The Use of GIS in the Analysis of Customers Mobility Routes

Ana Mascarenhas Quintas
Institute of Statistics and Information Management, New University of Lisbon
Portugal
G2001181@isegi.unl.pt

José Jesus Costa
Institute of Statistics and Information Management, New University of Lisbon
Portugal
G2001163@isegi.unl.pt

Victor Hugo Ribeiro
Institute of Statistics and Information Management, New University of Lisbon
Portugal
G2001164@isegi.unl.pt

ESRI Users Conference 2003
San Diego

The Use of GIS in the Analysis of Customers Mobility Routes

Abstract

The aim of this paper is to show the role of GIS when using it to study the mobility phenomena of consumers. This is a very important subject were we can interact many fields of knowledge such as spatial analysis, location based services and Spatial Data Mining. This type of analysis allows to define strategically approaches in order to maximize sales results and customer satisfaction, as well to build communications infrastructures service that support the mobile commerce.

Key words: Data Mining, Spatial Data Mining, Location Based Services, and Spatial Analysis.

1. Introduction

Mobility has always assumed a basic role in the history of the humanity, being one of the main causes for the development of our civilization. Presenting advantages, it also had many disadvantages throughout many thousands of years, namely due to the lack of communication resultant from the great distance between communities or the individuals. This fact was overcome with the telecommunications appearance, and has changed in a radical way with the appearance of mobile telecommunications.

The mobile communication devices revolutionized in a radical way the life style of the humanity. Voice became nowadays only one of the many possible ways of communication along with data and image, in this type of equipment characterized for its high portability and capacity of personalization. The high availability of these devices and the, increasing, capacity of localization of them came on the basis of allowing the sprouting of a type of services supported by the geographic localization of the equipment, the Location Based Services (LBS).

The high potential of application of the LBS to business has contributed for its enormous development. Something similar has happen with Geographical Information Systems (GIS) whose main characteristic is the capacity to establish connection between spatial data and non spatial attributes, associated to a strong capacity of analysis of these two types of data together. The capacities of use of the GIS in business it's well accepted today and the number of applications it's growing every day enclosing varied areas such as: Marketing, cash collection, logistics, and many others.

The mission of this paper is to show to the existing relation between the GIS and the LBS, both depending, on an unquestionable way, on the geographic localization, as well to demonstrate the capability of these systems in the determination and analysis of routes of

mobility of customers through the use of the Spatial Data Mining, allowing the forecast of future mobility routes, essential to the knowledge of the behavior of the customers.

2. Location Based Services

In global terms the Location Based Services can be defined as services that use the capacity to determine and to transmit in a dynamic way the localization of one given entity, through a mobile communication network. In the case of the user these services are materialized in the form of a service available in a mobile communication device.

From the analysis of several definitions of Location Based Services, available in literature, all of them have a common factor, which is the geographic location. In fact, the LBS are not more then a service based on the current localization of the user. The initiative of that service could come from the user, for example to know the localization of a certain restaurant, or could come from the server of the LBS (e.g.: when it detects the proximity between the user and a certain restaurant with a specific promotion advising the customer for that situation through a message).

The number of possible applications of the LBS is very high, because they are very connected with something basic to humans, the mobility. We can, however, classify this type of services in five main categories:

2.1. Traffic Management and coordination.

With the localization of the user, we can manage his route in a road, preventing for example traffic accidents or even defining, which is the fast and/or shortest path to arrive at the destination.

2.2. Advertising Contents.

Depending on the localization and on the profile of the user, we can send messages focalized in order to maximize the potential of purchase of the user. Although with a great potential, this type of services will raise problems of ethics and data protection. The initiative of these services, normally, does not come from the user, so the user must allow, normally by subscription this kind of advertising messages.

