Journal of Korea Multimedia Society Vol. 12, No. 6, JUNE 2009(pp. 777-784)

A Spatial Analysis Supporting System Based On CRM And Data Mining Technique

Jeong Min Seo[†], Hu Xiao Wei[††], Sang Moon Lee[†††]

ABSTRACT

Recently, the importance of geoCRM (geographic Customer Relationship Management) systems are growing rapidly. So, result of the recognition that their applications extend well beyond the traditional CRM systems with the advent of ubiquitous environment and generalized location based services. A majority of traditional CRM systems are either incapable of managing spatial data or are not user-friendly when doing so. On the other hand, the geoCRM systems can be built as providing the geographic-based functions about CRM, including spatial and market analyses and the visualization of customer data, etc. However, it lacks the specific model and implementation of the geoCRM systems, being caused by the incomprehension of needs, the absence of related standards and the difficulties of development, and so on. In this paper, we develop a new spatial analysis supporting system that to enhance productivity through the convenient use and management of spatial data. The functionality provided by our system includes a set of analysis functions based on data mining techniques which allow a user to affect powerful transformation on spatial data. Particularly, both spatial data and non-spatial attributes can be efficiently handled as an object through our OODBMS.

Key words: CRM, Spatial, geoCRM, GIS, Data Mining

1. INTRODUCTION

GIS (Geographic Information System) is pointed end information system that input geographical and attribute information about various feature that exist to real-world into computer system and make so that can support various plan establishments and decision-making and industry activity efficiently[1]. Past days, most of GIS worked manually that needed much expenses and time are cost and used in very limited field. But, recently, used in various application fields by technology development of wireless system, LBS (Location Based System), Image processing etc. Lately, spatial data through computer because needs that wish to do implemented becomes many the importance of GIS more increase. Also, new concepts of SIIS (Spatial Imagery Information System), GNSS (Global Navigation Satellite System), ITS (Intelligent Transport System) etc. are introduced by wireless and development of distributed system technology[2]. Also, geoCRM of that utilize in marketing operation establishment using spatial analysis ability of GIS is getting into the spotlight in management information technical field. geoCRM applies geographical element that influence in corporation business such as form of road or building location, floating population, transfer path as terminology that compose GIS and CRM in customer management[3]. In this paper, we design and implementation of GIS based spatial analysis support system that easy integrated of existing CRM system, such as circulation, banking, manufacture, lo-

※ Corresponding Author : Sang Moon Lee, Address : (380-702) 72 Daehak-ro Chungju-si, Chungbuk, Korea, TEL : +82-43-841-5267, FAX : +82-43-841-5260, E-mail : smlee@cjnu.ac.kr

Receipt date : Dec. 30, 2008, Approval date : Apr. 8, 2009

[†] Dept. of Computer Science, Konkuk University (E-mail : jmseo@kku.ac.kr)

[††] Dept. of Computer Science, Chungju Nat'l University (E-mail : huxiaowei2001@hotmail.com)

[†††] Dept. of Computer Science, Chungju Nat'l University

gistics, real estate, communication. And, we used spatial clustering method to search point spatial features such as store candidate locations[4]. Also, proposed system can be support that extract, retrieval, growth and reduction of spatial features.

2. CASE STUDY

2.1 CRMs

Customer Relationship Management system, or CRM, is how your company develops, manages and leverages an understanding of customers. Various technologies are used to aid in this process. Many of them center around Business Search or sometimes the term Business Intelligence is used instead. At the center of whatever name you wish are the capture, storage and analysis of information about customers. In the early days CRM was called Sales Force Automation or SFA. This meant pushing through information from first contact to signed purchase order, and that was all. Later it was learned that going deep on customer groups and eventually individual customers could reap large benefits in terms of loyalty, cross-sales and up-sales. But in any data capture about customers the information comes in two forms. Numerical indicators, codes or otherwise structured information and unstructured information. The unstructured information is notes or feedback from customers. The unstructured information are really just sentences but contain valuable information about customer wants, desires, gripes, etc. The trouble with CRM systems up to this time is they ignore the processing of these sentences for insight -either for individual records or across many records [3,5,6].

