Development of a Web-based Spatial Decision Support System for Business Location Choice in Taipei City

Chin – Te, Jung
Ph.D. student,
Dept. of Geography,
National Taiwan University
Taipei, Taiwan
d94228001@ntu.edu.tw

Chih – Hong Sun
Professor,
Dept. of Geography,
National Taiwan University
Taipei, Taiwan
chsun@ntu.edu.tw

Abstract

Motivated by the increasing accessibility of technology, more and more spatial data are being made digitally available. How to extract the valuable knowledge from these large (spatial) databases is becoming increasingly important to businesses as well. It is essential to be able to analyze and utilize these large datasets, convert them into useful knowledge, and transmit them through GIS-enabled instruments and the Internet so that the key information can be conveyed to business decision-makers effectively. In this research, we built a GIServices website, called *Location Analysis of Business Decision Support System in Taipei city (LABDSSiT)*, for the location evaluation of convenience stores (CVS) in Taipei city, Taiwan to achieve the following goals: (1) integrate databases from spatial and non-spatial datasets about the locations of businesses in Taipei; (2) combine GIS and SDM (spatial data mining) analysis to extract the knowledge from the integrated databases; and (3) develop a interactive interface for users who want to evaluate the suitability of a potential CVS location.

Keywords: Spatial Data Mining, GIS, Spatial Decision Support System, Location Analysis

There is no more there, everywhere is here.

Negroponte, N. *(1995) Being Digital*

1. Introduction

Users of geographic technology are both blessed and cursed by the growing availability of geospatial data, software, hardware, and services (Goodchild *et al.*, 2005). Due to the technology, more and more spatial data are being made digitally available (e.g. National Aeronautics and Space Administration or NASA can receive digital satellite images in 500 gigabytes per hour.) We can use such a lot of spatial data to research and understand a remote area, even not to put foot on it. This is what Negroponte (1995) meant when he said "*There is no more there, everywhere is here.*"

In the era of explosive geospatial data and information, it has become more difficult to extract the useful spatial knowledge from the large spatial databases. However, geography knowledge discovery (GKD) or spatial data mining (SDM) can intelligently and automatically analyze and utilize the large geospatial databases and transform geospatial data into useful geospatial information, which can further be combined with geographic information systems (GIS) (Buttenfield *et al.*, 2000; Miller and Han, 2001; Mennis and Liu, 2003).

After we extract some key spatial knowledge, how can we then distribute the spatial knowledge more efficiently so that Web users can use the knowledge? Using the Internet, we can achieve this goal more easily. Popular since 1990, the Internet has changed the field of GIS from the desktop system to web GIS services (Peng and Tsou, 2003). GISerivces refers to the service-oriented approach that allows users access, assemble and rent geoprocessing components that are distributed across a network via standard Internet browser (Tao, 2001).

In Taipei City, the business of convenience stores (CVS) is very competitive. The number of CVS is over 1000 in Taipei City, including 7-ELEVEN, OK, Family and others. All CVS want to occupy the suitable location to make more profit, because the key factor of CVS is the location (Nelson, 1968; Applebaum, 1968; Grune, 1987). However, there is no efficient way to find the suitable location using the large spatial database. In this research, we built a GIServices, called *Location Analysis of Business Decision Support System in Taipei City (LABDSSiT),* which used the technologies of ASP.NET (Microsoft, 2003) and ArcGIS Server (ESRI, 2003) to solve the location problem of CVS in real time on the Internet. *LABDSSiT* is designed to achieve the following three goals: (1) integrate databases from spatial and non-spatial datasets about the locations of businesses in Taipei, Taiwan; (2) combine the result of SDM analysis of extracting the knowledge from the integrated databases; and (3) develop a interactive interface for users who are looking for potential CVS locations and want to evaluate the suitability of the locations.

The remainder of this paper is structured as follows. In section 2 we describe the concept of GIServices and the result from our previous work used SDM to analyze the spatial databases. Section 3 discusses the development of *LABDSSiT*. In section 4 we present a conclusion and directions for future research.

