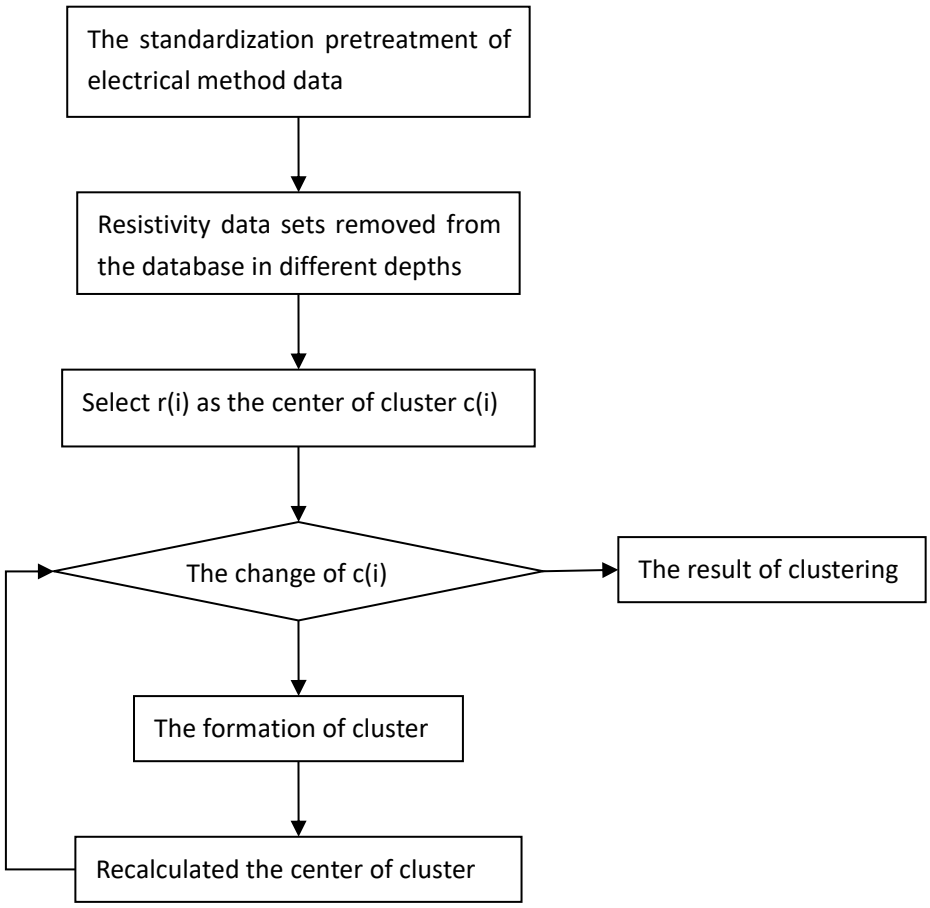

Fig.2 The data-mining process of high density resistivity method

First unified the resistivity value of the property to the range of [0,1], by the data standardization of high-density electric method, mainly to, and get the results of cluster by using the K-means algorithm. Below is the detailed process of cluster:

4.1.2 The preprocessing of high-density resistivity data

During the paleochannel ruins clustering analysis, according to the Chengdu Plain part rock, soil resistivity values of sample data, the data of high-density resistivity need primaries; Electrical method data is a longitudinal section of

data during the ancient river ruins outlined in the horizontal plane sketch, so different probing depth electrical method data preprocessing, in line with the requirements of data mining. According to the high density resistivity method for sample data the measured resistivity in the archaeological sites are shown in Table 1-1:

Table 1 The resistivity of rock, soil measured in the sacred area

Horizon	Lithology	Resistivity $(\Omega \cdot m)$
H2313	Silt layer	17.6
H2313	Silt layer	25.22
43	Clay	37.5
43	Clay	28.7
38	Gray-green clay	32.6
39A	Gray-green clay	37.0
39B	Gray-green clay	37.1
39A・B	Gray-green loam	26.3
H2313	Gray-green loam	24.8
41B	Gray-green loam	24.9
40	Gray-green loam	39.9
39B	Gray-green loam	37.2
35-37	Brownish yellow loam	33.2
35-37	Brownish yellow loam	33.6
The top surface of riverbed	Silt folder pebble	52.6

The basic characteristic of the rock and soil conductivity from the resistivity values in the Table 1-1 :

(1) The range of varies soil resistivity are 24.9-39.9, which is a low resistivity layer;

(2) The range of the silt layer resistivity is 17.6-25.22, which is the conductive best low resistance layer;

(3) The range of gravelly sand and muddy gravel layer resistivity is about 70-114, which is the high impedance layer;

(4) The resistivity of the gravel layer value is greater than 220, which is a high resistivity layer;

According to the requirements of data mining, we should create a database of high density electrical like table 4-4, in which the X, Y, Z are spatial data of detection point, value is the point resistivity values logarithmic of each point and lineno is the measure number of the point. To trace Paleochannel, we utilize a total of 55 survey line and 29,735 data points.