2.2 geoCRM

Lately, take advantage of GIS technology to CRM that supply information of simplicity textual information, competition store analysis, research about geoCRM that support multi-dimensional analysis and so on that use customer position information are consisting abuzz[7]. [8] is various spatial data mining technique and some relation such as connection spatial clustering, spatial character delineation, spatial classification, spatial direction retrieval, generalization, spatial operations on a geoCRM's function. [9] proposed MAP/CRM system to support CRM. MAP/CRM used OMEGA [10] that is OOSDBMS to store and manages map information and customer information to marketing aided. MAP/CRM has client/server structure and added spatial analysis function to existing CRM of text center like data warehouse. [11] use decision-making tree analysis that is one of data mining techniques, geoCRM shows customer specification class and test examples. For this, that system using NeedEye company's Marzer tool that concord to fit customer information and customer's administrative division, area customer ranges connecting customer information and map that do grouping and geographical changes and supply information about customer that is atomized visually presents geoCRM's logical modeling with the financial world as the central figure and investigated about the practical use[12]. Those proposed Modeling is collected from corporation's interior data and outside data, spatial data transform by data that application is possible in GIS through mapping. And, store to spatial data warehouse and that is G-TOOL using visualization functions to user variously viewpoint.

2.3 Spatial Data Mining

Figure 1. shows to a breifly about representative spatial data mining techniques[17,18,24]. The most important spatial data mining technique is spatial clustering as see in Figure 1. Spatial clustering technique is method to find interested structure or cluster justly from data that unuse some background knowledges such as conceptional hierarchical structure informations.

Table 1. Spatial Mining Tech's.

Methods	Examples
Clustering	DBSCAN[14], PAM[15], CLARANS[13,15]
Description	Neighbors[13,16]
Classification	Spatial decision tree[13,14]
Detection	Pseudo code[14], global trend, local trend[13]
Generalization	Hierarchical structure[15]
Association	<g_close_to, highway17>[14], X → Y(80%)[15]

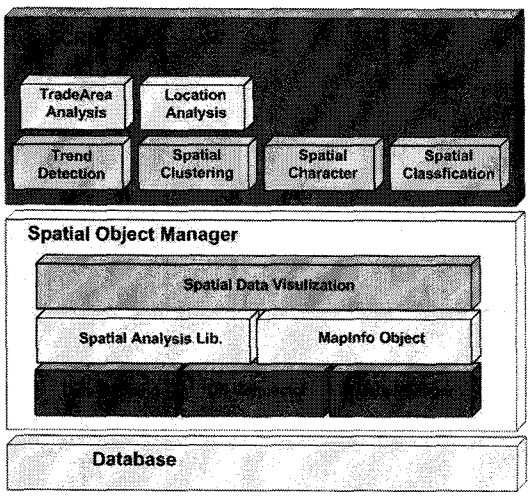

Fig. 1. System Architecture

The representative spatial clustering techniques that k-medoid clustering[13,15] such as CLARANS, single scan clustering to search neighborhood of all spatial features in a spatial database and, density base clustering[14]. Spatial characterization is method to find concise delineation or description about selected subset of database[13,16]. Spatial classification technique is method to allocate one feature in one class from set of given class based on attribute value of feature[13,14]. Spatial trend detection technique is method that defined by regular change of one or many non-spatial attributes[13,14]. [15] is introducing about techniques that investigate connection knowledge discovery of generalization, clustering and spatial association rule base. Techniques to discovered spatial association is much concerned with spatial character descript techniques. In concern with, [15] is presenting status relation, distance relation and direction relation by basic three types of spatial relations. And defined six extended operations using by these three types. [14,15,19,20] and [21] reify involved neighborship, and define and presenting "Neighborhood index" to avoid that approach space object confidence. Also, presenting "Cost model" to predict expense that "Neighbors" operation by existence and nonexistence of "Neighborhood index".