2.3. Integrated Tourism Services

This is one of the services that will have more expression in the near future, due to its big range of applications; they will be the new type of travel guides. They will consist in a common platform of tourist services, which in face to the geographical location of the user

will offer a set of functions, from the simple localization of a hotel or restaurant, until the trace of a tourist route on the basis of parameters defined by the user. Another application will be, for example, when visiting a museum, where the visitor through a dedicated device, will receive information concerning each masterpiece every time he is near them.

2.4. Health Services

This is an interesting category of services that thinks about the health of the users. There are many possible applications, but we will give as an example only two: the monitoring of old people or with a low level of health, or the monitoring of tourists in dangerous zones, as are mountains and deserts. This type of applications can inform the user on potential risks, or even support in cases of dangerous accidents.

2.5. Entertainment

In this category we can also develop countless applications that goes from mobile games to sports. For example a mobile game were the players moves depend on their geographical position or a subscribed service in a soccer game, that will be activated when the user enters in the soccer stadium, and automatically its mobile device will make a download of an application that will give him all the information about the game, such as: replay of certain moves, statistics of the game, information about the players, simulations, among others. When the user leaves the stadium this application will be extinguished automatically from its device in order to save memory space.

This classification is not rigid, and has the virtue of segmenting the LBS in function of the main needs connected to this type of technology. In the near future new categories or new sub-groups will have to appear, to break these five initials. Everything will depend on the acceptance of this concept, the available technology, and still on the degree of acceptance from customers to the different types of service offered.

3. Geographical Information Systems and Location Based Services

GIS can be defined, in general terms, as Information Systems that process spatial data. In the same way LBS works and processes spatial data, so it can be considered as a specialized type of GIS. These two systems had different evolutions GIS was developed, as an Information System for the analysis of spatial data in a continuous and sustainable way, while LBS appeared and grew up in a very fast way with the appearance of the mobile communications and according to the user needs.

After its birth, in the sixties of last century, the GIS currently present a very high degree of m ic Information, basic support to the GIS, allows a

specialized knowledge on spatial databases and modeling of spatial data, based in methods supported in mathematics, statistics and complex numerical analysis. Most of this type of analysis techniques existed before the existence of the GIS, having however implied some time until its fully integration in GIS, normally due to computational reasons.

Also concerning to its principles of use, the GIS and the LBS had distinct ways of evolution. The GIS had been created for the use for experienced users armed of vast informatics resources and making use of some dedicated software extensions varying in accordance with the intended use, on the other hand LBS emerged with user-friendly interfaces and intended to get the maximum of non-specialized users.

As you can see, LBS are part of a vast universe called GIS. Due to its specification the LBS use the best characteristics existing in the GIS used to define concepts such as mobility and dynamic localization of the geographical position of one given entity. On the other hand LBS gives dynamic skills to the GIS that in the majority of the cases refer to a certain scenario in a specific moment in time.

4. GIS as the support for LBS

If we consider LBS as a specialized GIS, certainly many of the characteristics of GIS will be used in LBS. Some of them assume a basic importance, which we will try to describe:

4.1. Data Collection and Data conversion

GIS works with many different types of spatial data, on the other hand the topology of this data it's very rich. The actual GIS, in general, support this multiplicity of formats, and consequently LBS will use also this important characteristic.

4.2. Management of Geographical Databases

The actual Geographical databases ought to be opened in standard interfaces, which is one of the basic characteristics to the richness of LBS.

4.3. Spatial Data Analysis

Concerning Spatial Data Analysis LBS will use the advantages of the analyze methods utilized in GIS, from the classical statistics to the advanced algorithms of Data Mining. We believe that one of the most important skills of LBS will be the data modeling of customer's scenarios.

4.4. Geographic Data Presentation

One of the most well known characteristics of GIS is the visual representation of the spatial data, due to the use of powerful computational techniques, allowing that 2D representations grew for three-dimensional representations (3D) and later on to spatial-temporal representations (4D). Easily we can understand the importance of the use of this type of visualization techniques in LBS.

