

공간적 의사결정을 위한 공간 데이터 웨어하우스 설계 및 활용

박지만 · 황철수

경희대학교 지리학과 대학원 · 경희대학교 지리학과 조교수

Ji-Man PARK and Chul-Sue HWANG

Abstract : The major reason that spatial data warehousing has attracted a great deal of attention in business GIS in recent years is due to the wide availability of huge amounts of spatial data and the imminent need for turning such data into useful geographic information. Therefore, this research has been focused on designing and implementing the pilot tested system for spatial decision making. The purpose of the system is to predict targeted marketing area by discriminating the customers by using both transaction quantity and the number of customer using credit card in department store. Focused on the analysis methodology, the case study is aiming to use GIS and clustering for knowledge discovery. The system is a key section of the research of multi-dimensional and spatio-temporal analysis in the internet environment.

Key Words : spatial data warehousing, GIS, cluster analysis.

1. 서론

급속하게 발전한 인터넷 환경과 자료의 증가는 수치화된 지리정보의 폭발적인 증가를 가져왔다. 이러한 자료의 대량화에 따라 토지이용, 범죄, 의료, 마케팅 전략 등 공공부분과 민간부분에서는 매우 상세한 수준의 지리학적 해결책을 요구하고 있다.

지리정보시스템(GIS : geographic information system), 위성항법장치(GPS : global positioning system)등 공간 표현 응용분야는 점, 선, 면과 같은 2차원, 3차원의 공간 자료를 효과적으로 표현하고 질의할 수 있는 방법이 필요하다. 더불어 래스터 모델을 포함한 소리, 움직임, 지리참조(Georeferenced)된 멀티미디어 자료, 위치를 인식하는 장치와 그것을 이용한 무선통신 장비, 고해상(high spatial and spectral resolution)의 위성영상 출현 및 방대한 양의 저장소는 이러한 현상을 가중시키고 있다(Miller, 1999). 또한 인터넷 사용자가 각각 개인적인 움직임을 시간, 공간적으로 추적하는 것이 가능하게 되었다. 지리학적 지식발견은 관심 있는 지역의 인구, 산업 특성과 공간적 특징을 고려하여 이루어지는데, 그 특징은 공간적 자기상관성(spatial autocorrelation)과 공간적 이질성(spatial heterogeneity)이라는 특성을 가지며, 시-공간적(spatio-temporal)관련성은 더욱 복잡하게 연관되어 있다. 이러한 맥락에서 분석을 위한

거리측정은 각 주제에 따라 다르게 하고, 지도학적 변형은 적절한 시간에 흥미 있는 패턴을 발견하여 의미있는 지식을 얻는데 이용된다. 결과적으로 이렇게 발견된 지식은 다양한 수준의 측정법을 이용하여 효과적인 의사결정 도구로 구체화된다.

따라서 본 연구에서 대용량의 지리공간자료는 정보분석 자료로 활용되어 지리학적 지식을 발견하고 기업이나 공공단체의 의사결정을 지원하는 형태로 발전할 것이다. 이러한 맥락에서 다차원적인 공간 데이터 웨어하우스를 설계, 구현하였다. 또한 시공간(spatio-temporal)적 특징과 더불어 사례분석을 통해 실험적으로 구현된 공간 데이터 웨어하우스를 통해 고객의 인구속성, 구매 패턴 등을 연구하고, 목표 마케팅(target marketing) 지역을 예측하였다. 또한 이 예측된 결과로 서울시의 공간, 비 공간 자료들 사이에서 각기 다른 차원과 대비하여 데이터 마이닝의 기법 중 군집기법과 GIS의 공간분석기법으로 시각화하였다.

II. 이론적 고찰

2.1 공간 데이터 웨어하우스

공간 데이터 웨어하우스는 공간, 비공간 데이터가 적용되고 점, 선, 면, grid-cell 등의 공간 객체를 저장, 색인 하는 방법을 지원한다(Morrison, 1999). 더불어 공간 객체(object)는 실세계의 지리학적 형상을 갖고 위상관계, 측정단위, 공간 해상도를 포함하며(Bedard, 1999), 일반적인 데이터 웨어하우스와 유사하게 통합된, 주제 지향적이며, 시간성을 갖고 비 휘발성의 공간, 비 공간 자료를 보유하는 저장소라고 정의한다. 공간 데이터 웨어하우스로 다차원 접근을 통한 공간 데이터는 기하학적인 사상을 갖고 측정을 하며 온라인 분석처리기법을 이용하여 다차원 분석을 수행한다. 공간 측정에는 세가지 타입이 있다(Chevallier and Bedard 1990; Chrisman 1997).

