

석사학위논문

공간적 의사결정을 위한  
공간 데이터 웨어하우스 설계 및 활용

지도교수 황 철 수

경희대학교 대학원  
지리학과

박 지 만

2004년 2월

공간적 의사결정을 위한  
공간 데이터 웨어하우스 설계 및 활용

지도교수 황 철 수

이 논문을 이학 석사 학위논문으로 제출함

경희대학교 대학원  
지리학과 GIS전공

박 지 만

2004년 2월

박지만의 이학 석사학위 논문을 인준함

주심교수\_\_\_\_\_□□

부심교수\_\_\_\_\_□□

부심교수\_\_\_\_\_□□

경희대학교 대학원

2004년 2월

## <목 차>

1장 서론 .....	1
1.1 연구배경 및 목적 .....	1
1.2 연구내용 및 방법 .....	3
1.3 논문의 구성 .....	6
2장 이론적 배경 및 개념 .....	8
2.1. 데이터 웨어하우징 .....	8
2.1.1 데이터 웨어하우징의 개념 .....	8
2.1.2 데이터 웨어하우스 .....	9
2.1.3 기본 구성 .....	12
2.1.4 온라인 분석처리 .....	13
2.1.5 데이터 마트 .....	15
2.2 공간 데이터 웨어하우징 .....	16
2.2.1 공간 데이터 웨어하우스 .....	16
2.2.2 공간 자료 집합체 .....	18
2.2.3 지식발견 논의 .....	20
2.3 인터넷 기반 공간 자료 활용 .....	22
2.3.1 지리정보학의 개념 .....	22
2.3.2 지리정보학을 활용한 의사결정 .....	22
2.3.3 지리학적 지식발견에 관한 방안 .....	23
2.3.4 인터넷 기반 공간 자료 저장소 .....	26

2.4 웹 기반 공간 데이터 웨어하우징의 활용 사례 및 연구 .....	27
2.5 기존 연구의 한계와 대안 .....	29
2.5.1 기 개발된 인터넷 기반 GIS 지리정보서비스 .....	29
2.5.2 웹 환경 공간 데이터 웨어하우스 스키마설계 .....	30
2.5.3 공간의사결정을 위한 다차원 모델링 대안 .....	31
<b>3장 실험적 시스템의 설계와 구현 .....</b>	<b>34</b>
3.1 실험적 시스템의 구현 요소 .....	34
3.1.1 시스템 환경 .....	34
3.1.2 백화점 공간 데이터 웨어하우스 구성 방안 .....	35
3.1.3 지식 탐색 및 발견을 위한 방안 .....	35
3.1.4 공간적 의사결정을 위한 인터넷 기반 상담 시스템 .....	37
3.2 실험적 시스템 설계 .....	38
3.2.1 요구사항 분석 .....	38
3.2.2 시스템 설계 .....	39
3.2.2.1 논리적 설계 .....	39
3.2.2.2 물리적 설계 .....	42
3.3 자료 생성 및 변환과 시스템 구현 .....	44
3.3.1 전처리 과정 .....	44
3.3.2 메타데이터 작성 .....	47
3.3.3 웹 기반의 시스템 구현 방안 .....	48
<b>4장 사례분석 .....</b>	<b>50</b>
4.1 카드 구매 고객의 공간적 분포특성 .....	50
4.1.1 전역적 공간분포패턴 .....	50

4.1.2 공간적 고객 세분화 .....	53
4.2 카드 고객의 구매 특성 .....	56
4.2.1 카드 고객의 일반 현황 .....	56
4.2.2 시공간적 거래패턴 .....	58
4.2.3 목표 마케팅 지역의 다차원적 분석 .....	60
4.2.4 평가 .....	63
5장 결론 .....	64
참고문헌 .....	69
Abstract .....	77

## <표 목차>

표 2.1 데이터 웨어하우스의 활용 .....	12
표 2.2 온라인 분석처리와 온라인 거래처리의 비교 .....	14
표 2.3 온라인 분석처리기법의 종류 및 특징 .....	15
표 4.1 고객 세분화에 따른 표준편차 타원체 .....	53
표 4.2 고객의 성별 자료 분석 .....	58
표 4.3 주거 형태 자료 분석 .....	58
표 4.4 지점별 거래량 .....	59
표 4.5 분기별 거래량 .....	59
표 4.6 일별 거래량 .....	59
표 4.7 시간에 따른 거래량 .....	59

## 〈그림목차〉

그림 1.1 연구의 흐름도 .....	7
그림 2.1 데이터 웨어하우스의 개념도 .....	10
그림 2.2 데이터 웨어하우스 구성도 .....	13
그림 2.3 다차원 성형 모델 .....	17
그림 2.4 공간 자료 집합체 .....	20
그림 2.5 지식발견 과정의 개관 .....	21
그림 2.6 의사결정 단계 .....	23
그림 3.1 실험적 시스템의 개념적 영역 .....	40
그림 3.2 공간 데이터 웨어하우징 논리적 체계 .....	41
그림 3.3 백화점 자료 논리적 설계 .....	42
그림 3.4 백화점 자료 눈송이 스키마 모델 .....	43
그림 3.5 연구지역의 A백화점 카드 고객 위치 .....	45
그림 3.6 영상자료 전처리 과정 .....	46
그림 3.7 메타데이터 문서 .....	47
그림 3.8 실험적 시스템 소프트웨어 흐름도 .....	48
그림 3.9 사용자 접근 화면 .....	49
그림 4.1 고객의 공간적 분포 패턴 .....	51
그림 4.2 공간이웃과 군집분석 .....	54
그림 4.3 목표 마케팅 지역과 각 지점 거리별 버퍼존 .....	55
그림 4.4 각 지점과 거리별 고객 수 .....	56
그림 4.5 구매 고객의 일반 현황 .....	57
그림 4.6 시간대별 거래고객 수 .....	59
그림 4.7 목표 마케팅지역의 다차원적 거래특성 .....	60
그림 4.8 목표 마케팅지역 다차원 분석을 위한 구조화된 질의 언어 .....	61

그림 4.9 초우량 1등급 고객 구매특성 .....	61
그림 4.10 초우량 1등급 고객구매 분석을 위한 다차원 접근 .....	62
그림 4.11 구매 브랜드의 연관 규칙 .....	63

# 1장 서론

## 1.1 연구배경 및 목적

급속하게 발전한 인터넷 환경과 자료의 증가는 수치화된 지리정보의 폭발적인 증가를 가져왔다(Miller, 2000). 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, 공간의 위상관계를 사용하는 멀티미디어 자료, 위치를 인식하는 장치와 그것을 이용한 무선통신 장비, 고 해상(high spatial and spectral resolution)의 위성영상 출현 및 방대한 양의 저장소는 이러한 현상을 가중시키고 있다(Miller et al, 1999). 또한 인터넷 사용자가 각각 개인적인 움직임을 시간, 공간적으로 추적하는 것이 가능하게 되었다. 이와 같이 공간 자료를 표현하고 저장 및 검색을 지원하는 데이터베이스를 공간 데이터베이스(spatial database)라 한다.