Table 2 The database fields of electric method

Name	Data types	Length
X	float	8
Y	float	8
Z	float	8
value	float	8

lineno	varchar	255

4.1.3 The data extraction of High-density resistivity method

The data of high density resistivity method including spatial coordinates (X, Y, Z) and resistivity values V and measured line number Lineno. The resistivity values used for cluster analysis.

Using minimum - maximum standardized formula for data standardization in resistivity value data preprocessing, Assumptions $\min_A$ and $\max_A$ as the minimum and maximum values of the resistivity V.

$$v' = \frac{v - \min_A}{\max_A - \min_A}(new\_\max_A - new\_\min_A) + new\_\min_A \quad (4\text{-}11)$$

Resistivity value of V is mapped to the interval $[new\_\max_A, new\_\min_A]$ of V'。The mapping interval is [0,1]。

4.1.5 Experimental results and analysis

(1) The toward of paleochannel

We cluster analysis the underground electric density resistivity data extracted from the database by K-means algorithm, which classed by the depth, such as 0-1 meters, 1-2 meters and 2-3 meters.The distribution of resistivity shows that the shape of the paleochannel is a U-shaped, if high resistivity on both sides while low intermediate, the high impedance is the banks of the paleochannel. Figure 1-2 is a sketched diagram of the paleochannel boundary according to the clustering results by the data of depth of 2-3 meters underground. Outline the paleochannel and geophysical interpretation of the ancient river position is basically the same, which achieve the requirements.

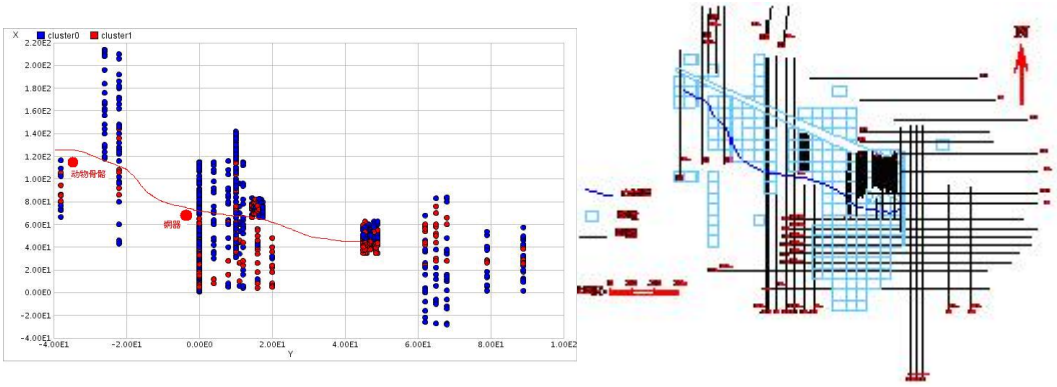

Figure 3 the comparison chart of clustering result and traditional interpretation in left river boundary

（2）The move of paleochannel

By conclusion of the density map of resistivity cluster result including 0-1 meters, 1-2 meters and 2-3 meters: the gravel layer in north is shallow, while the deep in south, which shows that the river is shallow in north and deeper in south, it can be inferred that the paleochannel migration from southeast to northwest.

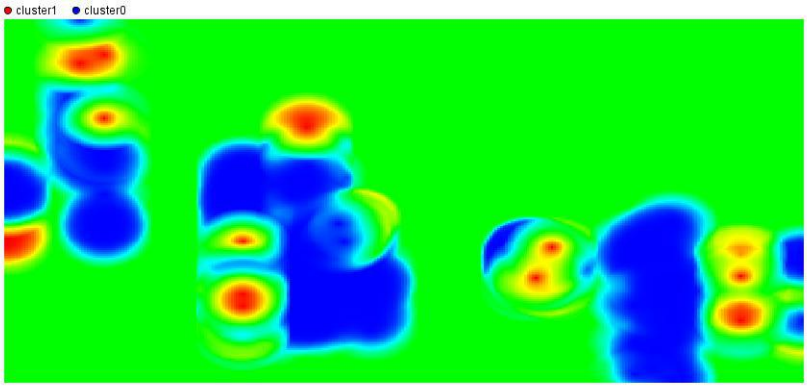

Figure4 The clustering density map of high-density resistivity in the deep of 0-1 meters