첫째, 비 기하학적 공간 차원(non geometric spatial dimension)으로서 명목척도(nominal scale)에 해당한다. 이 차원은 행정경계 관리의 주기처럼 공간의 수치적인 성질이 요구되지 않으며 지도학의 공간 사상 구분에 해당한다(caron, 1998).

둘째, 기하학-비 기하학적 공간차원(geometric-to-non-geometric spatial dimension)으로 점, 선, 면의 공간 객체를 사용하여 사상으로 표현하고 그에 관련된 속성 값이 해당된다. 사상은 기하학적 X, Y좌표체계를 가지고 있으며 그 속성 값은 비 기하학적 성질을 가진다.

셋째, 기하학적 공간차원(fully geometric spatial dimension)이다. 이 타입은 원 자료가 기하학적 성질을 갖는다. 예를 들어 면 자료로 생성된 지번데이터와 그 데이터를 일반화시 거리는 같다.

수치적(numerical) 측정법에는 수치적인 자료를 포함한다. 예를 들어 공간 데이터 웨어하우스에서 한달간 세입, 물품 구매량이 있다.

2.2 인터넷 기반 공간 자료 활용

웹을 통한 GIS상담 시스템에는 지리적 사상과 속성정보를 제공하며 주제에 따라 다양한 색

채 표현법과 분류법이 요구된다. 또한, 의사결정을 위한 분석가는 지리정보를 필요로 하는 각기 다른 수요에 맞도록 수정, 편집할 수 있는 기능이 필요하다. 특히 어떤 주제에 관해 상세한 정보를 필요로 하는 분석가는 자료의 정확도나 그 주제를 반영하는 축척과 함께 반복되는 분석과정을 위해 초기자료는 손상이 없고, 잘못된 초기 자료는 수정해야 한다. 또한 시간차원으로 인한 지리적 속성변화와 공간 자료 공급의 유통과 변화에 따른 공간자료와 관련된 속성 정보도 공급되어야 한다.

III. 실험적 시스템 설계 및 구현

3.1 시스템 환경

실험적 시스템은 공간 데이터 웨어하우스와 웹 서버가 분리된 분산 환경에서 구현되었다. 우선 공간 데이터 웨어하우스는 수치지도와 영상자료를 저장하기 위한 공간 데이터베이스 서버에 Oracle 8i Database system(8.1.7)에 미들웨어로써 ArcSDE 8.2를 사용하여 공간 자료와 관련된 속성자료를 저장하였다. 웹 환경에서의 개발도구는 인터넷 서버인 IIS, 외부 응용프로그램과 웹 서버와의 상호 연계는 Servlet 엔진인 Jakarta-Tomcat-3.2.3, 공간 자료의 로딩(loading)을 위해 ArcIMS 4.0, 환경은 Java 2 Standard Development Toolkit 1.3.6, 공간 자료 명세서인 메타데이터 문서는 ArcCatalog를 사용하여 XML문서로 작성하였다.

Microsoft SQL-2000의 확장 모듈인 Analysis Services는 구축된 공간 데이터 웨어하우스에 다차원적 접근을 위한 응용 시스템으로 온라인 분석처리에 사용하였다. 연구지역의 공간 자료 생성 및 변형과 분석 도구로는 ArcView GIS 기본모듈, Spatial Analyst등 확장 모듈, 공간 객체의 일반화를 위해 ArcInfo 8.1 Workstation을 포함하는 GIS 프로그램과 위성영상 전처리 과정을 수행하기 위해 EARDAS Imagine 8.5를 사용하였다. 데이터 마이닝 도구로는 SAS 8.2 확장 모듈인 Enterprise Miner 4.0를 사용하고, S+ Spatial Statistics와 S-Plus for ArcView GIS등 통계프로그램을 이용하였다.

3.2 시스템 설계

본 연구는 공간 데이터 웨어하우스의 효율성을 검토하기 위해 응용 시스템을 설계 및 구현하였다. 응용 시스템 설계를 위해 다음과 같은 요구사항을 설정하고 연구를 진행하였다.

첫째, 본 연구에서 실험적 시스템의 구현은 공간 데이터 웨어하우스의 관리적, 분석적 측면을 고려한 실증 분석의 활용을 목표로 한다.