3. A SPATIAL ANALYSIS SUPPORTING SYSTEM

3.1 System Architecture and Modules

Propose system is consisted of three portions greatly as appear in subsequent figure 1.

geoCRM using spatial feature manager to supporting user interface of proposed system that function of map's customer position information output and search, area population size, subway flow population size, area selection, position search, map registration etc. And, support GIS concerned functions that spatial data mining, commerce area analysis, spatial analysis of location selection.

Spatial feature manager is divided by six kind of module that manages data that need to spatial analysis, and support spatial analysis function. There are 'data mapping', 'spatial analysis lib.', 'spatial data visualization', 'DB connector', 'MapInfo object', 'data manager'.

Finally, database of most important in the geoCRM consisted by enterprise database, external database, spatial database. Enterprise database manages customer information, store information, corporation's interior data such as transaction information. External database stores data such as necessary statistics of population information, life style information in customer management. Spatial database manages feature

based map as like spatial data

3.2 Map and Spatial Area Search Operations

Proposed spatial analysis system in this paper, feature manager extracted spatial features from SHP typed map. Proposed system has two kind of operations that map and spatial operations. The map operations are extension, reduction, and move operations. The extension operations divided by map and windows based, and those operations co-operated with each other.

Coordinate conversion method between digital map and window screen for display uses size for ground and reduced scale about window screen size. When change map coordinate to window coordinate, map coordinate system and window coordinate system are difference, therefore, must consider following item. First, must consider Y axis direction of digital map and window. Second, must change digital map's coordinate values to window screen coordinate of pixel unit. Finally, must set height and width ratio of fixed digital map area and window screen. Use MBR (Minimum Boundary Rectangle) to calculate scale of digital map and window screen. Also, datum point uses center coordinate of map and display that is not influenced Y axis and size.

Basic principle of extend or reduce map data which controls size of map's MBR via digital map's central point. Also, if map's central point move to extend or reduce point then, spatial feature and relative distance are changed to screen coordinate from changed central point.

Spatial feature's retrieval in specification area can divide by point search of circle and point search of polygon. Point search of circle can examined that if small or same than radius length comparing with radius length and Euclid distance of target point from central point. Blow figure 2. shows which method.

To examined target point in specification position is situated inside polygon on plane that target

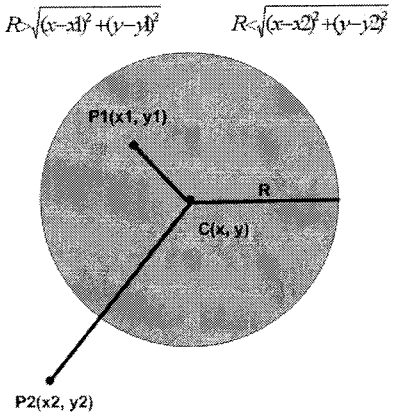

Fig. 2. Map Data Manager

point's Y axis by standard all line segments of polygon and count number of point that cross over polygon. If number of point that cross is odd number, is on interior of polygon.

3.3 Clustering Operations

The best known non-heirarchical clustering procedure is *k-means* clustering[13,15]. For n samples with j selected features, the goal is to select k centroids in X-dimensional space such that the when each sample is assigned to the cluster with the closest centroid, the sum of the distances-squared to their centroid is a minimum. Finding the optimum locations of the k centroids is not an easy problem, and again we are looking for the set of j features that produce the most homogeneous clusters.

$$X=\begin{bmatrix} x_{11}\, x_{12}\cdots x_{1n} \\ x_{21}\, x_{22}\cdots x_{2n} \\ \cdots \\ x_{n1}\ \ x_{n2}\ \ x_{nn} \end{bmatrix} \qquad \textbf{Formula 1.}$$

An outline of the *k-means* Clustering algorithm is as follows.

1. Randomly assign each of the N samples to one of the K clusters.
2. Determine the centroids (average coordinates) of each cluster and the sum of the distances-squared from all samples to their centroid.

3. For each sample
   a. Sequentially place the sample in each of the other clusters.
   b. Calculate the change in the centroid positions for the two affected clusters.
   c. Calculate the new sum of distances squared. This only has to be done for the two affected clusters as long as the sum-of-squares distance is stored for each cluster.
   d. Place this sample into the cluster that has the smallest sum of squares and update the stored, intra-cluster sums.
4. If any sample changes from one cluster to another, return to Step 3.