그러나 공간 데이터베이스는 실세계의 여러 차원을 고려하지 않았고 정보 분석을 공간적인(spatial)것에 한정하였다(Miller, 1999). 실세계의 여러 가지 차원을 고려한 데이터베이스 모델은 다차원적(multi-dimensional) 정보 분석을 요구하고, 모델의 구현, 처리방식, 그리고 처리과정은 정보 분석의 주제마다 다르다.

기존의 지리 자료를 생성하고 사용하는 공공, 민간단체는 정부의 정책에 따라 성공과 실패에 커다란 영향을 받았다. 그러나 각종 규제철폐와 세계화로 인한 확장된 정보 중심 사회에서는 분석을 위한 새로운 접근법을 지원하는 개방적인 방법과 해결책이 요구되고 있다. 구체적으로 특정 지역의 문화와 경제에 따른 전문적인 공간 의사결정은 경쟁력을 강화시키는데 필수적인 연구가 되었다. 위의 기술한 내용을 토대로 다차원의 공간적 속성과 변수를 고려하는 공간 데이터 웨어하우스<sup>1)</sup>가 필요한 것이다. 공간 데이터 웨어하우스는 분석가의 의사결정을 지원하기 위해 축적된 많은 자료를 분석 관점에서 주제별로 통합하여 운영시스템(operation system)과 분석가 사이의 별도 저장소에 저장해 놓은 데이터 웨어하우스로 지리학적 지식을 발견하기 위해 공간적인 특징을 고려한 데이터베이스다(Rawlings, 1997).

지리학적 지식발견은 관심 있는 지역의 인구, 산업 특성과 공간적 특징을 고려하여 이루어지는데, 그 특징은 공간적 자기상관성(spatial autocorrelation)과 공간적 이질성(spatial heterogeneity)이라는 성격을 가지며, 시공간적(spatio-temporal)관련성은 더욱 복잡하게 연관되어 있다. 그러므로 각 변수에 따른 관계, 방향, 연결성은 차원에 따라 다르고, 다양한 지리학적 차원에서 주요한 지식발견을 위해 시간, 공간적으로 추출해야 한다. 이러한 맥락에서 분석을 위한 거리측정은 각 주제에 따라 다르게 하고, 지도학적 변형은 적절한 시간에 흥미 있는 패턴을 발견하여 의미 있는 지식을 얻는데 이용된다. 결과적으로 발견된 지식은 다양한 수준<sup>2)</sup>의 측정법을 이용하여 효과적인 의사결정 도구로 구체화 된다. 즉, 대용량의 지리공간

---

1) Spatial Data Warehouse : ‘공간 데이터 웨어하우스’ 라고도 하며 의사결정을 지원하는 Data Warehouse와 공간요소를 통합하여 각종 분석을 하는 시스템으로 온라인 분석 처리기법을 확장하는 공간 분석기능이 포함된다.

2) 수준 (level) : 계층구조상의 거리나 위치를 나타낸다. 계층구조의 동일한 거리를 갖는 항목들의 집합을 의미하며 최대치가 동일할 경우 같은 수준에 속한다고 말한다.

자료는 정보 분석 자료로 활용되어 지리학적 지식을 발견하고, 기업이나 공공단체의 의사결정을 지원한다(UCGIS, 2003).

본 연구는 대용량의 공간 자료에서 의미 있는 지식을 탐색, 발견하고 의사결정과정을 지원하는 다차원적 공간 데이터 웨어하우스의 모델을 설계하고 구현하는 것이 목적이다. 데이터 웨어하우스는 실제 백화점 카드 고객 자료의 거래별 시-공간(spatio-temporal)적 특징과 신상자료를 활용하였다. 그리고, 사례분석을 통해 실험적으로 구현된 공간 데이터 웨어하우스에 다차원적 접근을 시도하여 고객의 인구속성, 구매 패턴 등을 연구하고, 서울시를 대상으로 목표 마케팅(target marketing) 지역을 예측하여 시스템의 효율성을 검토하였다. 또한 본 연구는 기존의 웹 환경의 공간 자료 활용 측면에서 고정된 사상을 제공하는 방식을 배제하고 분석가를 위한 양방향 통신을 제공한다. 더불어 소매업종의 민간 기업측면에서 공간자료를 활용하여 고객 속성에 대비한 시공간적인 차별적 마케팅 전략수립의 모델을 제시한다.

## 1.2 연구 내용 및 방법

본 연구에서는 분석가의 의사결정을 지원하는 공간 데이터 웨어하우징 기법과 실증분석을 위한 사례로서 문헌연구를 통해 이론적 기반을 마련한다. 그리고 공간 데이터 웨어하우스 설계 및 구현, 데이터 마이닝 기법 중 군집기법, GIS시각화 방법을 통해 얻은 지식을 기반으로 실제 서울시의 목표 마케팅 지역을 예측하기 위해 다음과 같은 절차에 따라 연구를 수행하였다.

첫째, 공간데이터의 가용성과 호환성을 가지고 공간 특성을 고려한 분석가의 공간분석을 위해 공간 데이터 웨어하우스의 각 단계별 프로세스를 설계하였다. 이를 통해 목표 마케팅 지역 예측을 위한 공간 데이터 웨어하우스를 구현하여 시스템의 개념적 기틀을 제공하였다. 둘째, 구현된 시

시스템을 통해 데이터 마이닝 기법 중 군집분석을 통해 의사결정시 요구되는 지식을 발견하도록 온라인 분석처리기법(OLAP : online analytical processing)을 이용하여 다차원적인 접근을 시도하였다.

본 연구는 주제에 따라 자료를 분류하고 실험적 시스템을 구현하는 단계와 효율적이고 합리적인 의사결정과정을 가능하게 하는 지식발견 단계로 구분할 수 있다. 전자의 단계에서는 특정 주제에 맞는 자료를 선별하여 각 차원<sup>3)</sup>에 따라 분류하고 정제하여 데이터마트<sup>4)</sup>를 설계한다. 그 이후 개별적인 각각의 데이터마트는 하나의 공간 데이터 웨어하우스의 요소가 되며 하나의 시스템으로 구성된다. 각 데이터 마트에는 서울시 백화점 고객 자료를 공간적 차원으로 동별, 구별로 분류하고 시간적 차원으로는 월, 분기, 년도별로 구분하여 시계열 분석이 용이하고 분석에 있어 요약, 세분화가 가능하도록 하였다. 후자의 단계에서는 구현된 공간 데이터 웨어하우스에서 데이터 마이닝의 탐색적 자료 기법 중 하나인 군집분석의 k-평균 군집방법(k-means clustering), 계층적 최근린 군집기법(nearest neighbor hierarchical clustering)을 활용하여 유사한 특성을 가진 몇몇의 집단으로 그룹화하여, 각 군집의 특성을 파악하고 지식과 패턴을 분석하여 서울시에서 목표 마케팅지역을 예측하는데 활용했다.

목표 마케팅 지역을 예측하기 위한 공간분석 수행방법은 공간분석기법에서 새롭게 대두되는 공간 데이터 마이닝의 여러 기법 중 군집분석을 사용했다. 군집분석은 대량의 자료에서 주어진 내부구조의 사전 정보 없이 의미 있는 자료구조를 발견하고 이후 모형화, 시각화할 경우 높은 성능을

---

3) 차원 (dimension) : 분석가의 관점에서 다른 항목과 독립적인 일련의 모든 항목들의 집합을 말하며 분석의 대상이 되는 변수차원과 목표의 관점항목을 가지는 유형차원으로 구분한다.