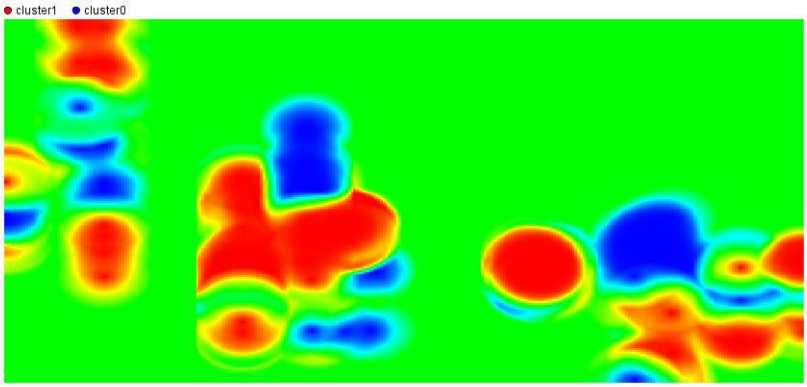

Figure5 The clustering density map of high-density resistivity in the deep of 1-2 meters

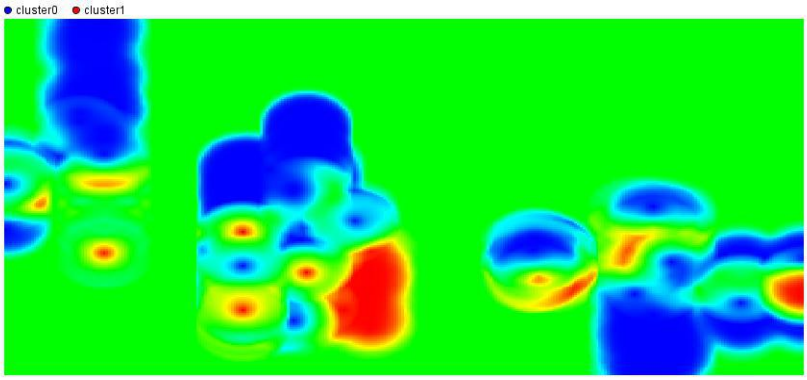

Figure6 The clustering density map of high-density resistivity in the deep of 2-3 meters

By using K-means algorithm on the archaeological sites of high density resistivity method different depth data clustering analysis of the results obtained migration and location of the paleochannel in the archaeological site, which consistent with the results obtained using the traditional geophysical interpretation methods. K-means clustering algorithm can automatically extract data for high-density resistivity method as a new technical support, which can greatly improve the efficiency of the archaeological probing interpretation and provide a new technical means of data processing.

5. GIS MANAGEMENT

The data of archaeological exploration have space properties. The location of archaeological sites and bound can be expressed by spatial data. The ancients left artifacts such as pottery, ivory and so on, are in specific position. Accurate detect data can be used to express their spatial location. Traditional archaeological research generally uses words description or chart. This approach is very simple, but often can not express all archaeological research, and does not analysis and simulation. Archaeological GIS can be easily multi-spatial analysis and modeling, manage detect data and integrate the data mining results.

The function modules of GIS are mainly divided into data management, layer management, spatial analysis and output functions. Data management is mainly responsible for detecting data input and editing and so on, layer management for different data layers to superimpose and edit. Spatial analysis modules for inquiry detect line and some unusual buffer analysis. Output module for thematic map print. Archaeological exploration GIS system functions into modules as shown in Figure 4.