To speed-up the final assignment of samples in clusters, different algorithms have been proposed, which includes H-Means Clustering, J-Means Clustering, and Variable Neighborhood Search. Though K-Means and its variants will generate a good set of clusters, they will not be the optimal set of K clusters for these NUMFSUSE features. The reason is that in K-Means Clustering final set of clusters depends on

- The initial distribution of samples to clusters
- The ordering of the samples.

Since there are KN ways to initially distribute the points amongst the clusters and N! orders of the points, finding the optimal set of clusters is computationally impossible in most cases. The procedure used in ClassCK can best be described as a modified H-Means Clustering algorithm. It can be outlined as follows.

1. Select a set of K samples to act as initial centroids.
2. Assign all samples to the cluster with the closest centroid.
3. Use the samples in each cluster to recalculate the position of its centroid.
4. If the cluster number of any sample changes, return to Step 2.

In the first pass of this algorithm the samples do not have an assigned cluster so Step 4 automatically returns the process to Step 2. After a few iterations, no samples change clusters and the process stops.

This procedure is independent of the order of the samples since all samples are assigned to clusters at the same time. The problem is therefore reduced to finding the best set of K samples to act as initial centroids. The total number of unique sets of initial centroids is

$$\frac{n\times(n-1)\times\ldots\times(n-k+1)}{k!} \qquad \textbf{Formula 2.}$$

This can be a large number of initial sets, so an Evolutionary Programming algorithm is used to search for the best set of initial centroids. Each set is then passed to this clustering algorithm and the cost of these clusters (the number of misclassified samples, the average cluster size or a combination of both) is used to search for the best set of initial centroids.

4. SYSTEM IMPLEMENT AND EXPERIMENT

In this section, we examine method to apply spatial analysis support system that proposes to geoCRM. Also, apply geoCRM using system that is operating in S company for an experiment.

4.1 System Experiment Environment

Apply proposed spatial analysis support system to [17] that we stored enterprise data, internal data, and spatial data in S company in an experiment system. [17] has various functions for customer management and business support by windows based business application that has client/server's architecture[22,23]. Enterprise data uses S company's database which use [17]. S company's DBMS uses MSSQL Server 2000, and server is Windows2003 environment. Spatial data constructs by proposed system's map data manager that use digital map and managed by Seoul city's SHP typed digital maps.

Table 2. Experiment Environment

H/W	CPU	Intel Core2 2.4GHz
	RAM	3582MB
	HDD	500GB
S/W	OS	MS Vista Home Premium Ko
	DB	MSSQL 2000, db4object
	Tool	Visual Studio.Net 2005
	P/L	C# 2.0

4.2 System Experiment

Area search and registration are available in digital map of necessary area by administrator side application which map data manager uses in geoCRM. Also, can input other information such as color and attribute data. Below figure 3. shows map data manager's execution result.

Figure 4. shows spatial feature type of building, road, block by result that input from digital map, and those data managed by its layer typed.

Figure 5. shows that attribute data input and update captured screen. This module managed attribute data such as layer type, record counts, and other attribute data about spatial features. Also, this module can input and update data of land lot number data that first you can clicked special area on the map and inputted land lot number.

Blow figure 6. shows that result of business point retrieval. According to figure 6., mark shop which select and denoted together radius 500m inner customer's number and competition shop.

