CHINESE GEOGRAPHICAL SCIENCE
Volume 12, Number 3, pp. 233 – 237, 2002
Science Press, Beijing, China

THE STRATEGIES OF INTEGRATING SPATIAL DATA ANALYSIS AND GIS

MAO Zheng-yuan[1,2], LI Lin[1]
(1. School of Resource & Environmental Engineering, Wuhan University, Wuhan 430070, P. R. China;
2. Department of Geography, Central China Normal University, Wuhan 430079, P. R. China)

ABSTRACT: The gap between SDA (Spatial Data Analysis) and GIS (Geographical Information Systems) existed for a long time. Presently this problem still remains in spite of a lot of theoretical and practical studies which try to find the solution for it. The research background and current situation about how to integrate SDA and GIS are introduced at first. The main idea of this article is to make sure what is the best scheme to bridge the gap between SDA and GIS and how to design it. There are a lot of factors to influence the standards to assess such a scheme, for instance, the attitude of users and GIS developers, the framework and related functions of current available GIS software in the market and so on. But the two most important ones of them are efficiency and flexibility of the scheme itself. Efficiency can be measured by the convenient extent and temporal length when it is used for carrying out SDA. Flexibility means users can define their own SDA methods. The best integration scheme should satisfy the two standards at the same time. A group of functions, which can be combined to implement any SDA method, are defined in order to design such an integration scheme. The functions are divided into five classes according to their properties.

KEY WORDS: GIS; Spatial Data Analysis; integrating strategy; open GIS framework

CLC number: P208, P91 Document code: A Article ID: 1002-0063(2002)03-0233-05

1 INTRODUCTION

Geographical Information Science developed rapidly during the several past decades. Its components include Geographical Information System(GIS), Remote Sensing(RS), Visualization and Spatial Data Analysis (SDA). There is considerable interest in integrating GIS with other three methods to meet the requirements of advanced applications. Since the integration among GIS, RS and Visualization is almost a fact, in the last decade more and more attention from the community of GIS and geographers focused on the integration of SDA and GIS.

SDA is a set of technique devised to support a spatial perspective on data (GOODCHILD, 1987). More specifically it is concerned with the spacial process, through which observational data are available, and methods that are sought to describe or explain the behavior of this process and its possible relationship to other spatial phenomena. The object of analysis is to increase our basic understanding of the process, assess the evidence in favor of various hypotheses concerning it, or possibly predict values in areas where observations have not been made (BAILEY and GATRELL, 1995). The main purpose of SDA is to explor and explain spatial patterns. Its content consists of Spatial Statistical Analysis and Spatial Autocorrelation Analysis. In some cases, SDA is abbreviated as SA (Spatial Analysis), which often results in confusion because SA is also used to name an important category of functions within GIS such as surface partitioning, map overlay, buffering, network analysis, and location-allocation analysis. Actually the history of SDA is much earlier than that of GIS (LIVINGSTON, 1992; BARNES, 2001). But SDA is very inconvenient without the supports of GIS, so with the rapid development of both, more and more geographers recognize that it is desirable to integrate both of them. On the one hand, GIS can provide various data sources for SDA, store and visualize the analytical results. On the other hand, GIS needs SDA if it is to reach the potential implied by many of its definers and proponents (GOODCHILD *et*

Received date: 2002-05-17
Biography: MAO Zheng-yuan(1964 –), male, a native of Hunan Province, Ph. D. candidate at School of Resource & Environmental Engineering, Wuhan University, associate professor. His research interests focus on GIS and spatial analysis.

al., 1992).

Although much progress has been made in the last decade in improving the spatial data analytical capabilities of Geographical Information Systems, the problem of integrating SDA and GIS still remains because the following two reasons. First, all the integration schemes published in academic journals were only compromise of the present GIS and statistical software and didn't refer to the feasibility of how to extend the schemes and thus keep them efficient when new GIS software or SDA methods appear. Second, they avoided discussing the contradiction of GIS and statistical software in system framework and data model, which will cause serious limitation in some application cases with respect to their integration.

The purpose of this article is to offer a strategy to realize a full integration of SDA and GIS. The main idea of the strategy is to create an open GIS framework. In the framework, only the most frequently used SDA methods are implemented and what are provided are a group of SDA functions. So the users can enjoy the freedom to develop their own more complicated SDA functions. The article will not get involved with too much of how to implement the current available SDA methods since there has been a great deal of discussion about them.

