A New Visual Associated Analysis Framework and Design Method for Statistical Data

Kaiyong Wang

Institute of Geographic Sciences and Natural Research, Chinese Academy of Sciences, Beijing, 100101, China

wanky@igsnrr.ac.cn

Abstract. According to the analysis to the characteristics of socio-economic data, referring to the information expressing method in Geographic Information System (GIS), this paper designs a visualization analysis system to the socio-economic data with attribute, graphics and charts associated style based on Flash technology. According to the requirement, administrative unit spatial graphics database, socio-economic database is designed and the chart component library and model library are constructed. Analysis result shows that the system can improve the efficiency in socio-economic and other relevant data analysis.

1. Introduction

Socio-economic data is a significant index which reflects the national economy and social development of a country or an area. Every country or area will report its current socio-economic data regularly. Nowadays, statistical departments have already accumulated abundant social data. It becomes more and more important to discover and explore the information and knowledge of the socio-economic data, for it can not only be used to system research and comparative analysis of areal disparity, but also provide reference for local governments' dynamic area monitoring, policy regulation and future economic development program. Deeply analyzing socio-economic data is also the important basis to carry out national macro strategy and make areal development policy. With the rapid development of the current economy, how to dig out information and knowledge from the abundant socio-economic data rapidly is the most urgent demand.

The socio-economic data was almost published in the form of the Statistical Yearbook annually before. Currently the CD version of the yearbook has been attached to the paper version. However, for those researchers, one of the way to get the statistics they need is to buy the statistical yearbook and input the data to the computer by hand, another way is to get digitalized data through Internet, and then analyze and classify them by GIS or other statistical software. The process takes so much time that it cannot serve the rapidly developing economic situation. The methods to analyze socio-economic data mainly include chart analysis and spatial distribution thematic analysis. Chart analysis conducts mathematical statistics with the help of statistical software, while spatial distribution thematic analysis combines attribute data with administrative area special vector graphics by GIS, and then generates thematic map on the basis of attribute fields. Although both of the two methods connect data and software together, they are still researchers' individual conduction which cannot achieve the sharing of network resources, so that different researchers has to install software, organize and analyze statistics repeatedly. Meanwhile, the two analytical methods cannot realize the dynamic visualization analysis with attribute, graphics and charts associated style. They can only express indirectly in the

way of step by step realization and integration, and cannot provide researchers with overall dynamic expression and thinking with visualization effect.

Visualization technology provides new way of thinking and development direction for the quick analysis and visual expression of socio-economic data. Because the incomparable advantage visualization technology has on processing and analyzing information, the technology catches academic circles and government's attention, and the theories and application of visualization become wider and wider. Besides, the development of geographic information visualization technology also promotes the development of the world information science. Visualization, in geographic field mainly refers to the process that transforms information such as abstract concept, changing process and the drive of the object, complex observation and statistical data into graphs that can be directly used for visual understanding and thinking, and most of the research objects have the feature of spatial distribution. Socio-economic data has its spatiotemporal pattern, which is generally divided into statistics combining with time and space, statistics combining with time and attribute and statistics combining with space and attribute. The visualization of spatiotemporal data can be divided into static visualization and dynamic visualization two aspects. Rui Xiaoping and some other researchers think that 3D visualization will be the research focus in the future by dividing it into three aspects: terrain visualization, geological entities visualization and multidimensional visualization adding time dimension and other special dimensions, and also summarizing the latest research process of its three aspects. Nowadays, many scholars have done researches on the visualization expression of spatiotemporal data, the grading of GIS visualization thematic charting elements by means of Arc GIS, TGIS annual ring visualization, 3DGIS prototype system, evolutionary algorithm GP, GA as well as GIS. They have made big progress especially on the population information space visualization. Wang Yingjie has concluded China's visualization research of electronic map into four periods. In general, researches that have been done realize the visual expression of data from different aspects and with different methods, but they are static visual expressions based on individual research. They have some limitations when dealing with social economic statistical data accumulated for many years and can hardly achieve data sharing.

According to the analysis above, this paper designs a visualization analysis system to the socio-economic data with attribute, graphics and charts associated style based on Flash technology, in order to realize the net sharing of socio-economic data visualization, and save the time spent by government officials and scientific researchers sorting basic data. The aim is to provide convenience for researchers to publish, get, organize and visually analyze socio-economic data online and make technological foundation of online data analysis for the establishment of GIS research environment.

2.   Features of socio-economic data
Socio-economic data has four features as follow:

(1). Different levels of administrative regions as basic statistical units. The feature makes the organization of socio-economic data formal and the content easy to search.