4) Data mart : 분석가들을 위해 특정 주제에 맞는 데이터 저장 공간을 말하며 그 범위는 선택된 주제에 한정되며 특별한 영역에 중점을 두어 구현되는 데이터 웨어하우스의 일부라고 할 수 있다.

발휘한다. 이를 통해 구매고객의 속성과 거래패턴을 파악하고 마케팅 지역을 예측하였다. 마지막으로 발견된 지식을 효율적인 의사결정에 반영하고 이를 설명하기 위해 GIS에서 제공하는 공간 분석과 정보시각화 기법으로 표현하고 설명하였다. 또한 공간 데이터 웨어하우스의 효율성 제고와 호환을 위해 메타데이터를 구현하였다.

설계된 공간정보 객체를 저장하는 공간 데이터 웨어하우스 구현은 Oracle 8i(8.1.7)와 미들웨어<sup>5)</sup>로서 ArcSDE 8.2를 사용하고 실험적 시스템에 다차원 접근을 위해 MS-SQL 2000의 확장 모듈인 Analysis Services를 사용하여 온라인 분석처리를 수행하였다. 연구지역의 공간 자료 생성 및 변형과 분석 도구로는 ArcView GIS 기본모듈, Spatial Analyst 등 확장 모듈, 공간 객체의 일반화를 위해 ArcInfo 8.1 Workstation을 포함하는 GIS 프로그램과 래스터 자료의 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 GIS 등 통계프로그램을 이용하였다.

최종 분석가의 지식발견을 위한 시스템의 접근과 데이터의 호환을 위해 인터넷 환경에서 서버 사이드에 ArcIMS 4.0, IIS, 서블릿 엔진(servlet engine) 인 Jakarta-Tomcat-3.2.3과 시스템 개발 환경으로는 Java 2 Standard Development Toolkit 1.3.6을 사용하였다. 또한 자료의 호환성을 위한 메타데이터 구현은 FGDC의 ‘디지털 지리 공간 데이터에 대한 메타데이터 표준’<sup>6)</sup>에 따라 중앙집중식 모형방법으로 제공하였다.

연구 대상으로는 1개월 시간 간격의 2000, 2001년 백화점 고객데이터 170만건 중 카드로 물품을 구매한 고객 41,263명의 신상 자료를 대상으

---

5) Middleware : 자료 또는 미리 요약한 자료 등을 이동하기 위해 두 시스템을 연결하는데 사용되는 하드웨어 또는 소프트웨어.

6) FGDC Content Standard for Digital Geospatial Metadata, 1992

로 하고 예측지역으로는 서울시로 국한하였다.

### 1.3 논문의 구성

본 논문은 총 다섯 개의 장으로 구성된다. 1장에서는 연구의 수행배경, 목적, 방법 및 내용에 대해 서술하였다. 2장의 전반부는 공간 데이터 웨어하우징, GIS를 통한 의사결정의 조사단계로서 문헌을 통해 이론적 배경을 검토하고 기존의 공간 데이터 웨어하우징의 개념을 재정립하였다. 후반부는 웹 환경에서 지리학적 지식발견을 위한 공간 자료 활용의 기존 사례의 한계와 대안을 설명하였다. 3장에서는 재정립된 이론적 배경을 기반으로 실험적 시스템을 설계 및 구현하였다. 4장은 구현된 시스템을 통해 목표 마케팅 지역을 설정하기 위하여 실제 백화점 자료를 활용하여 분석하였다. 5장에서는 이상의 연구내용과 결과를 종합하고 연구의 한계를 제시한다.