2. Literature Review

2.1. GIServices

The development of GIS has been greatly influenced by Information Technology (IT) which changes GIS from mainframe GISystem to Desktop GISystem to Internet GIServices. Internet GIServices are network-based geographic information services that utilize both wired and wireless Internet to access geographic information, spatial analytical tools and other GIS resources (Peng and Tsou, 2003). For the desktop GISystem, users have to purchase the whole GIS package to solve the simple spatial tasks. It costs a lot and is not efficient for the users. If users just want the simple spatial function like “buffer”, why can we just provide the spatial functions on the Internet, where users can choose and combine them for their need? Therefore, GIServices have emerged using the service-oriented approach that allows users access, assemble and rent geoprocessing components that are distributed across a network via standard Internet browser (Tao, 2003). So, GIServices include the services of geodata and geoprocessing for users to access the geodata remotely and combine the geoprocessing components on the internet to do spatial tasks.

2.2. Spatial Data Mining in the Location Analysis of CVS in Taipei City

With the advance of remote and survey technologies, more and more geospatial data have been collected everyday. GIS researchers have urgent need for new methods and tools that can intelligently and automatically transform geographic data into information and, furthermore, synthesize geographic knowledge (Buttenfield, *et al.*). Spatial data mining (SDM) is a new method meted for GIS researchers to extract interesting, implicit, and useful spatial patterns from large spatial databases (Roddick and Spilioporlou, 1999; Shekhar and Chawla, 2003).

In our previous work, we use the spatial association rule (Han, 1995) and the apriori algorithm (Agrawal *et al.*, 1994) to extract interesting and useful rules for location knowledge of CVS from large spatial and census databases in Taipei city (Jung, 2005). For example, a rule "Distance to schools < 300m and Distance to banks < 400m → CVS (10.5%, 92.9%)" means that the statement that a CVS is located within 300m of schools and within 400m of banks is supported by 92.9% and 10.5% records of entire databases. We can find that some variables in every rule are repeated and consider these variables are more important for the location issue of CVS. Therefore, these variables would be re-organized in advance to form the location indexes of CVS (Table1) and validate these indexes with results of Kuo et al. (2002).

Table 1: the location indexes for CVS

Census Index	Spatial Index
The number of people in male > 1900	The distance to stations of Mass Rapid Transit (MRT) < 600m
The number of people in female > 2000	The distance to Parks < 400m
The number of people in age0~6 > 400	The distance to Schools < 300m
The number of people in age6~25 >1200	The distance to Foods < 900m
The number of people in age25~65 > 2100	The distance to Main Roads < 10m
The number of people in single > 300	The distance to Banks < 400m

The number of people in industry > 2000	The distance to Hotels < 600m
The number of people in business > 2000	The number of competitions < 2
The number of people in service > 2000	

These location indexes would form the knowledge databases for location of CVS, which can then be used in the GIServices for location analysis of CVS in Taipei City and for validating the suitability of locations chosen by users.

3. A GIService for Location analysis of CVS in Taipei

In order to achieve the real-time location analysis of CVS in Taipei City on the Internet, we have implemented a GIService to achieve this task. We have assembled the GIService by adopting the technologies of ASP.NET (Microsoft, 2003) and ArcGIS Server (ESRI, 2003) that offer the geoprocessing components for the service.

3.1. The Framework

We follow the client-server architecture to implement the GIServices (Figure 1), called *Location Analysis of Business Decision Support System in Taipei city (LABDSSiT)*. Users can connect to the system via an internet browser and click points on the map to be the potential locations of CVS. Next, the server would focus on the points to evaluate whether the points are suitable for CVS or not by the following steps: spatial join, buffer, intersect with building area, area interpolation and match up with previous CVS knowledge extracted from spatial data mining. These steps would be processed in concert with the geodatabases to temporally save the result or to get the spatial layer in real time. The following section discusses further about these steps. After system finishes the evaluation, the system would present the result map to the user and provide the suggestion about the points.