2 CURRENT AVAILABLE SCHEMES TO INTEGRATE SDA AND GIS

2. 1 Advantages and Disadvantages of Different Integration Schemes

Following GOODCHILD *et al.* (1992), four strategies can be identified: 1) stand-alone Spatial Data Analysis software, 2) loose coupling of existing GIS software with statistical software, 3) close coupling of GIS software with statistical software, and 4) full integration of Spatial Data Analysis in a GIS. Advantages and disadvantages are summarized in Table 1 (Please pay attention to the conception "spatial analysis" in the table, here it is the abbreviated form of "spatial data analysis").

2. 2 Two Practical Schemes to Integrate SDA and GIS

UNGERER *et al.* (2002) put forth a coupling strategy which was based on Component Object Model (COM) components, and is implemented using one module of code which uses the software components of both the analysis software and the GIS. The code module programmatically opens the analysis software package, transfers the data from the GIS, performs the analysis, transfers the data back to the GIS and finally closes the analysis software package.

ZHANG *et al.* (2000) put forward another practical instance to integrate SDA and GIS. Their integration strategy is embedding a limited set of GIS components into a software system that can be easily extended to incorporate spatial data analysis functions. A module with the functions mentioned above is designed and implementated by using Microsoft Access, a proprietary Database Management System (DBMS).

2. 3 The Current Research Situation in China

There are large differences between China and western countries in terms of the adoption of conceptions, theoretical and practical research relateing to the topic in this article in spite of not being devoid of communication. In English literature, SDA is a clearly defined conception that includes Spatial Statistical Analysis and Spatial Autocorrelation Analysis. SA has three different definitions, the first one defines SA as the abbreviation of SDA; the second one means a kind of data processing methods in GIS which extract spatial information from spatial location data (to avoid confusion, this kind of methods are also named "Spatial Data Manipulation" in English literature); the third one can be expressed as the quantitative studies of phenomena that are located in space, which incorporates the above two ones and Spatial Modeling(BAILEY and GATRELL, 1995). Among the three mentioned definitions of SA, the meaning of the third definition is too broad and thus has only been used in very little situations. The first one is mainly adopted in the geographical academic circle, and the second one is mainly in all kinds of literatures relating to GIS software. To avoid confusion, more and more people use "SDA" instead of "SA" to express the first meaning, and use "SA within GIS" for the second one since both of them are frequently used. In English literature, the articles on the topic of "integrating SDA and GIS" are too many to be listed here one by one (ABEL and KILBY, 1994; BOOT 2000; GETIS, 2000; GOODCHILD, 2000; MARBLE, 2000). In the contrasts, hardly can we find any Chinese literature on the just mentioned topics because of the following two reasons. First, Chinese people know SA from literatures relating to GIS. For them, SA has only one meaning, i. e. the method to extract information from spatial location data, the second definition in this paragraph. So SA is the part of GIS, and

Table 1 Coupling strategies for linking Spatial Data Analysis and GIS (adapted from GOODCHILD *et al.* 1992).

	Isolated	Loose	Close	Integrated
Description	Analysis and output display directly in spatial analysis software	Analysis in spatial analysis software, output display in GIS, facilitated by online file database exchange	Analysis method varies; GIS and analysis package share a common	Analysis and output display directly within GIS
Advantages		Less work in terms of code creation	Spatial analysis can be done within the GIS environment	No file import or export, no code creation required
Disadvantages	Abundant GIS data layers cannot be used	Consuming time to import and export data	Much work in terms of code creation	Possible lack of specialist insight in spatial analysis

the relationship between GIS and SA is just like the whole thing and a part of it, of course there is no integration problem between them. Second, SDA is still not accepted as a technical term in China, its meaning is the same as SA if translated literally. So the idea of integrating SDA and GIS doesn't exist in China either. A part of SDA—Spatial Statistical Analysis appears in some Chinese literatures that have nothing to do with GIS(XU, 1996). In fact, there is no GIS software platform developed independently by Chinese companies incorporating the implementation of SDA. For example, MAPGIS 6.0, the latest version of Chinese top one GIS software, only includes four kinds of function implementations, i.e. buffering, overlay, DTM analysis, and network analysis within its SA module. In China, the research on SA is always related to GIS and concentrates on the design and implementation of algorithms of SA(GUO, 1997), and the researches about "integration" are often involved in "integrating '3S' (RS, GIS and GPS)" (LI and GUAN, 2000).