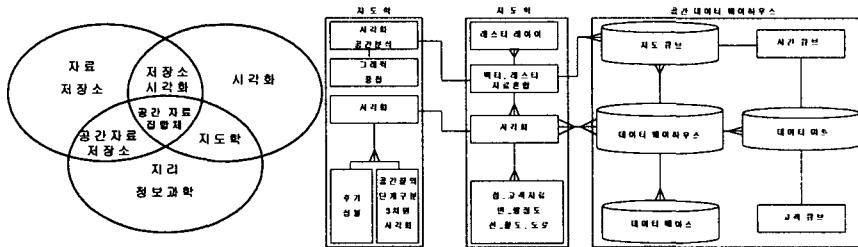
둘째, 본 연구에서 고려한 실험적 시스템의 주 대상은 의사결정을 하는 분석가로 한정하고 시스템의 보안과 인터페이스의 설계는 이 연구에서 배제하였다.

셋째, 본 연구는 웹 환경을 통한 분석가의 의사결정을 지원한다. 그러므로 지리적 사상과 속성정보는 주제에 따라 다양한 표현법과 분류법이 요구된다. 즉, 공간분석의 수요에 맞도록 시스템은 반복되는 분류, 분석을 위해 초기 자료에는 손상이 없어야 하고, 정확하지 않는 자료와 시간에 따라 변형된 지리적 사상과 속성 변화는 수정, 편집할 수 있는 기능을 삽입하였다.

넷째, 본 연구에서 사용하는 수치화된 공간 자료와 비 공간 자료는 연구지역인 서울시의 실 세계를 추상화(Abstraction), 단순화(Simplification)가 개입된 것으로써, 자료의 내용과 형식, 용어와 개념을 정리한 메타데이터 문서를 제공한다.

1) 논리적 설계

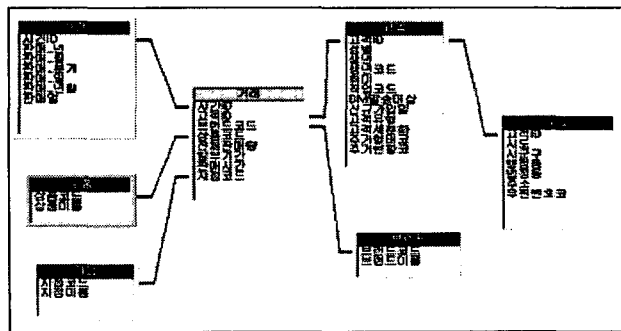
실용적 시스템은 데이터 웨어하우스, 지리정보학, 정보 시각화의 개념적 영역을 가진다(그림 1). 각 영역은 결합관계를 이루고 있는데 데이터 웨어하우스와 지리정보과학의 개념적 결합으로 공간 데이터 웨어하우스, 지리정보과학과 정보 시각화 영역에서 지도학의 공간 자료 시각화, 데이터 웨어하우스와 정보 시각화 영역으로 데이터 웨어하우스의 스키마 모델을 시각화한다. 마지막으로 이 세가지 영역의 결합은 공간 자료 집합체로 개념적 모형을 설명한다.



<그림 1> 공간 데이터 웨어하우징 논리적 체계

2) 물리적 설계

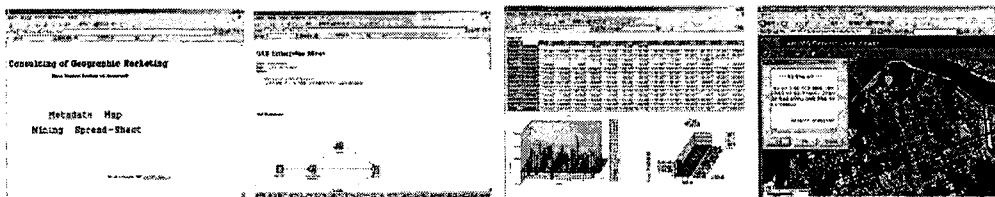
자료 저장소의 물리적 설계는 거래량을 사실 테이블로 설계하였다(그림 2). 이 데이터 마트는 시간, 고객, 브랜드, 상품, 지점을 차원 테이블로 구성하였다. 각 차원의 시간ID, 고객ID, 브랜드 코드, 상품코드, 지점코드를 주키(primary key)로 각 차원들과 다차원 테이블 조인을 수행하도록 구현했으며, 고객 차원은 고객ID 키를 사용하여 주소 차원과 조인이 가능하도록 눈송이 스키마 모델로 설계했다. 그리고 주소차원의 고객ID는 공간 데이터베이스와 연결하여 지리 정보과학에서 제공하는 공간 분석을 수행한다.