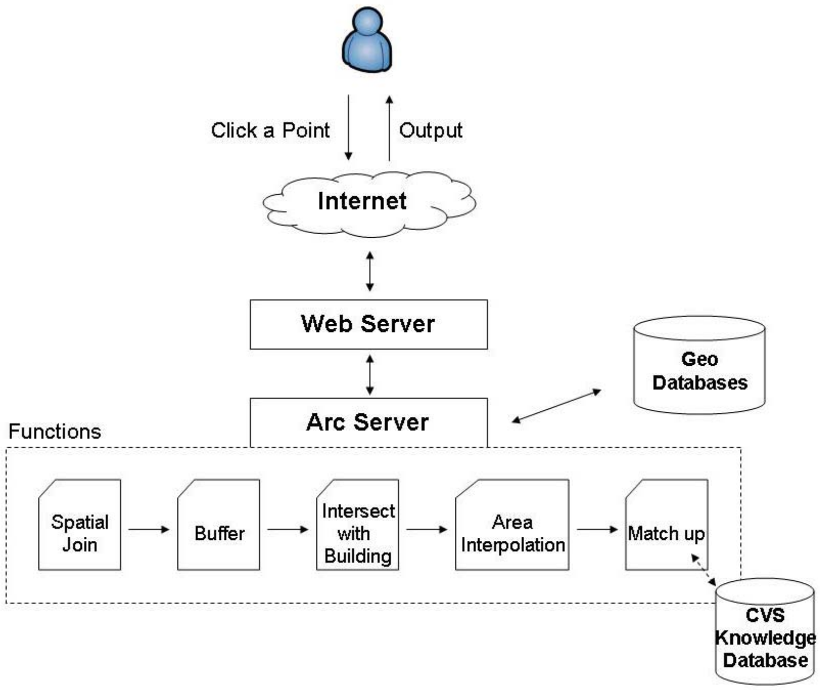

Figure 1: the architecture of LABDSSiT

3.2. The Spatial Functions

3.2.1. Spatial Join

Because the spatial relationship is implicit in the spatial data, we have to extract spatial relationships first from the spatial databases and record them in the attribute table. At this step, we just extract the distance, one of the spatial relationships, from the point users want to set a convenience store to the selected spatial features (e.g. roads, parks, schools and other features) and record distances to these features in the attribute table of point.

3.2.2. Buffer

Trade areas are needed in every business to access the properties of consumers and provide the new strategy to attract consumers to convenience stores. However, setting trade areas depends on the kind and the place of business (e.g. the trade area of a convenience store is different from a bank; the trade area of the convenience stores is different in Taiwan and the United States). Therefore, we follow previous studies in setting the trade area of a convenience store in Taiwan and set up the trade area with a radius of 250 meters to get the attributes of consumers (e.g. age, sex, and education).

3.2.3. Intersect with Building Area

In order to derive the properties of consumers, we have to overlay the trade area with the census data in Taipei City. However, the census data are aggregated by "census tract" in Taipei City (Figure 2) and people do not live evenly within a census tract. So, we first hypothesize people just live in the building area, one of the landuse types (e.g. the area of a1 + a2 in Figure 3), and will not live in other landuse types (e.g. parks, business, and river areas). We intersect the buffer and census tract layers to get where the building areas are for every feature in buffer layer and calculate the proportion of the building area using area interpolation (Figure 3).