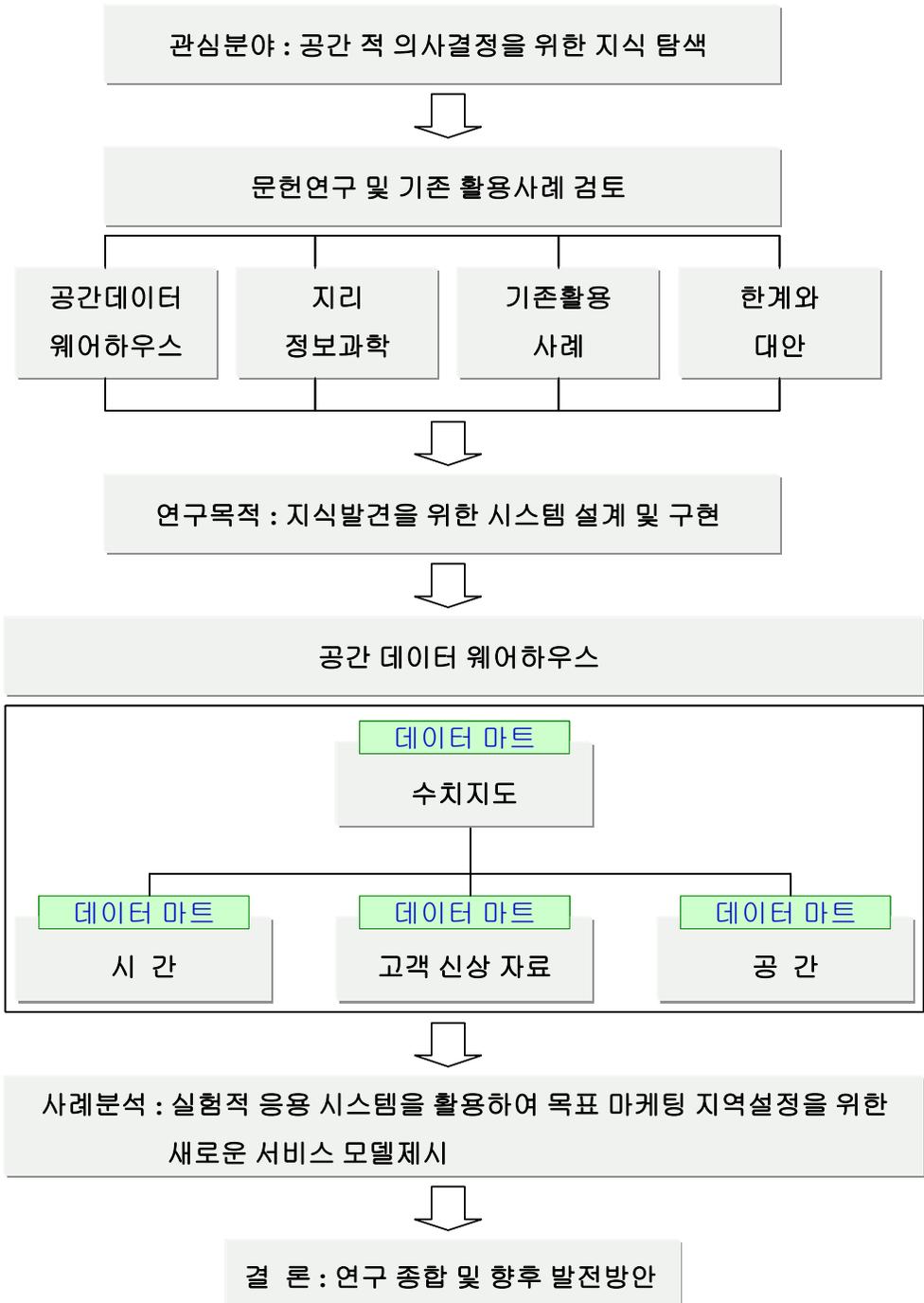


그림 1.1 연구의 흐름도

## 2장 이론적 배경 및 개념

### 2.1 데이터 웨어하우징

#### 2.1.1 데이터 웨어하우징의 개념

과거의 의사결정은 각 부서별로 개별적이고 단편적인 필요성에 의해 산발적으로 의사결정시스템<sup>7)</sup>이 구축되었다. 이러한 원인으로 자료들은 통합적이지 못하고 다양한 시스템 상에 산재되어 존재했다. 즉, 각각의 정보시스템은 조직이라는 제한을 받게 되며, 개별적이고 독립적인 시스템으로 다루어졌다(Oracle, 2001). 의사결정시스템에는 여러 개의 시스템 상에 유사한 자료가 존재하고, 같은 이름을 가진 개체가 서로 다른 의미를 나타내거나, 동일한 개체가 다른 이름으로 표현되기도 한다. 각각의 시스템 상에서 자료는 그 의미와, 표현상의 차이를 갖고 있으며 이러한 문제점은 결국 자료 신뢰성과 생산성, 정보생성의 측면에서 여러 장애를 가져왔다. 이러한 맥락에서 과거의 많은 의사결정시스템이 기대만큼 효과적으로 활용되지 못했으며, 축적된 자료는 기업의 중요한 의사결정과정에 사용되지 못했다 (Peterson, 1998).

이러한 기존의 의사결정지원 시스템의 한계를 극복하기 위해 나타난 개념이 데이터 웨어하우징(data warehousing)이라고 할 수 있다. 데이터 웨어하우징은 데이터 웨어하우스(data warehouse)를 구축하고 활용하는 일련의 과정으로, 전체적인 시스템 관점에서 의사결정을 지원하기 위한 환경을 구현

---

7) DSS (Decision Support System) : 준 구조적 더 나아가 비구조적인 문제를 해결하는데 의사결정자가 자료와 모형을 활용하는 것을 돕는 상호작용적인 컴퓨터 시스템을 말하며 사용되는 방식은 원형개발접근법으로 가능한 빠른 시간 내에 사용자의 기초적인 요구를 최대한 반영한 소규모 모형을 구축한 후 이를 점진적으로 개선해 나가는 시스템 개발방식을 사용한다. 여기서 말하는 시스템은 정보를 제공하는 응용 프로그램을 의미하는 것으로서, 일상적인 업무 운영을 통해 데이터를 수집하는 등의 운영 시스템과는 구별된다(조재희, 1999).

하는 것이다(Kamber et al., 1998). 데이터 웨어하우스는 분석가의 의사결정을 지원하기 위해 축적된 많은 자료를 분석가 관점에서 주제별로 통합하여 별도의 장소에 저장해 놓은 데이터베이스이다.

### 2.1.2 데이터 웨어하우스

데이터 웨어하우스는 의사결정을 위한 주제 지향적이고 통합적이며 시계열성을 가진 비 휘발적인 자료 집합이며 저장 공간을 제공한다(Inmon, 1994; Kamber, 2001). 이러한 저장 공간은 사용자가 원할 때는 언제나 자료를 조회할 수 있도록 중앙의 저장소에 위치한다. 그리고 하나의 중앙 집중적인 데이터 웨어하우스를 자료 원본으로서 이용하는 것은 조직 내에 자료의 품질과 일관성을 증가시킨다(Devline, 1996). 왜냐하면, 최종적인 자료 검증 절차는 거의 하나의 시스템에 의해 처리되기 때문이다. 대량의 자료처리 작업을 하는 대부분의 주요 공공기관이나 기업은 데이터 웨어하우스 모델을 필요로 한다. 그러나 이 자료를 관리하는 조직은 자료가 너무 일반적이거나 혹은 그 반대로 다루기 어려워 특정 또는 특화된 의사 결정의 지표로 삼을 수 없었다(Kimball and Merz, 2000).

또한 데이터 웨어하우스가 자료를 구조화하는 작업을 수행함에도 불구하고, 기반 자료로부터 유용한 정보를 추출하기 위해서는 데이터 마이닝을 필요로 한다. 이러한 데이터 마이닝 분석을 수행하기 전에 어떠한 유형의 자료가 데이터 웨어하우스에 저장되고 구조화되어야 한다.

데이터 웨어하우스의 장점 중 하나는 분석을 수행하는 설명 또는 발표하는 자료와 온라인 거래 처리(OLTP : online transaction processing)<sup>8)</sup>자료를

---

8) OLTP : 원시 데이터가 실제로 발생하고 기록되는 시스템으로 무엇(what)에 초점을 맞추고 있다. 이 시스템의 일차적인 목표는 현재 거래 상태를 정확하게 기록하고 갱신할 수 있도록

분리할 수 있다는 것이다. 이러한 처리과정을 분리하는 것은 동일한 자료 처리과정에서 대부분 다른 결과를 얻기 위해 다른 자료 구조를 사용하며, 그 구조는 어떻게 자료를 추출하고, 질의응답을 하는지에 대한 해결책을 제시한다. 이와 같은 구조는 간단한 질의로 다 차원적이고 복잡한 응답을 얻을 수 있으며, 필요한 발표나 설명을 위해 자료를 어떻게 구조화할 것인가 결정할 때 분석가는 사용자 중심의 편리함과 단순성을 고려할 수 있다.



그림 2.1 데이터 웨어하우스의 개념도

그림 2.1은 데이터 웨어하우스가 의사결정을 지원하기 위해 운영적 데이터베이스(operational database)와 구분된 데이터베이스이며 분석 관점에서 사용자에게 통합되고 단일화된 플랫폼임을 보여준다. 데이터 웨어하우스의 특징은 다음과 같다(Kimball, 1996; Rawlings and Kucera, 1997; Chaudhyri and Dayal, 1997; Han, 1998; Han and Kamber, 2001; Microsoft, 2001; Oracle, 2003)

첫째, 주제 중심적(subject-oriented)이다. 전통적인 데이터베이스는 운영 시스템의 특성에 따라 특화된 기능을 제공하는데 반해, 데이터 웨어하우스는 주제를 중심으로 그 주제와 관련된 자료들로 구성되며, 일상적인 운영이나 처리과정(transaction processing)에 대한 것이 아니라, 의

---

하는 것이며 일상적인 운영을 지원하고 은행의 창구 업무, 항공사의 예약업무등을 예로 들 수 있다.

사결정자를 위하여 자료를 모델링하고 분석하는데 초점을 둔다.

둘째, 통합적(integrated)이다. 전체적인 시스템 관점에서 운영시스템으로부터 공급되는 다양한 자료와 외부정보로부터 데이터 정제와 통합기술이 적용되어 일관성을 유지하며 신뢰할 수 있는 하나의 시스템으로 통합한다.

셋째, 시간성(time variant)을 가진다. 과거와 현재의 자료를 동시에 유지하여 시간성을 가지고 있다. 기존의 운영 데이터베이스는 현재를 기준으로 해서 정확한 값을 가지는 반면 데이터 웨어하우스는 역사적인 관점(historical)에서 과거의 시간 항목을 포함하고 있다.