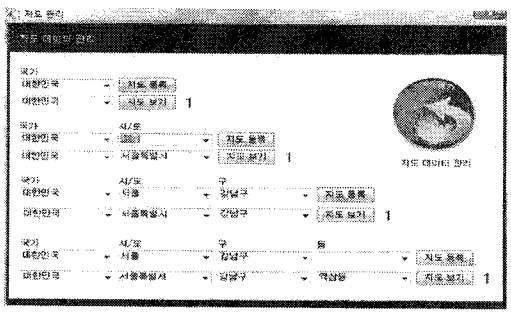

Fig. 3. Map Data Manager

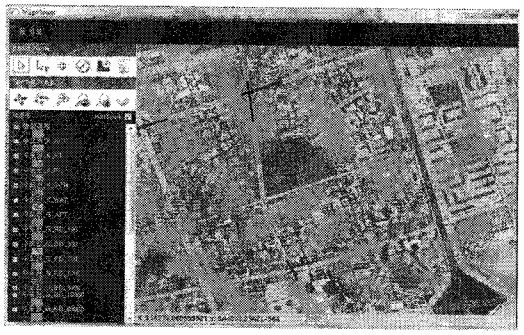

Fig. 4. Map Data Input

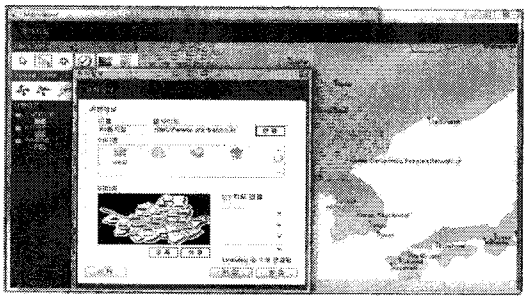

Fig. 5. Attribute Data Input and Update

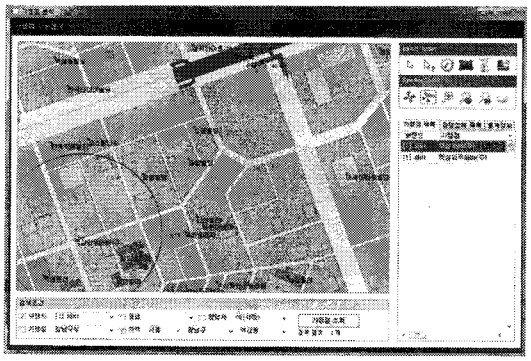

Fig. 6. Map Data Extension and Reduction

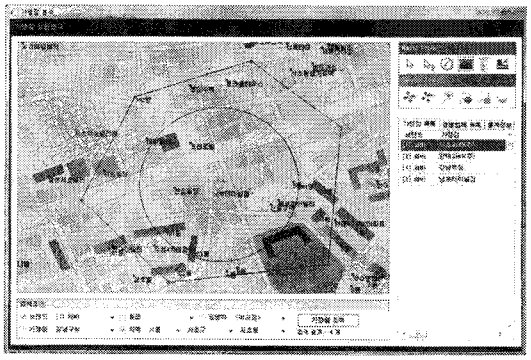

Fig. 7. Commerce Area Analysis

Figure 7. is captured screen that inspect customer's number in area that user chooses specification selection area. Its function can be available if select user area selection menu that mark by box on screen right side. And, area that select also using a mouse in form of polygon transformation possible.

5. CONCLUSION

We, in this paper, design and implemented spatial analysis support system for geoCRM. Proposed system has following characteristics.

First, spatial feature and attribute data can manage integration used spatial data manager.

Second, can get various result because offering various spatial analysis operations that need to geoCRM.

But proposed system can be extract spatial features in SHP format's digital map but impossible from various digital map such as DXF or MIF. So, in future, we must extends to extract spatial features from various typed digital map. Also, do that can be examined various spatial feature extending point, polygon and polyline.

REFERENCE

[1] Bruce E. Davis, GIS: A Visual Approach, OnWord Press, Santa Fe, New Mexico, USA, 1996.

[2] J. Y. Oh, New GIS Paradigm in The Web 2.0, KISA, Report Vol.2, Seoul, Korea, 2007.

[3] DMReview Co., "Enterprise Content Management: Conducting Your First Text Mining Project," http://www.dmreview.com/article_sub.cfm?articleId=8049

[4] Han, H., et al., Spatial Clustering methods in Data Mining, A Survey, Geographic Data Mining and Knowledge Discovery, Taylor& Francis, London, UK, 2001.

[5] Gartner Group, "Financial services firms must look at new ways to maximize CRM, says Gartner," http://www3.gartner.com/press_releases/pr3sept2003c.html

[6] CRMBuyer Co.,"Why Customers Are Still Angry," http://www.crmbuyer.com/story/33883.html

[7] S. Y. Jin, G_CRM, The Keyword of CRM Solutions, Management & Computer, Seoul, Korea, 2002.