(2). Integral time sequence. Regularly published socio-economic data every year forms ordered time sequence of the index data.

(3). Multi-style statistical types and indexes. There are not only statistical data, but also sampling survey data, which may change every year.

(4). Appear in the tabular form, monotonous and boring. There's no graphic data so that users cannot form the concept with spatial distribution.

Statistical data has attributed such as statistical index (system), time and space, which can be concluded to three dimensions of statistical data: index dimension, temporal dimension and spatial dimension. Generally, socio-economic data not only shows its normative and historical inheritance, but also has the feature of complexity. The complexity mainly reflects in the annually change of the statistical index, which makes it hard to design uniform table to manage in the process of statistical analysis and poses challenges for the organization and management of the socio-economic data. At the

same time, those features also and the complexity of the information searching. Because of those features, it becomes hard to study and collect effective information.

The application of GIS and other statistical analysis software has made it easier to search and visualize the information of socio-economic data, greatly enhanced the intuition of the data usage, and continuously improved the method of data visualization. However, most of the researches that have been done are based on individual study, which need hand-on operation in the processes of obtaining, inputting, analyzing, conducting data, and cannot realize network sharing. To solve the socio-economic data, this paper designs a visualization analysis system to the socio-economic data with attribute, graphics and charts associated style based on Flash technology through rapid visualization and convenient network sharing. The system will enhance and improve the visualization methods and approaches of socio-economic data, enrich the connotation and expression of digital map, and realize the deep study and network sharing of socio-economic data.

3. Framework design of the visualization analysis system with attribute, graphics and charts associated style

3.1 Technical analysis

According to the features of socio-economic data, the realization of friendly visualization associated analysis, requires designing a good database structure for convenient data retrieval and collection at first, choosing good visualization analysis technologies for multi-style demonstration of information secondly, and confirming technical proposal with powerful network for convenient function expansion, web publishing and network sharing thirdly. Considering the convenience of visualization and fulfillment of the powerful demand of net support, Flash technology is undoubtedly the best choice. Flash itself is a kind of software used to make network animation, which has powerful graphic visualization rendering capabilities and is created for supporting the network. With the flourish of RIA, Flash/Flex RIA technical framework has become the mainstream of the development of internet application program. Especially after the Flash ActionScript3.0 scripting language, support of the network application function has been greatly strengthened. Adobe Flex SDK provides abundant chart control that can be directly used to analyze the attribute data on the two-dimension table, simplifying the workload of Flash program.

In order to give the socio-economic data visualization analysis system with attribute, graphic, chart associated style flexibility and generality, instead of making it only suitable for certain database just like normal visualization analysis systems, this paper integrates database technology with APS database access technology based on the Flash technology, dividing the whole system into four modules, namely administrative unit spatial graphics database, socio-economic database, chart analysis component database and model database. Every module is associated with each other by means of event trigger to achieve timing analysis. The relation of the functional modules is shown in the Fig. 1. below:
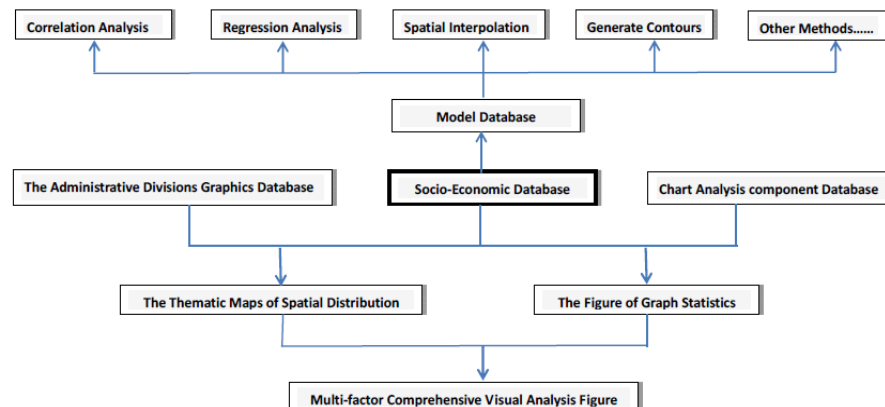
Fig. 1.   Working principle and the relation of the four modules

3.2 Socio-economic database

Socio-economic database is the core of the four modules. Almost all the expression of other modules is achieved directly or indirectly depending on the data from the socio-economic database. According to the features of socio-economic data, and to make the system more suitable for different database structures and different database servers, the socio-economic database designed by this paper adds two tables on the basis of existing database: one is subject partition table index_group, which is used to classify attribution and quickly search by subject; the other is table concordance list table_list_all, which is used to record the metadata information of every attribute list in detail for users' precise search. Following is the structure of table_list_all(Table 1).