넷째, 비 휘발적(non-volatile)이다. 물리적으로 분리되어 있으며 각 시스템 사이의 상이한 자원사용 패턴을 나타낸다. 운영 시스템에서 시간성 자료는 비 휘발성 자료의 집합을 말한다. 또한 다양한 이(異)기종의 데이터베이스에 대한 통합의 관점에서 데이터 웨어하우스는 먼저 자료를 통합 후 웨어하우스에 저장하므로 매우 유용하다. 데이터 웨어하우스의 활용측면은 표 2.1과 같다.

표 2.1 데이터 웨어하우스의 활용

산업	활용의 기능적 영역	전략적 활용
항공	운영과 마케팅	승무원 지정, 항공기 배치, 요금 혼합, 항로에 대한 이윤 분석, 빈번한 비행 프로그램 촉진
의류	유통과 마케팅	효과적인 판매촉진 및 보충
은행	제품 개발, 마케팅	고객 서비스, 추세 분석, 제품과 서비스 촉진
신용카드	제품 개발과 마케팅	고객 서비스, 요금에 대한 새로운 정보 서비스, 부정수단으로부터의 보호
투자&보험	제품개발, 운영 및 마케팅	위험관리, 시장 이동 분석, 고객의 동향 분석, 포트폴리오 분석
건강 보조	유통과 마케팅	유통 결정, 제품 촉진, 판매 결정, 가격 책정
공공부문	운영	지능형 수집
소매 유통	유통과 마케팅	추세분석, 구매패턴분석, 가격책정, 재고관리, 판매촉진, 최상의 유통채널
철강	생산	패턴분석
통신	제품개발, 운영 및 마케팅	신제품 서비스 촉진, 이윤 분석

자료 : Kimball(1996)

### 2.1.3 기본 구성

데이터 웨어하우스는 이기종(heterogenous)의 다른 운영시스템과 별도로 구축되며 분석가의 필요와 수요에 따라 자료를 정제하여 데이터 웨어하우스에 적재된다(그림 2.2). 그래서 ‘협력적 구성’(corporated architecture) 또는 ‘포괄적인 구성’(generic architecture)이라 표현된다(Poe, 1995; Weldon, 1997). 이 과정은 ETT(Extraction, Transformation, Transportation)<sup>9)</sup>프로세스를 통해 추출되고, 데이터 웨어하우스 서버에 적재되어 통합된 데이터는 다차원 모델링(Multi-Dimension Modeling)<sup>10)</sup>이

9) 운영 데이터베이스 시스템에서 데이터 웨어하우스로 자료를 가지고 오는 일련의 프로세스를 의미하며 ETL(Extraction, Transformation, Loading)이라고도 한다.

10) 분석가의 요구 관점에 따라 발생하는 다차원 분석을 위한 것으로, 기본적인 데이터를 테이블에

적용된다.

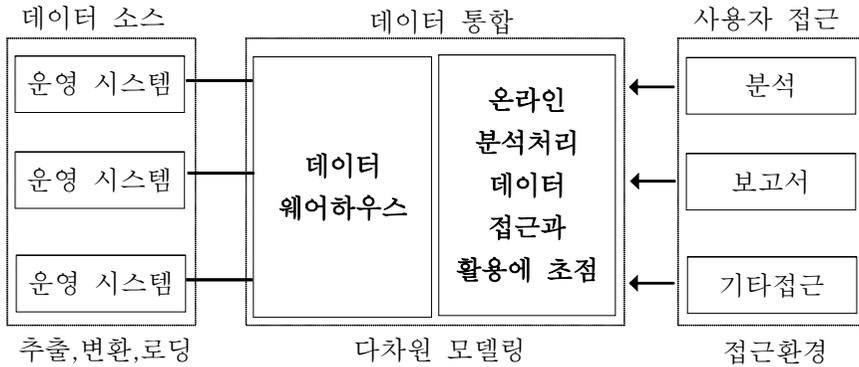


그림 2.2 데이터 웨어하우스 구성도

모델링이 되어있는 서버에, 분석가는 온라인 분석처리 또는 다양한 수준과 기법으로 접근한다. 데이터 웨어하우스의 접근환경은 온라인 분석처리 기법과의 관계를 표현한다. 온라인 분석처리 기법은 정보의 효과적인 활용 측면에 초점을 맞춘 것으로 자료 접근을 강조한 기법이다. 데이터 웨어하우스의 방대한 정보의 가치를 인식하고 활용의 측면을 고려하면서 온라인 분석처리 기법은 데이터 웨어하우징 환경으로 통합된다.

#### 2.1.4 온라인 분석처리

온라인 분석처리(OLAP : online analytical processing)는 온라인 거래 처리(OLTP : online transaction processing)와 상대적인 개념으로, 기존 저장소의 주요 작업이 자료 저장 및 저장된 자료 처리면, 온라인 분석처리는 저장된 자료의 분석으로 실제 유용한 정보를 생성하기 위해 사용된

---

저장한 후, 각 주제별 하나의 차원을 설정하고 차원에 맞는 데이터를 저장하는 것을 말한다(정보통신부, 1999).

다(OLAP Council, 1995). 온라인 분석처리 기법은 기존의 온라인 거래 처리(OLTP : online transaction processing)의 정형화된 양식의 자료 생성 작업이나 간단한 질의가 수행되는 정적 모델로 설명하고, 정적 모델과 구분되는 동적 모델로서 정보를 생성, 조작, 활성화 및 종합하는데 필요한 역동적 분석기법이라 정의한다(Codd, 1993). 즉, 온라인 분석 처리 기법은 최종 분석가나 사용자가 다차원 정보에 직접 접근하여 대화식으로 정보를 분석하고 지식을 발견하여 의사결정에 활용하는 과정으로, 그 특징은 표 2.2와 같다.

표 2.2 온라인 분석처리와 온라인 거래처리의 비교

구분	온라인 거래처리	온라인 분석처리
대상업무	운영업무	정보 생성 및 분석 처리
사용자	사용자계층	관리자, 분석가, 지식계층
자료 성격	업무 기능 위주	자료 중심 구성
자료의 갱신	실시간 갱신	주기적 갱신
자료 보유기간	단기간	장기간
사용자 요구사항	사전에 정의가능	사전에 정의하기 어려움
질의의 성격	정형적	비정형적

자료 : Han and Kamber, 2001, "Data Mining Concept and Techniques", p.43

첫째, 분석을 위해 사용할 정보의 형태는 다차원적(multi-dimensional)이다. 기본적으로 업무구조나 정보의 형태는 정보의 가치를 가지고 정보의 다양한 각도에서 비교해야만 가능하고 대부분 단순히 표현할 수 없다.

따라서 데이터 항목들 사이에 내재된 상호 연관성이 높을수록 이러한 상관관계에 대한 분석은 가치 있는 정보가 될 것이며, 이러한 정보를 데이터베이스로 구조화하기 위해서는 다차원적 구조체가 필요하게 된다.