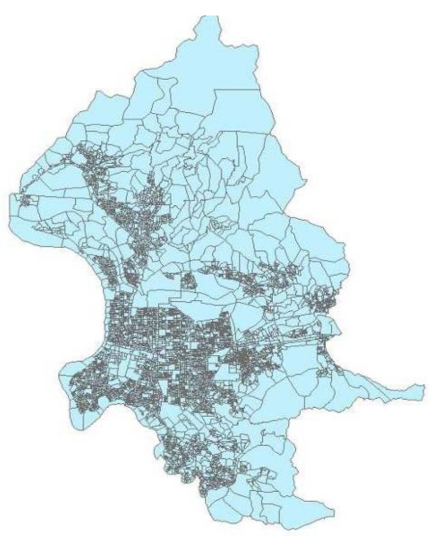

Figure 2: Census tract in Taipei city

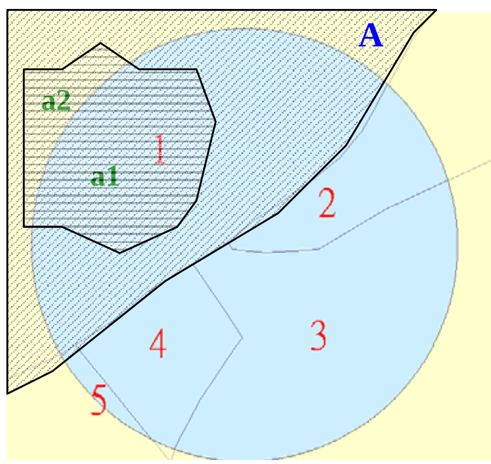

Figure 3: Intersect with Building area

zone 1,2,3,4,5: the small factors in a buffer of CVS; ***Area a1***: the building area intersected in zone 1; ***Area a2***: the building area not intersected in zone 1.

Zone A: a census tract.

3.2.4. Area Interpolation

The result from overlay shows the difference between boundaries of the two input layers. In order to summarize the census data in every buffer zone, we use the proportions of the building area in the census tract and buffer layer to be the weights for area interpolation (Figure 3; Formula 1) and summarize the census data for every feature (Table 2).

$$P_1 = (BA_1 / BA_A) * P_A \qquad \textbf{(Formula 1)}$$

P_1: The population of zone ’1’

BA_1: the building area of zone ’1’ (a1)

BA_A: the building area in census tract’A’ (a1+a2)

P_A: the population of census tract ‘A’

Table 2: Summarized census and spatial table

Buffer ID	CVS Name	Distance to Police Office	…	Male	Female	Sum People	…
0	7-11	573.3341	…	1288	1364	2652	…
1	7-11	386.1735	…	2895	3268	6163	…
2	Hi-Life	529.649	…	4223	4538	8761	…
3	Hi-Life	152.3975	…	3492	4041	7533	…
4	7-11	82.22532	…	137	63	200	…
5	7-11	160.7047	…	2062	2271	4334	…
6	Family	1086.15	…	3437	3920	7358	…
7	7-11	1424.382	…	1029	972	2001	…
8	OK	466.1352	…	1067	1022	2089	…
…	…	…	…	…	…	…	…

Table 2 describes every attribute in the buffer layer which has been summarized after area interpolation (e.g. Buffer ID 0 presents a “7-11” CVS: its distance to Police Office is about 570 meters and the number of people is about 2650 within its 250 meter trade area,.

3.2.5. Match Up

After we summarize the census data and the distances to the select spatial features to be the index of the convenient store, we have to match up with the previous results from spatial data mining (Table 1) to see which index is suitable for the location chosen by the user.

3.3. A Demo of Location Analysis of CVS in Taipei

Figure 4 shows a demo of LABDSSiT which is implemented by ASP.NET and ArcServer software and divided in four parts:

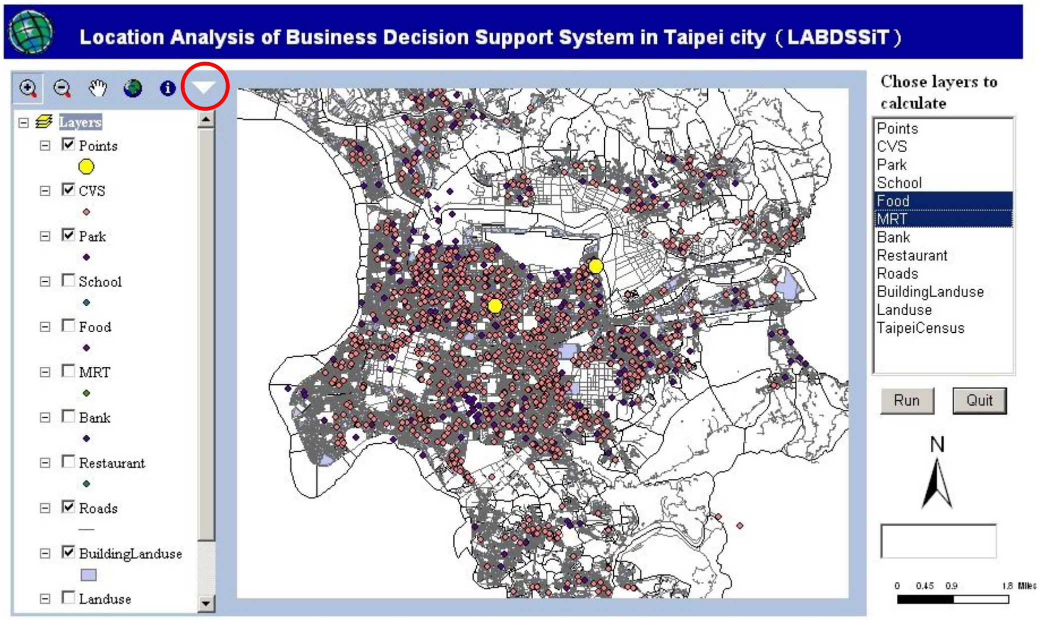

Figure 4: a demo of LABDSSiT

(1) Map: It is the major part of GIServices to show the result and to interact with the user in the spatial view.

(2) Toolbox: we offer the some simple spatial functions (e.g. zoom in, full extent) and customize a function which is circled on Figure 4. The customized function helps the user click on the map to generate points to be the potential locations of CVS and use these points to evaluate if these locations are suitable or not.

(3) Table of Contents (TOCs): It provides users to decide which spatial layer is shown or not.

(4) Select Spatial Layer: It is the main part of LABDSSiT, which allows the user to select which spatial layer they want to analyze and to calculate the distance from the chosen location to the selected spatial layers. After selection, the user can click the button of "Run" to run the system and get the suitability analysis of the chosen location (Figure 5).

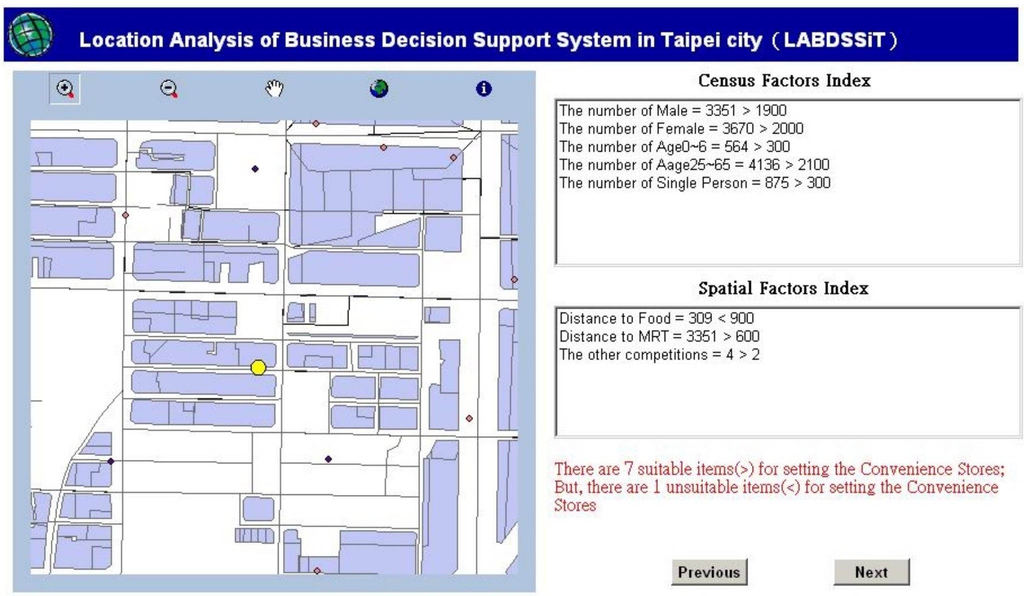

Figure 5: the result of location analysis

Figure 5 shows the results of location analysis of every point. The evaluation is divided into two parts:

(1) Census Factor Index: if the number of people is more, the CVS can attract more people to shop. Therefore, if the number of every census factor is larger than the number of previous index from spatial data mining, we consider the location suitable for CVS.

(2) Spatial Factor Index: if the distance to every spatial layer which users select is less, the CVS can attract more people by being adjacent to these spatial features. Therefore, if the number of every spatial factor is less than the number of previous index from spatial data mining, we consider the location suitable for CVS.

However, we have to pay more attention on the competition of other CVS. If there is more competition from other CVS, consumers will shop in other CVS. Accordingly, if the number of competition of other CVS is less than the number of previous index from spatial data mining, we consider the location suitable for CVS.