Table 1.   Structure design of table_list_all

Name of Child Segment	Type of Child Segment	Meaning of Child Segment
ID	Integer	Current table ID
Table_Name	String	Table name
Stat_level	String	Statistical level of economic data as well as regarding country, province, city and county as statistical levels
Content	String	Main data content of the current table
Content_Type	String	Type of table data sources; statistical data or survey data
Cover_Area	String	Area covered by the data
Data_Source	String	Data sources
Stat_year	String	Statistical year
Table_Structure	String	The table structure of the data, and the detail information of fields

Database access mainly involves data reading and querying, which can be achieved through ASP, JSP and PHP technology. This system adopts the ASP technology. Since Flash possesses XML handling feature pack, this system will output the data read by ASP in XML format,   to provide convenience for ASP to transmit the socio-economic data to Flash, as well as to further associate with administrative unit spatial graphics and carry out relevant chart analysis processing.

3.3 The administrative divisions' graphics database

Flash is a kind of software specializing in vector animation with good spatial graphics support function. It has distinguished advantage in designing vector WebGIS system. The administrative divisions' graphics database mainly comes from the basic geographic data provided by Earth System

Science Sharing Website, including the administrative boundaries of 340 level cities and 2400 counties in 33 provincial administrative units all over the country. The original data is in the format of ESRI Shape Document. In order to realize the expression of spatial graphics in Flash, it is necessary to transform the ESRI Shape document into Flash's SWF document, for network to call and express directly. Therefore, this paper using Flex SDK open source framework provided by Adobe Company (References3) for reference, specially develops a converter that can convert Shape document into SWF. The converter has been uploaded on the relevant website (http://www.gdrsgis.cn/zjqweb/index.asp), for users' free downloading.

In the process of converting data into Flash SWF vector spatial graphics document, the attribute information of the original Shape document will be completely remained. Every graphic spot has two attribute values: area name and area code so that it can be associated with attribute data in the socio-economic database. The purpose of build administrative divisions' spatial graphics database is to realize the spatialization process of socio-economic data and the making of spatial distribution thematic map. Administrative divisions' spatial graphics database will be organized and managed in document format, and every document will be named with the administrative code of the area.

3.4 Chart analysis component database
Chart analysis component database is mainly used to extract data from socio-economic database for chart analysis. The types of chart include histogram, line graph, scatter diagram, bar chart, pie chart, area chart, etc. The open source Flex SDK provided by Adobe has already contained these components, but it still requires program realization combining with Flash Actioscript3.0 to realize the free call and expression of the chart control.

3.5 Model database
The development of model database needs the participation of professional users. To provide more freedom for users, the system only integrates commonly used analytical model, regression model, inverse distance weighting reciprocal interpolation model, isoline generative model into the system as it designs the model library, and only provides model design interfaces for users to design other models and do real-time calculation.
The coordination and association of the four modules above are conducted by virtual control and calculation center written by Flash script program language ActionScript3.0, and the conducting result will be presented to users with the data, chart and map associated pattern.

4.   System implementation and its preliminary practice
According to the framework design above, which adopts the model-view-control center organizational structure in application process, the model and control center are mainly realized by using Actionscript3.0 programming. View is realized by imbedding webpage with Flash SWF objects. SWF view objects will send requests to socio-economic database and spatial graphics database, and demonstrate the result in forms of attribute data, chart and map associated pattern. The whole socio-economic data visualization analytical process is shown in the Fig. 2. below.
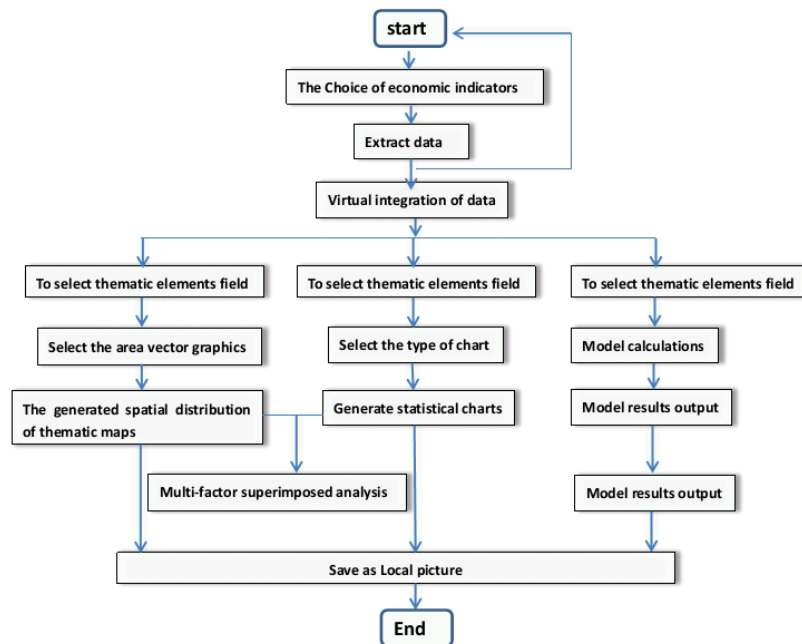
Fig. 2. Socio-economic data visualization analytical flow-process chart