[8] K. J. Lee, et al, "gCRM and Spatial Data Mining," KISS, Proc.. Spring Conf., Vol.29, No.1, pp.38-44, 2002.

[9] J. W, Jeong et al., "Map/CRM: GIS based CRM Used by Customer's Geographic Information," KSIGDB, Vol.17, No.1, pp. 85-92, 2001.

[10] H. K. Kang. et al., "Design And Implementation of OMEGA Spatial Feature Management System," KSIGDB, Winter Conf. Vol.14-1, pp. 197-202, 1998.

[11] S. S. Lee. et al., "gCRM Used Customer Information," JKSS, Conf., Vol.18-3, pp. 567-581, 2005.

[12] Y. H. Jeong., A Study of CRM, G_CRM Used by GIS, Kookmin Univ. Master Thesis, Seoul, Korea, 2002.

[13] Ester M. et al., "Spatial Data Mining:Database Primitives, Algorithms and Efficient DBMS Support," In Data Mining and Knowledge Discovery, Vol.4, pp. 193-216, 2000.

[14] Ester M. et al., "Spatial Data Mining: A Database Approach," Proc. 5th Int'l. Symp. on Large Spatial Databases, pp. 47-66, 1997.

[15] Koperski K. et al., Spatial Data Mining: Progress Challenges, TR 96-08, Univ. of British Columbia, Vancouver, Canada, 1996.

[16] Ester M. et al., "Algorithms for Characterization and Trend Detection in Spatial Databases," Proc. 4th Int'l. Conf. on Knowledge Discovery and Data Mining, pp. 44-50, 1998.

[17] Stormsoft Co. Ltd., "FX2", http://www.stormsoft.co.kr/FX2.html

[18] J. M. Seo., A Conceptual Feature-based Spatial Data Modeling Based on Groundwater Metadata and Attributes, Hankyong Nat'l Univ. Doctor Thesis, Ansung, Korea, 2006.
[19] Ester M. et al., "Database Primitives for Spatial Data Mining," Proc. Int. Conf. on Databases in Office, Engineering and Science, Freiburg, Germany, pp. 23-26, 1999.
[20] "GeoMiner: A Knowledge Discovery System for Spatial Databases and Geographic Information System.," http://db.cs.sfu.ca/GeoMiner/
[21] "GeoVRML Visualization: A Tool for Spatial Data Mining," http://www.geog.psu.edu/ica/icavis/
[22] Shashi S., et al., Spatial Database: A Tour, Prentice Hall Press, Englewood Cliffs, New Jersey, USA, 2003.
[23] Zhong-Ren Peng et al., Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks, John Wiley & Sons Press, Hoboken, New Jersey, USA, 2003.
[24] J. M. Seo., et al., "A Spatial Analysis Supporting System Based on CRM," Proc. MITA2008, Int. Conf. on MITA, Chiang Mai, Thailand, pp. 37-40, 2008.

Jeong Min Seo

He received the B.S degree in Computer Science from Chungju National University and M.S. degree in Computer Science from Chungbuk University, Korea, in 1996 and 2000. And Ph.D. degree in Computer Engineering from Hankyong National University, Korea, in 2006. He is currently a teaching professor at the Dept. of Computer Science from Konkuk University, Korea. His research interests include Image Processing, Spatial DB, and Multimedia System.

Hu Xiao Wei

He received the diploma degree in Accounting Computerization from Xuzhou Institute of Technology, China, in 2007, and B.S degree in Computer Science from Chungju National University, Korea, in 2008. He is currently a M.S. course at the Dept. of Computer Science from Chungju National University. His interested research is Multimedia System and Spatial System.

Sang Moon Lee

He received the B.S and Ph.D. degree in Computer Science from Hongik University, Korea, in 1980, and 1993. And M.S. degree in Computer Science from Yonsei University, Korea, in 1984. He is currently a professor at the Dept. of Computer Science from Chungju National University, Korea since 1985. His interested research is Ubiquitous DB, Multimedia System and Spatial System.