둘째, 분석가는 대화식 정보를 분석할 수 있어야 하고, 의사 결정을 지원받을 수 있어야 한다. 정해진 질의를 넘어 대화식(interactive) 정보 접근에 의한 처리는 분석가의 연결 흐름에 끊이지 않고 시스템과 상호작용을 통해 분석하며 원하는 결과를 얻을 때까지 계속해서 질의할 수 있는 방법을 제시한다. 이러한 질의 결과는 주제별 전반적인 상황을 파악하고 의사결정을 지원할 수 있다. 온라인 분석처리기법의 종류 및 특징은 표 2.3과 같다.

표 2.3 온라인 분석처리기법의 종류 및 특징

종 류	특 징	장 점
ROLAP	관계형 DBMS와 분석가 사이에 위치하는 중계 서버	관계형 DBMS에 이용되며 안정적
MOLAP	배열형태의 다차원적 저장엔진	요약데이터를 사용하여 빠른 계산능력
HOLAP	ROLAP과 MOLAP장점 혼합	안정적이고 빠른 계산 능력

### 2.1.5 데이터 마트(Data mart)

데이터 마트는 사용자 중심의 특화된 저장소로서, 동질적인 사용자 집단에게 유사한 모델로 구현하여 데이터 가용성에 초점을 맞춘다. 데이터 웨어하우스와 사용자 사이에 데이터 마트가 존재해야 하는 이유는 첫째, 데이터 웨어하우스는 최종 사용자와의 인터페이스보다는 대량의 자료를 효율적으로 통합하고 관리하는 측면에 초점을 맞추고 있다. 그래서 사용자 측면에서 편리한 형태로 설계되지 않을 수 있다. 둘째, 데이터 웨어하우스는 전체적인 시스템 용도로 구축되기 때문에 각 개별부서나 사용자 집단에 적합한 형태로 자료가 저장되지 않는다. 따라서 사용자에게 최적의 성능을 제공하지 못할 수 있다. 셋째, 분석가는 데이터 웨어하우스의

전체 자료 중 일부분만을 사용할 것이다. 분석가들이 데이터 웨어하우스에 직접 질의를 수행하는 것은 많은 시스템 자원을 필요로 하며 전체 시스템 성능에 부하를 줄 수 있다.

## 2.2 공간 데이터 웨어하우징

공간 데이터 웨어하우스는 공간 과학에 데이터 웨어하우스를 적용한 공간 자료 집합장소이다. 이 저장소는 공간의사결정을 위한 지식발견의 개념으로 공간 자료나 차원을 적용한다. 본 연구에서 공간 차원을 고려한 데이터 웨어하우징은 공간 의사결정과정의 일부이며, 이를 공간 데이터 웨어하우징이라 한다.

### 2.2.1 공간 데이터 웨어하우스

공간 데이터 웨어하우스는 공간, 비공간 자료가 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 위상 관계, 측정단위, 공간 해상도를 포함하며, 일반적인 데이터 웨어하우스와 유사하게 통합된, 주제 지향적이며, 시간성을 갖고, 비 휘발성의 공간, 비공간 자료를 보유하는 저장소라고 정의한다(Bedard, 1999). 공간 데이터 웨어하우스로 다차원 접근을 통한 공간 데이터는 기하학적인 사상을 갖고 측정을 하며 온라인 분석처리기법을 이용하여 다차원 분석을 수행한다. 다차원 모델링은 시간, 공간을 포함한 여러 차원을 사용하여 분석한다.<sup>11)</sup> 특히, 분석방법은 주제에 따라서 차원을 세분(drill-down)<sup>12)</sup>하거나, 요약

---

11) pivoting : 분석가는 데이터의 차원을 변경하며 분석을 수행한다. 이를 pivoting이라 한다.

12) drill-down : 한차원의 계층구조를 따라 단계적으로 요약된 형태의 데이터 수준에서 보다 구체적인 내용의 상세 데이터로 접근하는 기능을 말한다.

(drill-up)<sup>13)</sup>할 수 있다(Wachowicz, 2000). 그림 2.3은 중심으로부터 성형으로 라인이 뻗어 있다. 하나의 선은 각 차원을 뜻하며 선에 있는 점은 계층별 수준을 표현한다.

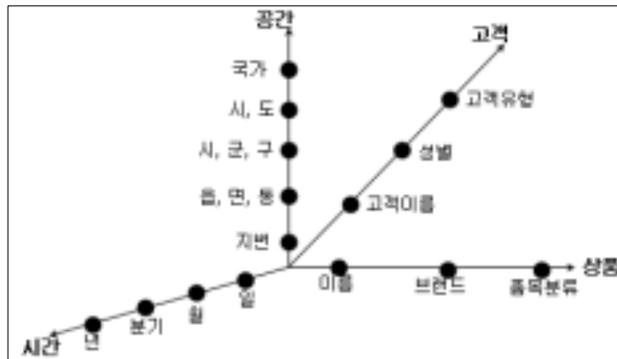


그림 2.3 다차원 성형 모델(star schema model)

그림 2.3은 4개의 차원으로 구성되며, 공간차원은 가장 세분화된 수준인 지번에서부터 국가수준으로 갈수록 일반화 되며, 시간차원은 일, 월, 분기, 년별로 구분하여 계층구조에 따라 설계되었다. 각 수준을 압축하고 있는 점은 해당수준의 관련속성과 연결되어 있다. 그래서 특정 품목에 대한 고객들의 선호 품목, 선호 브랜드별 공간적 군집 패턴을 시계열적으로 분석할 수 있을 뿐 아니라 각 차원별로 세분화, 일반화하여 분석 목적에 따른 차원변경이 가능하다.

공간 데이터 웨어하우스의 다차원 분석 패러다임 측정방법에는 공간적 (spatial) 측정법과 수치적(numerical) 측정법이 있다. 전자의 공간 측정에는 세가지 타입이 있다(Chevallier and Bedard 1990; Chrisman

13) drill-up : 한 차원의 계층구조를 따라 단계적으로 구체적인 내용의 상세 데이터로부터 요약된 형태의 데이터로 접근하는 기능이며 roll-up이라고 한다.

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.

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 벡터타입의 면자료로 생성된 지번자료와 그 자료를 일반화시 거리는 같다.

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

### 2.2.2 공간 자료 집합체

분석가는 공간 데이터 웨어하우스에 온라인 분석 처리기법으로 다차원적 접근을 한다. 이러한 접근은 질의과정을 통해 이루어지는데 시간과 공간에 따른 분포나 패턴을 분석하기 위해서는 자료집합 모델이 필요하다(Mumick, 1997).

공간 데이터 웨어하우스는 위성영상, 항공사진을 포함하는 래스터 자료와 점, 선, 면의 벡터 자료를 포함하는 공간 자료와 비 공간 자료를 포함한다. 예를 들어, 비 공간 자료는 인구 통계 자료나 수치화된 교통체증

자료 등이 있다. 전통적인 데이터 웨어하우스의 온라인 분석 처리기법은 자료집합으로서 글자, 숫자 등의 테이블을 말한다. 반면에 공간 데이터 웨어하우스는 결과로서 지도에 공간 객체가 중첩된 레이어를 사용하는 공간 자료 집합체로서 맵큐브(map cube)<sup>14)</sup>를 사용한다(Stonbraker et al., 1993; ESRI, 1997; Han, 1997; NCGIA, 1999; Bay, 2001).