The Location Based Services are still on an initial phase of its development, which means that much of its potential remains unexplored, by many reasons, especially for a technological question and on other hand we still need to discover the full potential that these services can have near the consumers. It's easy to believe that in the near future, due to the similarity between GIS and LBS, that this later will be one of the many functionalities of the Geographical Information Systems.

5. Spatial Data Mining

Every day, the organizations receive and store data, whose origin can be internal, normally resultant from the operational component of the organization, and external, resultant from the interaction with the environment where it's inserted. These data warehouses constitute important resources, due to the information that we can get from them. The knowledge that we can get from those data must be used in the improvement of procedures, in order to get a more competitive and proactive organization. This improvement results from the identification of standards and behaviors, allowing the organization to take corrective measures in order to foresee more competitive future positioning face to the market. Although these resources are in almost all organizations, not all of them are prepared to use this advantage, because this knowledge it's implicit, and cannot be extracted directly through the conventional database management systems. The solution for this challenge is called Data Mining (DM).

DM is applied in many areas of knowledge and consists of the discovery of information eminently predictive and strategically in great amounts of data. It's a process wide automated that actuates, not only describing facts of the past, but mainly to foresee future trends, originate knowledge that can be translated in important competitive advantages.

As in other sciences, in Geography the work ambience evolved from a poor situation in data and computational resources to another one rich in these two variables. For this alteration

has contributed the great development that GIS had during the last 40 years, allowing, together with the information systems, to create great data warehouses. In this context it is necessary to create mechanisms of analysis for the great flows of data that elapse in general from the development of the Information Systems and especially from the Geographic Information Systems. It is in this situation that appears the concept of Data Mining, which with the contribution of many areas of knowledge has created tools to explore the data, in order to detect standards between it or to foresee the emergency of future standards.

With the development of the Geographical Information Systems, a new type of data has appeared: the spatial data that has add the variable location to the other variables of characterization. More recently, has been given an increasing attention to another interesting characteristic that is the introduction of the dimension time, that allows the study of the evolution of the events, it can assume to predictive behavior or a regressive one in a study of an event.

The Spatial Data Mining (Gahegan 2001, Openshaw 1999), must be understood as a special type of DM that tries to carry similar generic functions to the ones of the conventional DM, modified to warn the aspects of the geographical information (Openshaw 1994, 1999). However, what it has of truly important in the Spatial Data Mining it's the field of study, the spatial and the Spatial-Temporal patterns distributions, once this provides a unique vision of the reality and confers this kind of Data Mining singular characteristics. The relevance of the Spatial Data Mining as increased with the increasing spreading and consequent globalization of spatial data in many fields of knowledge, specially in digital maps, remote sensing images and in some countries with the Census spatial data.

The Spatial Data Mining its based essentially on no parametric models, that we can define as models that depend essentially on the use of the data, for opposition to the use of specific knowledge of the domain of the problem in study. The no parametric models are also many times appointed as data-driven models. This type of model has known great success especially in the resolution of complex problems, knowing that normally they use great data sets that contains a high number of examples. The essential premise of these methods is that relations that occur in a consistent form in the data set will happened again in future observations of the event; we can call this an inductive approach. So, if we get enough set of examples and model its behavior we will be able to adjust a model of arbitrary complexity that will allow to repeat the behaviors observed.

An important benefit of this type of models is that they do not demand a deepened knowledge of the phenomena in study, which is a particularly useful fact in the treatment of spatial problems. However, although theoretically it's possible to solve any problem, independently of its complexity, there are still some problems that continue to be too much

complex. This fact it's connected with limitations in terms of the data collection and time of processing to construct the non-parametric model. The inductive approaches, which characterize the Spatial Data Mining, guided for the data (data driven) applied to the modeling and Spatial Analysis, could be the form to facilitate the creation of new knowledge and to help to extract excellent information on mobility routes proportionated by LBS. The tools of Spatial Data Mining that have been developed can play a central role attending the process of exploration of the great amounts of data produced by LBS systems, in the search of recurrent standards and persistent relations.