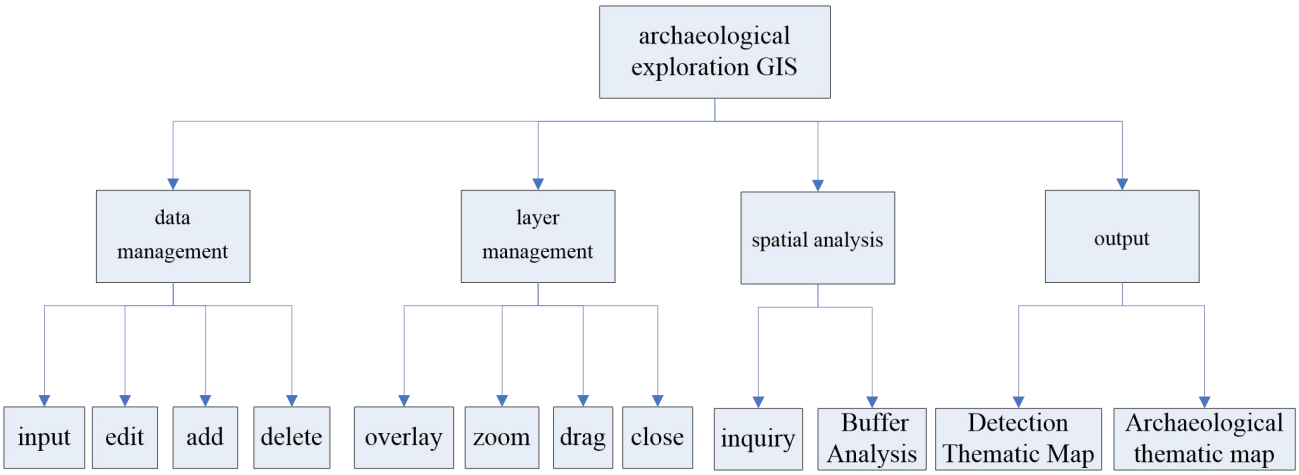

Figure7 Archaeological exploration GIS system functions

The database manages Jinsha base map data, archaeological exploration pit data, detection work area data, electrical data, magnetic data, radar data, as well as artefacts and to detect unusual data management, detect data in the database structure shown in Figure 5 as follows:

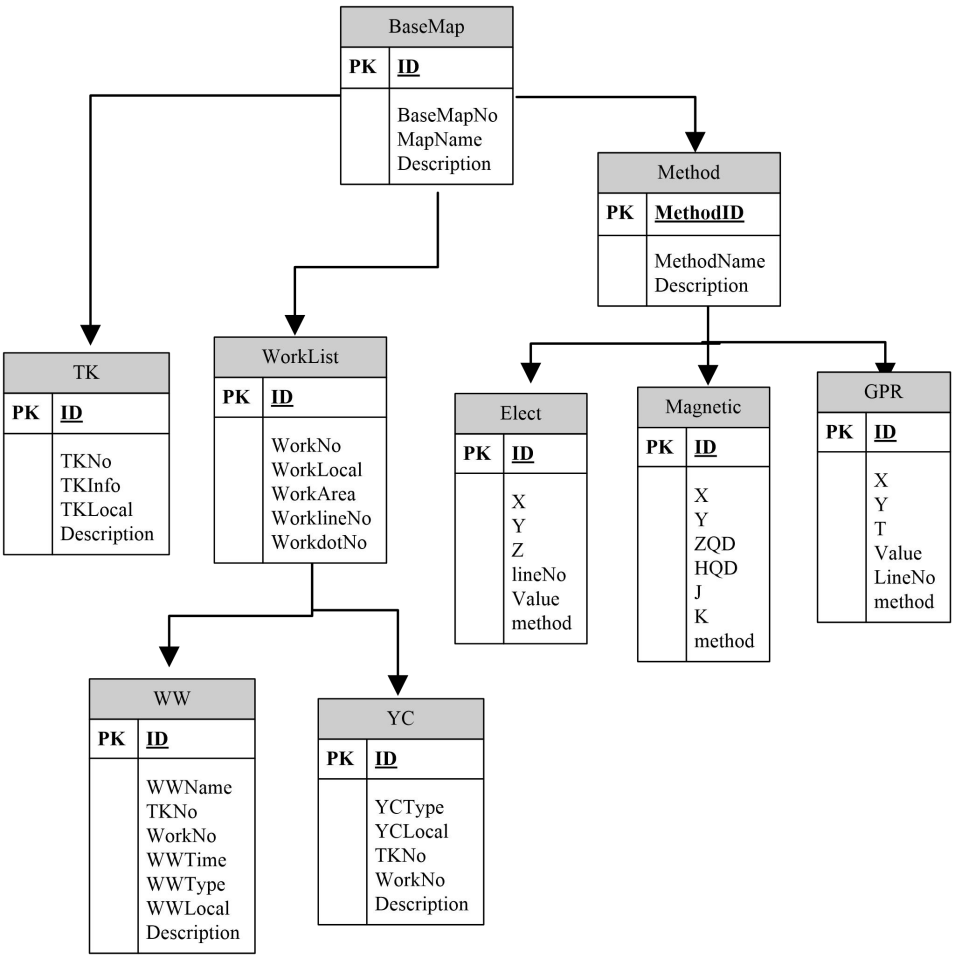

Figure8 archeological exploration database structure diagram

Archaeological exploration GIS system development using C / S framework, use Visual Studio.NET as a system development language and Intergraph's Geomedia Professional 6.0 as a software development platform. According to function of the system design, implementation of the map of basic operations, such as zoom, move, output, etc.; spatial analysis, such as query and analysis, imagery analysis; and layers of management and interactive interpretation of development. Figure 6 for high-density electrical method to detect the GIS Manager screenshot, Figure 7 for the GIS analysis and management of buffer screenshot, Figure 8 for the electrical data mining management screenshot, Figure 9 for the magnetic data mining management screenshot .