3 THE STRATEGY OF FULLY INTEGRATING SDA AND GIS

The advantages and disadvantages of different integrating strategies have been mentioned in the above context. Developing stand-alone packages, though a sensible choice in some situations, is generally not considered a good strategy because it seems a waste not to exploit either the powerful spatial data handling, displaying, querying and analysis capabilities of current GIS software, or the sophisticated statistical data analysis routines available in statistical software. Loose coupling approaches leave GIS and Spatial Data Analysis as two separate entities with an ability to transfer files in a common format between them. This is probably the oldest method that has been used in practice. It is feasible but very inefficient (ZHANG and GRIFFITH, 2000). Close coupling scheme is involved in a lot of extra programming, and that's really difficult for most users. By comparing and contrasting the advantages and disadvantages of the four scenarios mentioned in the previous context, it is evident that the last scheme, full integration of SDA in a GIS is the best one from the theoretical perspective. This point of view is supported by the following reasons: 1) GIS is the most efficient tool to manipulate (including input, retrieve, manage, visualize, display and so on) spatial data up to date; 2) usually there is a database with abundant spatial data within GIS; 3) other components of geosciences such as RS and Visualization have been integrated with GIS; and 4) the design of a GIS software and the accumulation of its data are very expensive and time-consuming, thus it is not sensible to repeat such a challenging job again and again only for implementing a relative simple Spatial Data Analysis or for making full use of an existent statistical software and avoiding involving in any change of the present commercial GIS. In spite of this, many articles still try to find other inefficient alternatives because they anticipate that GIS developer will not embed SDA functions into their products in the near future.

Perhaps it is true that GIS developers will pay no attention to the mentioned integration, but this should not be the excuse for the academic circle not to try that way. We believe that GIS developers will adopt the full integration scheme sooner or later since it is the permanent way to bridge the gap between SDA and GIS. The best way to perform the full integration is to design an open GIS framework. New modules will be allowed to insert that framework. It should also include an interface, which consists of a group of functions, to offer the chance for users to finish their own specific task that hasn't been implemented in the system. Since the interests of each kind of users narrowly focus on some specific phenomena or processes, it's hard to estimate and satisfy so various requirements from users with heterogeneous background and tasks. It's also unnecessary to do so.

In terms of SDA, the interface is a group of functions. Each of them provides a very basic manipulation to the specific data such as retrieve, store, calculation, etc. Please pay attention to the fact — what we really need to do is not too much since a lot of required functions have been available in the GIS software. A series of new functions will form through combining the basic functions with various orders and styles. All the basic functions that are necessary to perform almost all the present existent SDA methods are listed in Table 2.

Most of the functions have several formats that are similar to the polymorphism of methods in JAVA or C + + language. The relationship between the data type of the parameters and the functions has been automatically constructed according to the definitions of functions. I am not going to describe the specific form of every function in the table. The users don't need to worry about the details of the function in the practice since they will be under the elaborated instruction of a visualized interface integrated with other modules. The interface will lead the users to finish the selection of the proper data type for each parameter. So the users can focus on the design of their own Spatial Data Analysis instead of the technical implementation of their ideas.

Let's take the implementation of the F function (it can quantitatively measure the proximity of events) as an instance. It refers to six basic functions in table. The following are the descriptions of their formats and corresponding parameters at first. Then the procedure of implementing F function by combining the six basic functions in a pseudo-language is demonstrated.

Table 2 Functions to perform SDA in the full integration scheme

Types	Arithmetic	Logical	Geometric	Matrix	Comprehensive
Function Names	Max Min Mean Product Sum Deviation	Not And Or Boolean	Distance Length Area Relation Count Random Get Interpolation	Addition Subtraction Multiplication Inversion Transposition	Read Recur Process Manipulate Visualize Memorize

(1) Count

Format1: Count(filename, object type, logical expression);

Format2: Count(filename, object type, region, logical expression);

Explanation: Return the counts of the objects, which satisfy the logical expression in the file (Format1) or in the defined area of the file (Format2).

(2) Distance

Format1: Distance(filename, array name);

Format2: Distance(filename, region, array name);

Format3: Distance(filename, point identity, array name);

Format4: Distance(filename, region, point identity, array name);

Explanation: Orderly return the distances between each two points in the file(Format1) or in the defined area of the file(Format2), or orderly return the distance between the specific point and each of other points in the file (Format3) or in the defined area of the file (Format4) with the array.

(3) Get

Format1: Get(filename, arrayname1, arrayname2);

Format2: Get(filename, region, arrayname1, arrayname2);

Format3: Get(filename, point identify, variablename1, variablename2);

Explanation: Return the number of points in the file (Format1) or in the identified region of the file (Format2) and keep the coordinates of the points with two arrays; or return the coordinate of the identified point with two variables (Format3).