6. The Use of Data Mining in LBS Systems

The use of DM in LBS offers a big range of chances for knowledge discovery, in specially related with routes of mobility of people and goods, allowing for example to try to answer, amongst others, to the following questions (Smyth et. al. 2001):

- Which type of important information can be removed of the mobility routes?
- What will be the next location of a customer?
- In the present location of the customer x, which is the type of information that he wants to know?
- Better planning of the mobility infrastructures.
- Foresee and avoid traffic zones in the rush hour.
- Planning marketing campaigns more focalized in the potential buyers.

While the traditional LBS answer to a request of a user, and the service given to him is in function of his location, LBS's supported by the Spatial Data Mining try to anticipate the situations that can occur in a future based in the current localization, a predictive behavior. We call this type of model as a Sb-Pb model, in other words "stored behavior - predicted behavior", this model is summarized it two steps (Smyth et. al. 2001):

- Discover Spatial-Temporal patterns that could be parameterized and stored;
- Use these patterns in order to create predictive models.

The Sb-Pb model has cover the existing imperfection in the conventional model of LBS between the universe of the simulation and the real world lived by the user of LBS. Another great advantage of the Sb-Pb is the fact that the type of information kept it's equal to the type of information that we want to get, which prevents loss of time data conversion, normalization and errors resultant from the analyses of these differences. The main

indicators to apply in order to evaluate the success rate in the results of this two steps SB-PB model are:

- Density of the participation;
- Previsions quality.

The first one has to do with the number of introduced mobility observations in the model, which must be very high for a good use of the Data Mining tools. The second one evaluates the level of satisfaction of the consumer when he uses the service supplied by LBS systems: as much bigger it is, more efficient is the service and consequently more generalization possibilities it has.

We can say that LBS systems are, in last instance, advanced Data Base Management Systems, whose functioning depends from the geographical location of the users, or the goods, that can be optimized in function of the right implementation of the Spatial Data Mining tools.

In the field of the Data Collection systems, Location Based Services (LBS), are assuming today a vital importance and will have a fundamental role in the next years, as a system by excellence for data collection. The type of data in use assumes a great importance due to its Spatial–Temporal characteristics. Although the initial idea of LBS it was the offer of services on the basis of the localization, the fact is that the resultant information of the analysis of the data collected it's also very important to build predictive scenarios. Building these predictive scenarios it's the main role of the Spatial Data Mining, in order to be able to maximize the power of LBS.

7. Conclusions

The sources of data in the organization have grown substantially in the last years. It's notorious that these data come each day more, geographical referentiated to a specific position. When that position develops through time we are in a presence of a mobility route. It is in this context and Geographical Information Systems environment that Location Based Services emerge, making it possible to answer, in a certain way based on a specific geographical localization. To know or to be able to predict/anticipate specific mobility routes can be of major interest, in the sense that we are able to understand the behavior of the consumer and thus, their acquisition patterns through a specific good. With all this in mind we come to the concept of Spatial Data Mining as a tool for determining possible future routes, allowing companies to adopt the necessary strategic measures/actions in order to maximize their effectiveness.

As we were able to verify in this presented paper, there is a direct relation between GIS and LBS. The later need crucial characteristics for GIS. LBS need the infrastructure of existing data in one or ore GIS. In this sense, GIS can be considered both as the raw material and as the infrastructure for any kind of LBS that intends to maximize its effectiveness.

This domain of research, LBS, foresees huge investments and developments in the coming years, involving the most varied areas of knowledge, that goes from communications equipment, to the analysis algorithms, visualization techniques, database systems, marketing, behavioral psychology, organizational behavior, logistics, among others. After Industrial Revolution and more recently the Information Revolution, we are facing a Knowledge Revolution, which is fundamented on this type of technologies and concepts. The knowledge that it will be possible to obtain on the behavior of the consumers, are about to revolutionize in a historically unique way, the means and forms on how we offer specific goods and services, allowing to adequate them with extremely accuracy and effectiveness to their needs.