Finally, we summarize the number of suitable indexes and not suitable indexes to inform the user about the suitability of the CVS location. Additionally, we can click the button of "Next" or "Previous" to see the condition of every point the user has chosen.

4. Conclusions & Future work

In the era of explosive geospatial data and information, we should use more efficient ways to extract the information from the large datasets and use this information to solve more complex problems or provide more accurate support for the decision makers. Additionally, GIS plays a key role to cooperate, analyze large spatial datasets and represent spatial information on a spatial platform. GIS from system to service would make more and more people understand what GIS is and use GIS to solve the daily spatial tasks. In this research, we use association rules, one of the methods in spatial data mining, to extract the location knowledge of CVS and implement a GIServices by ArcServer, to help users choose potential locations for CVS on the map and evaluate these locations with previous results from spatial data mining to see whether these locations are suitable or not.

In the LABDSSiT GIService, we validate and consult the result with experts to see the location index or the GIService is correct and suitable for analysis for location issues of CVS. If it works, we will implement more GIServices for locations of businesses (e.g. banks, hotels, gas stations and *et al.*) to enforce the importance of GIS and make GISerivces more powerful and useful for everyone to solve the spatial tasks.

References

Agrawal, R. and Srikant, R. 1994. Fast Algorithm for Mining Association Rules, *In Proceedings of the Twentieth VLDB Conference*：487-499

Applebaum, W. (1968) *Guideline to Store location research-with emphasis on supermarket*, Massachusetts: Addison-Wesley.

Buttenfield, B., Gahegan, M., Miller, H. and Yuan, M. (2000) Geospatial Data Mining and Knowledge Discovery, *UCGIS White Paper on Emergent Research Themes*, http://www.ucgis.org/priorities/research/research_white/2000%20Papers/emerging/gkd.pdf

ESRI (2003) ArcGIS Server, http://www.esri.com/software/arcgis/arcgisserver/index.html

Ester, Martin, Kriegel, Hans-Peter, Sander, Jörg 1999. Knowledge Discovery in Spatial Databases, Lectures Notes in Computer Science：64-71.

Goodchild, F. M., Kryiakidis, P., Rice, M., Schneider, P. (2005) Spatial Web, *Report of the NCGIA Specialist Meeting on Spatial Webs*: 73-78.

Gruen, V. and Smith, L. (1960) *Shopping Town U.S.A.: the Planning of Shopping Centers*, New York: Van Nostrand Reinhold.

Han, J. and Fu, Y. (1995) Discovery of Multiple-level Association Rules from Large Databases , *Processings of the 21st International Conference on Very Large Databases*：420-431.

Jung, C. (2005) *Spatial Data Mining on Census Data — A Case Study for Location Analysis of Convenience Stores in Taipei city*, Master dissertation, University of NTU, Taiwan.

Kuo, R. J., Chi, S. C. and Kao, S. S. (2002) A decision support system for selecting convenience store location through integration of fuzzy AHP and artificial neural network, *Computers in Industry*, 47: 199-214.

Mennis, J. and Liu, J.W. (2003) Mining association rules in spatial-temporal data, *7th International conference on Geocomputation*

Microsoft (2003) ASP.NET, http://asp.net/Default.aspx?tabid=1

Miller, J. H. and Han, J. (2001) *Geographic Data Mining and Knowledge Discovery* , London: Taylor & Francis.

Negroponte, N. (1995) *Being Digital*, New York: Knopf.

Nelson, R. L. (1968) *The selection of retail location*, New York: McGraw-Hill.

Peng, Z and Tsou, M. (2003) *Internet GIS: Distributed Geographic Inofrmation Services for the Internet and Wireless Networks*, New Jersey: John Wiley & Sons.

Roddick, J. F. and Spiliopoulou, M. 1999. Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research , *SIGKDD Explorations* **,** 1(1): 34-38

Shekhar, S. and Chawla, S. 2003. Spatial Databases：A Tour , *Prentice Hall*

Tao, V. (2003) Online GIServices, *Journal of Geospatial Engineering*, 3(2): 135-143.