그림 2.4에서 공간 자료 집합체는 3단계의 분석관점을 나타낸다. 구체적으로 1차원 면은 하나의 차원을 고려하며, 하나의 면은 2개의 차원을 고려하여 분석하는 것을 의미한다. 또한 3차원은 시간, 공간, 고객에 따른 분석을 수행한다. 데이터 웨어하우스의 온라인 분석 처리기법을 이용하여 공간 데이터 웨어하우스의 공간 자료 집합체로 접근, 다차원 분석을 수행한다. 예를 들어, 3차원의 기본 면은 1999년 강남의 여성고객들을 의미한다.

---

14) 맵큐브는 기본지도와 그 공간 객체, 객체에 연관된 속성자료가 계층구조를 이루고 있다. 이 계층구조는 데이터 웨어하우스의 온라인 분석처리기법에서 공간 개념이 추가되어 확장된 자료 집합체라 정의한다.

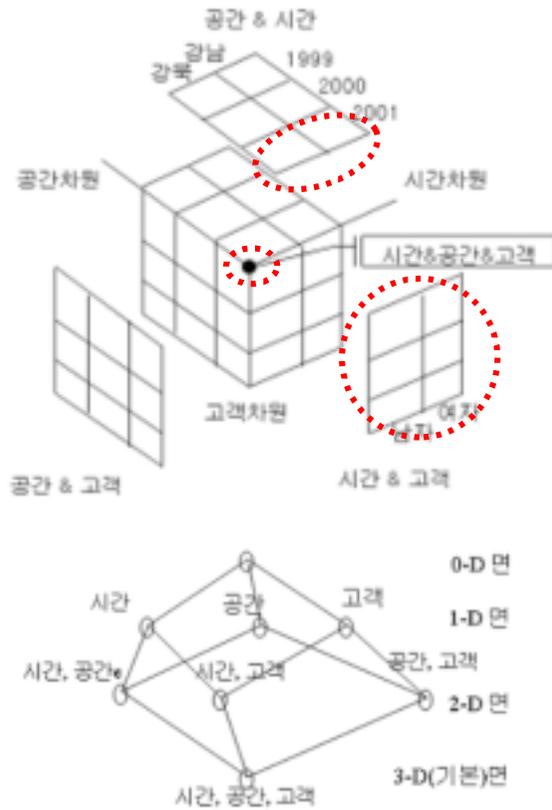


그림 2.4 공간 자료 집합체  
(수정인용: Shekhar et al., 2001)

### 2.2.3 지식발견 논의

지식발견 방법론에는 크게 두 가지 접근방법이 있다. 첫째, 확정적 분석방법(confirmatory analysis)으로 특정 자료의 상관관계에 대해 가정을 세우고 나서 그 가정을 검증하거나 오류를 지적하는 방법으로 기존의 자료 처리 기법에서 사용하고 있다. 둘째, 발견적 접근방법(discovery driven approach)으로 자동화 된 도구를 이용하여 자료 속에서 의미 있는 패턴이나 인과관계, 상호 연관성을 밝혀내는 방법으로 공간 객체를 이

용한 공간 데이터 마이닝(spatial data mining)<sup>15)</sup> 분석방법에 주로 사용된다.

지식발견은 하나의 분석기법을 의미하는 것이 아니라 여러 기법과 방법들의 적절한 조합으로 이루어진 일련의 과정(process)이라 말할 수 있다. 또한 통계적인 관점에서 보면 대용량 자료에 대한 탐색적 데이터 분석(Exploratory Data Analysis)이라고 할 수 있다(Kimball, 1996).

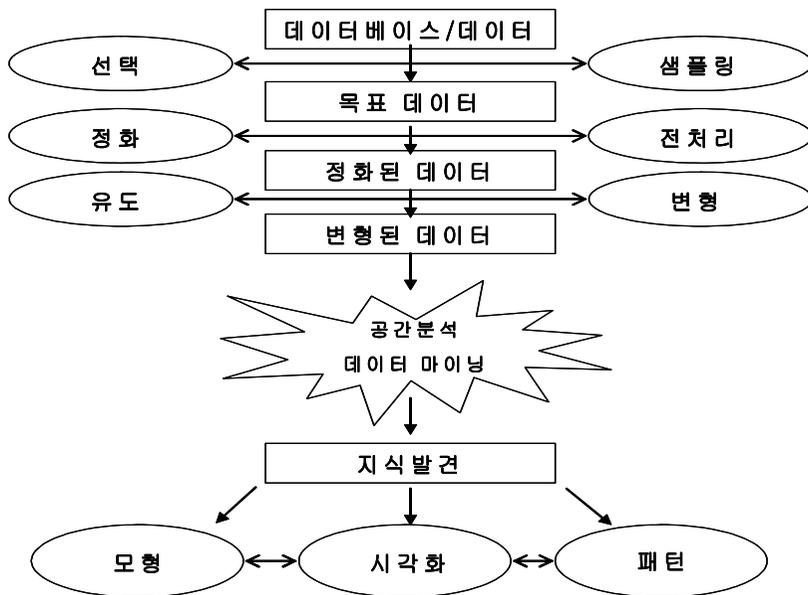


그림 2.5 지식발견 과정의 개관

그림 2.5는 지식발견의 과정으로서 첫째, 데이터 웨어하우스 또는 데이터베이스에서 주제에 맞는 지식을 얻기 위해 관련된 자료를 선택하여 목표 자료를 추출한다. 둘째, 추출된 목표 자료에서 부정확한 값, 결손 값

15) 공간 데이터 마이닝은 흥미있는 정보와 패턴을 발견하는 과정으로 통계학, 데이터베이스 시스템, 지리학과 원격탐사, 공간통계학을 포괄하며 설명적 공간 데이터마이닝에는 일반화, 공간 군집, 공간 연관성 규칙이 있고 예측적 공간 데이터 마이닝에는 공간 회귀 기법이 있다.

(missing value), 불일치(inconsistency), 잡음(noise)등을 제거하고 자료 및 특이 값을 추출하는 전 처리 과정을 수행한다. 셋째, 정제된 자료에서 이미 존재하는 자료에 새로운 자료를 생성하거나 더 많은 정보를 포함 하도록 몇 개의 필드(field)를 하나의 필드로 변환하는 등의 작업을 수행하고 변형된 자료가 분석 알고리즘 수행에 적합하도록 유도한다. 넷째, 패턴을 추출하고 분석을 수행하기 위해 하나 이상의 기법을 적용하여 분석한다. 마지막으로 분석된 결과는 결과를 해석하고 발견된 패턴이 가치가 있는지, 실제로 적용 가능한지를 평가하며 발견된 지식을 표현하기 위해 모형과 시각화 기법을 통하여 지식을 표현, 설명한다.

## 2.3. 인터넷 기반 공간 자료 활용

### 2.3.1 지리정보학의 개념

GIS는 공간상에 존재하는 지리자료(geographic data)와 관련된 속성 자료(attribute data)를 통합하는 정보시스템으로서, 다양한 형태의 지리정보를 효율적으로 수집, 저장, 갱신, 처리, 분석, 출력하기 위해 이용되는 하드웨어, 소프트웨어, 지리자료, 인적자원의 총체적 조직체라 정의한다.