<그림 2> 눈송이 스키마 모델(snowflake schema model)

3.3 시스템 구현

본 연구는 구현된 데이터 웨어하우스에 공간 차원을 고려한 공간 데이터 웨어하우스로서 시간, 공간차원에 따라 다양한 수준에서 분석을 지원한다. 따라서 시간차원에서 년, 분기, 월별로 자료를 변형하고 공간 차원은 구별, 행정동으로 변형하였다. 데이터 마트와 고객ID키로 연결된 주소차원은 GIS의 공간 데이터베이스의 점사상의 속성 테이블로 연결하였다. 공간 데이터베이스의 고객 점 사상 위치는 고객의 위치와 위상관계를 담고 있으며 고객 데이터베이스와 연관되어 공간의사결정의 도구로 사용된다.



<그림 3> 시스템 사용자 접근화면

그림 3의 첫 번째는 사용자 접근화면이다. 그리고 사용자 화면에서 메타데이터, 비 공간 자료의 분석을 위한 데이터 마이닝, 분석된 자료를 공간 자료의 형태로 표현하기 위한 인터넷 기반의 수치지도, 공간 사상의 관련 통계자료를 테이블과 차트 형식으로 제공한다. 의사결정과정을 위해 웹 기반 수치지도와 기하 보정된 위성영상은 관심지역의 분석을 수행한다.

IV. 결론

정보기술과 인터넷의 급속한 성장에 따라 자료의 양적 증가는 공공부분과 민간부분에서는 의미있는 정보를 발견하기 위한 정교한 분석이 요구된다. 특히 지리학적 지식발견은 지역의 공간적 특징을 고려하여 이루어지는데 시계열적 변화에 따라 그 특징은 변화한다. 그러므로 지리학적 지식을 발견하기 위해서 공간적, 시간적 차원을 고려한 관계, 방향, 연결성은 적절한 지도학적 변형과 더불어 주제별 상이한 측정법이 필요하다. 따라서 본 연구에서는 대량의 공간 자료를 분석하기 위한 공간 데이터 웨어하우스의 이론적 개념을 정리하고, 실험적 응용 시스템을 설계 및 구현하여 효율성을 검토하였다. 실험적 시스템을 통해 얻은 결과는 다음과 같다. 공간 자료를 위한 저장소는 구축 목적, 배경, 대상 사용자를 대상으로 자료 공유를 중심으로 하는 관리적 측면과 공간 데이터 웨어하우스에 시간, 공간 등 다차원적 접근을 시도하여 분석을 수행하는 분석적 측면을 고려했다. 이 측면은 실험적 시스템에서 기존 웹 기반 지리정보과학에서 제공하는 고정적인 위치, 관련속성 정보제공에서 분석가와 관리자사이의 양방향 통신이 가능하다. 특히, 주제에 따른 다차원 분석은 특정 차원별로 세분화, 일반화가 가능하며 의사결정을 위한 지식을 발견한다. 본 연구에서 분석적 측면의 사례분석은 군집분석의 k-평균 군집(k-means clustering), 최근린 계층 군집(nearest neighbor hierarchical clustering)을 사용하고, 관련 속성의 밀집과 분포패턴의 공간 이웃을 정의하는 티센 다각형(thiessen polygon)을

생성하여 지식을 탐색하였다. 또한 특정지역에서 발생하는 사건의 연관규칙을 설명하기 위해 기존의 자료를 사용하여 지역에서 발생하는 사건의 패턴을 발견하여 지식을 발견할 수 있다.

참고문헌

- 1) Lee, J. and Wong, W.S., 2001, "Statistics analysis with ArcView GIS", pp.5-56.
- 2) Han, J., 2002, "Spatial Data Warehousing and Design for Enterprise GIS", DBMiner Technology, pp.5-8.
- 3) Oracle, 2003, "Slowly Changing Dimension with Oracle 9i warehouse Builder: An Oracle White Paper", pp.5-23.
- 4) Boyles, D., 2002, "GIS means Business: Marketing Communication, Targeting Marketing, Market Reserach, Internet Map Publishing", ESRI Press, p.18