At the end of this paper we want to leave a question: "What will come next?"

8. Bibliography

Burrough, Peter A.; McDonnell, Rachael ; McDonnell, Rachel A.. 1998. *"Principles of Geographical Information Systems"*. Oxford University Press.

Chang, G. Healey, M. McHugh, J. Wang, J.2001, "*Mining the World Wide Web - An Information Search Approach*", *The Kluwer International Series on Information Retrieval*, Volume 10, Kluwer Academic Publishers

Davis, David E.. 2001. *"GIS for Everyone"*. Environmental Systems Research Institute (ESRI).

Heywood, Ian; Cornelius, Sarah; Carver, Steve. 1999. *"An Introduction to Geographical Information Systems"*. Prentice Hall.

Jensen, Christian S., Friis-Christensen, Anders, Pedersen, Torben B., Pfoser, Dieter, Saltenis, Simonas and Tryfona Nectaria. *"Location-Based Services – A Database Perspective"*. Department of Computer Science, Aalborg University.

Longley, Paul A.; Goodchild, Michael F.; Maguire, David J.. 1999. *"Geographical Information Systems: Principles, Techniques, Applications and Management"*. John Wiley & Sons.

Longley, Paul A.; Goodchild, Michael; Maguire, David J.; Rhind, David W.; Lobley, Joe. 2001. *"Geographic Information Systems and Science"*. John Wiley & Sons.

Mountain, David and Raper, Jonathan. 2002. *"Positioning techniques for location-based services: caracteristics and limitations of proposed solutions"*. Geographic Information Science Group, Department of Information Science, City University, Northampton Square, London.Davis, G., *Management Information Systems: Conceptual Foundations, Structure, and Development*, McGraw-Hill, 1974.

Openshaw, S. (1994), "*Two exploratory space-time-attribute pattern analysers relevant to GIS"*, In: Fotheringham, A. S., Rogerson, P. A. (eds) Spatial analysis and GIS, Taylor and Francis Ltd. 1900 Frost Road, Suite 101, Bristol PA 19007, pp 83-103

Roddick, John F.; Lees, Brian G.. *"Paradigms for Spatial and Spatio-Temporal Data Mining"* in Miller, Harvey J.; Jiawei, Han. 2001. "Geographical Data Mining and Knowledge Discovery". Taylor and Francis

Smyth, Carl Stephen. "Mining mobile trajectories" in Miller, Harvey J.; Jiawei, Han. 2001. "Geographical Data Mining and Knowledge Discovery". Taylor and Francis

Shiode, Narushige; Li, Chao; Batty, Michael; Longley, Paul; Maguire, David. 2002. *"The Impact and Penetration of Location-Based Services"*. Centre for Advenced Spatial Analysis, University College London.

Upendra Shardanand and Pattie Maes. 1995. "*Social Information Filtering: Algorithms for Automating "Word of Mouth"*", in Proceedings of ACM Conference on Human Factors in Computing Systems, CHI'95.

Vários. 2000. *"What are Location Services? The GIS Perspective"*. Environmental Systems Research Institute (ESRI).

Vários. 2000. *"A Request for Technology In Support of an Open Location Services (OpenLS™) Testbed"*. Open GIS Consortium, Inc.

Vários. 2001. *"Location Based Services. Identifying and Capitalizing on Service Opportunities"*. Cambridge Strategic Management Group.Vários. 2000. *"What are Location Services? The GIS Perspective"*. Environmental Systems Research Institute (ESRI).

Weiss, Sholom M. 1997, "*Predictive Data Mining: Web Site Access"*, Morgan Kaufmann Publishers

Wolfson, Ouri. 2001. *"The Opportunities and Challenges of Location Information Management"*. Department of Computer Science, University of Illinois, Chicago.