(4) Min

Format1: Min(expression1, expression2[, ..., expressionN]);

Format2: Min(arrayname, integer1[, integer2, integer3, ..., integerN]);

Explanation: return the smallest value in the expressions or the array members.

(5) Sum

Format1: Sum(expression1, expression2[, ..., expressionN]);

Format2: Sum(array name, integer1[, integer2, integer3, ..., integerN]);

Explanation: Rreturn the total sum in the expressions or the array members.

(6) Visualize

Format: Visualize(type, datasets);

Explanation: Visualizing the result of a process or manipulation.

```
Begin
n = Get(filename, array1, array2);
For (i = 0; i < n; + + i)
{
    Distance(filename, p[i](array1[i], array2[i]),
    array3);
    array[i]Count(filename, point, Min(array3,
    (n - 1)) < d);
}
F(d) = Sum(array, n)/Count(filename, point);
Visualize (type, d, F(d));
end.
```

4 CONCLUSIONS

(1) The problem of integrating SDA and GIS still remains in spite of many attempts.

(2) The full integration is the best one in the four strategies put forward by GOODCHILD *et al.* in 1992.

(3) There are two schemes to finish the full integration. The first one tries to implement each current available Spatial Data Analysis method and updates regularly so as to keep pace with the development of SDA. The second one is only to implement the most frequented used Spatial Data Analysis methods and meanwhile to provide a group of functions, each of which finishes a basic manipulation to a specific data type. Of course, resorting to that group functions users can explore the SDA methods they are interested in.

(4) The advantages of the second scheme is that it offers a possibility for the users to design and implement their own needed SDA methods, and sometimes this is very important because the GIS developers can't anticipate the various requirement from users with diverse background in terms of interests, application cases and purposes.

REFERENCES

ABEL D J, KILBY P J, 1994. The system integration problem [J]. *International Journal of Geographical Information Systems*, 8(1): 1 - 12.

BAILEY T C, GATRELL A C, 1995. *Interactive Spatial Data Analysis*[M]. Edinburgh (UK): Pearson Education Longman, 1 - 413.

BARNES T J, 2001. Retheorizing economic geography: from the quantitative revolution to the "culture turn" [J]. *Annals of the Associate of American Geographers*, 91(3): 546 - 565.

BOOT B, 2000. Using GIS to promote spatial analysis[J]. *Journal of Geographical Systems*, 2(1): 17 - 21.

GETIS A, 2000. Spatial analysis and GIS: an introduction[J]. *Journal of Geographical Systems*, 2(1): 1 - 3.

GOODCHILD M F, 1987. A spatial analytical perspective on geographical information systems [J]. *International Journal of Geographical Information Systems*, 1(4): 327 - 334.

GOODCHILD M F, HAINING R P, WISE S M, 1992. Integrating GIS and spatial data analysis: problems and possibilities [J]. *International Journal of Geographical Information Systems*, 6 (5): 407 - 423.

GOODCHILD M F, 2000. The current status of GIS and spatial analysis [J]. *Journal of Geographical Systems*, 2(1): 5 - 10.

GUO Ren-zhong, 1997. *Spatial Analysis*[M]. Wuhan: Wuhan University of Science and Technology of Geodesy, Photogrammetry and Cartography Press, 1 - 236. (in Chinese)

LI De-ren, GUAN Ze-qun, 2000. *Integration and Implementation of Spatial Information Systems*[M]. Wuhan: Wuhan University of Science and Technology of Geodesy, Photogrammetry and Cartography Press, 1 - 244. (in Chinese)

LIVINGSTONE D N, 1992. *The Geographic Tradition: Episodes in the History of a Contested Enterprise*[M]. Oxford (UK): Blackwell.

MARBLE D F, 2000. Some thoughts on the integration of spatial analysis and Geographic Information Systems[J]. *Journal of Geographical Systems*, 2(1): 31 - 35.

UNGERER M J, GOODCHILD M F, 2002. Integrating spatial data analysis and GIS: a new implementation using the Component Object Model(COM)[J]. *International Journal of Geographical Information Science*, 16(1): 41 - 53.

XU Jian-hua, 1996. *Mathematical Approaches in Modern Geography* [M]. Beijing: Advanced Education Press, 1 - 278. (in Chinese)

ZHANG Z and GRIFFITH D, 2000. Integrating GIS components and spatial statistical analysis in DBMSs[J]. *International Journal of Geographical Information Science*, 14(6): 543 - 566.