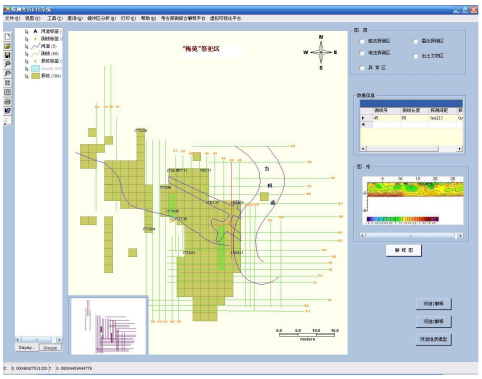

Figure 9 high-density electrical method management

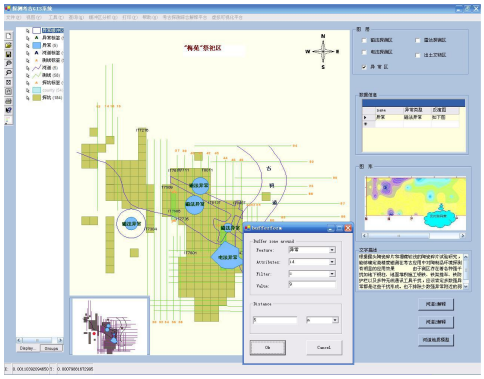

Figure10 GIS analysis and management of buffer

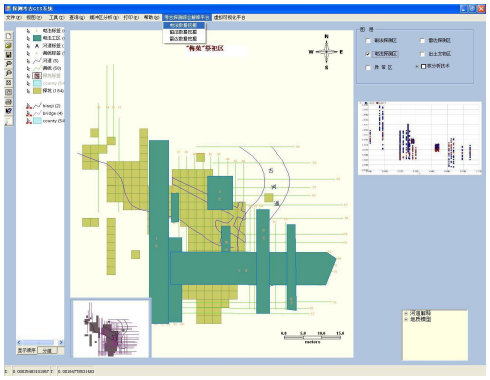

Figure11 the electrical data mining management

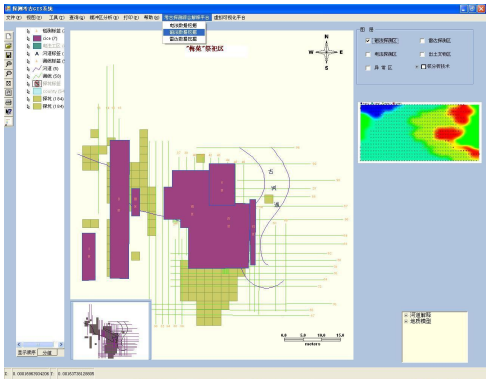

Figure12 the magnetic data mining management

After the data mining of archaeological exploration and knowledge discovery, the obtained results can be used archaeological virtual reality technology to show. The virtual reality display technology make the archaeological results from boring text descriptions to in simple, intuitive way to express them. The using of virtual reality technology can be put ancient archaeological sites in the form of three-dimensional model, this is a novel visual expression of data mining outcome.

6. CONCLUSION

The digital system of archaeological based on GIS integrates archaeological detection technology, data mining technology and geographic information systems technology. Using non-destructive detection of archaeological artifacts we can greatly improve work efficiency in the case of non-excavation. We will be able to conduct archaeological research when conditions are not ripe for excavation. Popularization of this technology system will provide a broad sense in the heritage conservation.

REFERENCES

Cao,Ligang and Wang,Xuben., “Research on interactive 3D scene in digital archaeology“ Computer Engineering & Design.

Li,Hairong. “Integrated interpretation of archaeological detection system support by GIS” papers Chengdu University of Technology (2006).

Micheline Kamber, Jiawei Han [Data Mining: Concepts and Techniques] Morgan Kaufmann; 2 edition (November 3, 2005)

Papers 2008-3

Pang-Ning Tan, Michael Steinbach, and Vipin Kumar [Introduction to Data Mining], Addison Wesley, US ed edition, (May 12, 2005)