The system adopts the method of configuration files, which means that after finishing the function development, it can fulfill the demand of different application just by modifying the configuration files. System configuration information is stored in XML document format, which is mainly divided into two parts: system initialized configuration part and data demonstration configuration part. The configuration content of XML document is as follow:

```
<?xml version="1.0" encoding="utf-8"?>

<Configuration>

    <filename>province.swf</filename>

    <initRawIndexId>103</initRawIndexId>

    <initDataYear>2007</initDataYear>

    <chartSize>120</chartSize>

    <chartIndex>

        <raw_indexID>103</raw_indexID>

        <raw_indexID>104</raw_indexID>

        <raw_indexID>105</raw_indexID>

    </chartIndex>

    …

</Configuration>
```

The information above has respectively configured the spatial graphics province. swf which present visualization needs to load, initialized loaded economic index code: 103, initialized loaded data year 2007, initialized size of the chart 120 pixels and the displayed element codes of the chart: 103, 103 and 105.

After the system initialization configuration, the system can be published and used online. Now, the visualization analysis system with attribute, graphics and chart associated pattern has been successfully applied to the E-Geosciences project funded by Chinese Academy of Sciences and the project of online visualization to the China regional economic monitoring and evaluation. Fig. 3. is the application rendering showing online data spatial interpolation; Fig. 4. is the demonstrative rendering of the attribute, graphics and chart associated pattern of China regional economic monitoring and evaluation index.
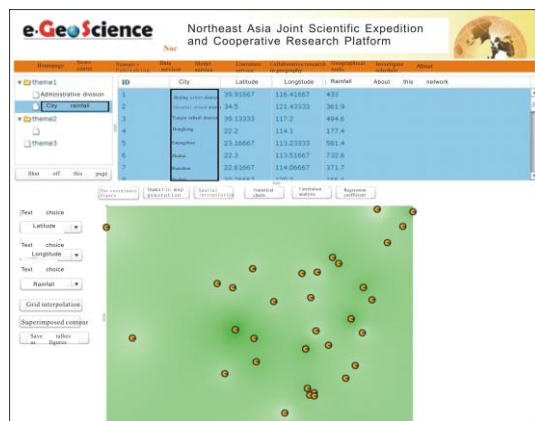

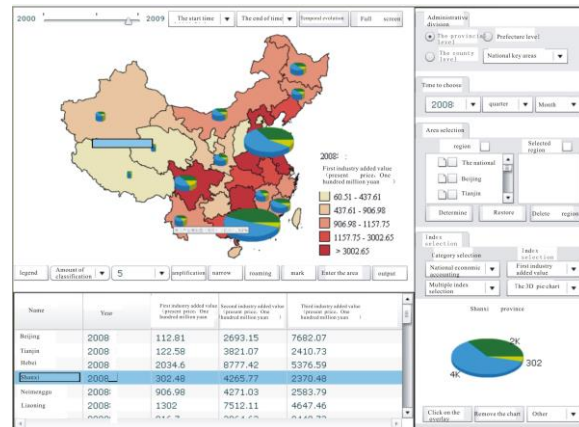Fig. 3. Online data interpolation rendering

Fig. 4.   System visualization interface

5.   Conclusion

The acquiring, storing, spatial analyzing and information inquiring of the geographic information technology has laid the foundation for the imitation, analysis and prediction of the Earth information. To some extents, the visualization of geographic spatial data can be seen as the cartography in the digital age. This paper takes advantage of the Flex SDK open source framework and Flash technology provided by the Adobe company to systematically design a solution to associate socio-economic data, charts and map together, and also applies it to the visualization analysis of the China regional economic monitoring and evaluation, realizing the multi-style expression for the developing indexes in areas of national economy, population, employment, education, science and technology and medicine in every province, area county and key region in China based on the geographic spatial location, as well as   the economic, social development indexes, innovation indexes and regional comprehensive developing indexes, providing intuitive, vivid and easy techniques for the researchers in studying the spatial distribution and trend dynamic changes of China regional development.

This paper's research has achieved: (1) A set of data, shown with multi-style expression, including charts and graphics (2D, 3D pie charts, histograms) and maps which show the regional economic monitoring index and evaluation data. (2) Convenient use and flexible interaction. Users can search and visualize any monitoring and evaluation indexes in any region (by province, area, county, key region, area in the province, etc.) at any time (by year, season, month). Providing flexible interaction for the first time allows users to mark word and special symbols on the maps. (3) Static status display combining with dynamic simulation. Users can not only browse and see the geographic spatial distribution pattern of the monitoring index or evaluation index in pointed region at the certain time, but also make use of timing simulation function to check the dynamic variation tendency of the developing index or evaluation index in certain areas. (4) Simple configuration, primary development for multi-stage application. The system can quickly set up the primary map layer and its display scale, display location, rendering method to spatially express the index, year, the size of histogram or pie chart, comparative index and the key regions of the country or every province. Therefore, no code needs to be changed to meet the applying demand of the country lever, province lever and city lever. It has realized the simple configuration. Primary development has achieved multi-stage application.

Currently, spatial data mining begins to arouse more and more researchers' interest, and has become the hot issue among the international researches. Making use of GIS visualization technology to visualize the statistical data is an important way for statistical data analysis and data mining. This paper creatively designs the visualization analysis system with attribute, graphics, and chars associated pattern for socio-economic data using Flash technology, based on the basic framework of GIS, and puts it into practice and application, to provide users with convenient, quick and direct platform of visualization and net sharing. Problems such as the generation and download of the visualization charts with high resolution, the integration and graphic expression of the statistical data measured by city and county, still need to be further studied.

Reference:
[1]   GENG Sheng-ling. Research of statistic information data mining based on complex network[J]. *Journal of Liaoning Normal University*(Natural Science Edition), vol. 32, pp. 64-66, 2009.
[2]   GAO Jun. Visualization in Geo-Spatial Data [J]. *Engineering of Surveying and Mapping*，vol. 9, pp. 1-7, 2000.
[3]   CHEN Mian, LIU Xiao-mei, LI Guo-yan.. Analysis of spatial visualization methods for statistical data[J]. *Science of Surveying and Mapping*，vol. 32, pp. 65-68, 2007.
[4]   DAI Changda, JIANG Xiaoguang, TANG Lingli. The application processing and analysis of RS Image Beijing: Tsinghua university, 2004.
[5]   RUI Xiaoping1; ZHANG Yanmin2; YANG Chongjun1. An Overview on Study of Key Technologies in Visualization of Geographical Information [J]. *Computer Engineering*，vol. 29, pp. 1-5, 2003.
[6]   XIA Hui-qiong, LI De-ren, SHAO Zheng-feng, ZHENG Chun-yan. A static visualization representation method for spatio-temporal data. , *Science of Surveying and Mapping*，vol. 33, pp. 116-119, 2008.
[7]   LIU Yang; LI Xin. Research and application of spatial data visualization in 3DGIS [J]. *Computer Engineering and Design*，vol. 27, pp. 1090-1093, 2006.
[8]   LIAO Yilan, WANG Jinfeng, MENG Bin, LI Xinhu. A Method of Spatialization of Statistical Population [J]. *Acta Geographica Sinica*，vol. 62, pp. 1110-1119, 2007.
[9]   Dang Anrong. STUDY ON THEMATIC MAPPING ELEMENT CLASSIFICATION OF GIS VISUALIZATION [J]. *Acta Geographica Sinica*，vol. 53, pp. 7-10, 1998.
[10] SU Ying; WANG Ying-jie; YU Zhuo-yuan; TAN Yu-qi. The design and research on geo-visualization system for population [J]. *Science of Surveying and Mapping*，vol. 30, pp. 38-40, 2005.
[11] WANG Yingjie, SU Ying,CHEN Xiaogang,YU Zhuoyuan,MENG Liqiu. Visualization issues in the development of electronic atlas in China. *Journal of Geographical Sciences*. vol. 15, pp. 87-96, 2005.