### 2.3.2 지리정보학을 활용한 의사결정

의사결정과정은 문제의 파악(intelligence), 설계(design), 선택(choice)의 단계를 거쳐 상호작용을 통해 최종대안을 선택한다(그림 2.6). 각 의사결정과정의 단계마다 요구되는 정보들은 다양하기 때문에, 공간 의사결정과정에서 필요로 하는 정보를 GIS의 공간 분석기능들이 상황에 따라 요구되며, 탐색적 자료 분석기능이 중요한 역할을 수행한다(임은선, 2001).

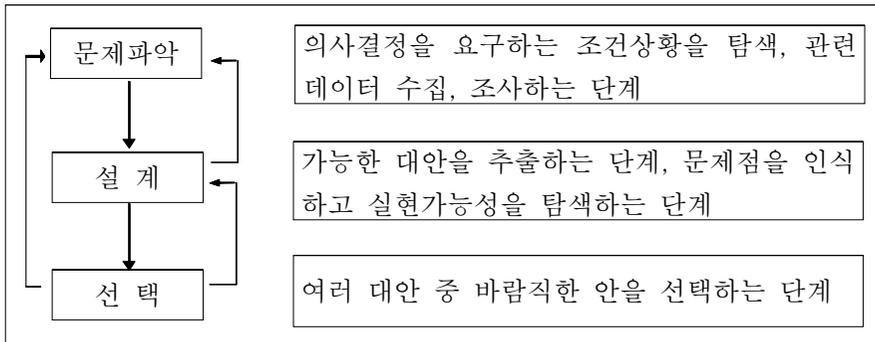


그림 2.6 의사결정 단계(자료: 임은선, 2001)

### 2.3.3 지리학적 지식발견에 관한 방안

지난 수 십 년간 위성영상과 탐사기법의 발전으로 지리학적 자료는 테라바이트급(terabyte)으로 증가되고 있다. 그러나 풍부해진 공간 자료에서 명백하게 관심 있는 정보를 얻는 것은 어려웠다. 그래서 정보 과학자들은 마케팅, 신용카드거래, 의학 분야, 통신, 증권 등 인간 활동 전반에 걸친 디지털 혁명(digital revolution)의 결과에 의사결정을 위한 지식발견을 위해 학제 간 상호 협력하고 있다(Gardner, 1996; Bhandari et al., 1997; Hedberg 1996). 또한 지리정보학자들은 지리학적 지식을 얻기 위해 공간 자료를 변형하고 지능적, 자동, 그리고 통합적으로 가공 처리 후 질의처리, 공간분석, 가시화 등 지리학적 접근하는 해결책을 요구받고 있다(Yuan, 2003; Harvey and Han, 2000; Gahegan, 2003).

지리학적 지식을 얻기 위한 지리학자들의 노력은 1970년대부터 컴퓨터의 도입으로 정량적인 발전을 거듭해왔다. 그 이후 1980년대를 거치면서 표현과 분석법이 강화되면서 컴퓨터지도학의 진보를 이끌었다. 또한 지리학자들은 지리학과 컴퓨터과학의 발전과정에서 풍부한 지리정보의 증대로 다양한 표현법과 방법론을 시도하고 있다. 예를 들어 컴퓨터 구조와 수행능력은 수

치화된 지리정보의 패턴인식, 검색, 분류 등 기존의 처리하기 어려운 정보의 해결책을 제시하고 있다. 최근에는 통계, 데이터 마이닝의 방법에 공간적 특성을 혼합하여 지리학적 지식을 얻고 있으며 이러한 지식은 공간군집, 세분화, 일반화과정을 지도학적 표현법으로 가시화 한다(Gahegan, 2003).

지리정보의 형태와 컴퓨터의 표현방법은 위상과 기하학을 함축하고, 이 틀 (framework)은 지리학적 속성의 측정을 요구하며 결과적으로 패턴을 추출할 수 있다. 그래서 지리학적 정보는 고차원일 뿐 아니라, 다른 차원의 측정을 제공하며 거리에 따라 상호 연관되어 있다.<sup>16)</sup> 즉, 공간적 자기 상관성이 적용된다. 더불어 실제 세계에서 공간요소를 측정시 시간과 공간에 따라 분석방법은 달라진다. 공간 자료는 특정한 공간에 위치하는 객체와 관련된 자료이다. 공간 데이터베이스는 공간 자료 타입과 공간 사상들 사이에 위상관계에 의해 표현되는 공간 자료와 비 공간 자료로 동시에 저장되고 거리정보를 포함하며 대부분의 공간 데이터베이스는 일반적으로 공간 인덱싱 구조에 의해 구성되는 공간 접근 방법(SAM: Spatial Access Method)에 의해 접근된다. 예를 들어 전염병의 확산이나 도시지역에서 교통시스템의 추세나 패턴 등이 있다.

분석을 위한 지도학적 기법은 적당한 거리측정과 각 주제에 맞는 틀에 따라 측정법이 달라지며, 지도학적 변형은 적절한 시간에 흥미 있는 지리학적 패턴을 발견하고, 효과적인 결정을 하는데 이용되며, 결과적으로 지리학적 지식에서 다양한 버전의 공간 측정법을 이용하여 구체화된다. 측정된 공간 속성들은 항상 공간적 의존성과 이질성을 제시한다. 공간적 의존성은 다른 공간적 속성과 연관되어 있는 것을 의미하며 주로 유클리드 거리내의 인접한 공간에 영향을 받는다. 공간적 이질성은 서로 다른 입지의 특성으로 공간적

---

16) "Everything is related to everything else, but things are more related than distant things", (Tobler's First Law of Geography)

또는 지역적 차이점이 발생한다. 이는 불안정한 형태로 표출되어 나타난다. 공간 의존성의 측정기법은 공간통계의 Moran's I 와 Geary's C방법이 있고 국지적인 공간적 이질성의 측정기법으로는 LISA(Local Indicators of spatial Association)가 있다.

크기, 사상, 가장자리 효과 등 시공간적 객체의 관련성은 비 공간 객체보다 복잡하고 더욱 관련성이 크다. 거리에 따른 관계, 방향, 연결성은 객체의 차원에 따라 다양하며 다양한 지리학적 객체에서 주요한 지식발견을 위해 시간, 공간적으로 추출해야 한다. 이러한 방법에는 일반화(generalization), 연관 규칙(association rule), 세분화(segmentation)등이 있다.

공간 데이터 웨어하우징은 공간 데이터 웨어하우스로부터 암시적인 지식, 공간관계, 또는 명시적으로 저장되지 않은 규칙들을 추출하여 분석가에게 원하는 지식을 제공하기 위한 저장 공간으로 단순히 자료의 저장 및 관리를 지원하기 위한 기존의 GIS(Geographic Information System) 데이터베이스와 차별된다. 또한 비 공간 자료를 위한 데이터베이스로부터 지식발견(KDD : Knowledge Discovery in Database )과 데이터 마이닝(data mining)은 확률적 모델링(probabilistic modeling), 규칙 발견(rule discovery), 예측 모델링(predictive modeling), 기술적 통계(descriptive statistics)를 통해 공간 분석을 수행한다(Bennett, 1999).

### 2.3.4 인터넷 기반 공간 자료 저장소

지리정보과학은 웹 환경에서 상호 연동하여 이루어짐에 따라 크게 진보되었다. 그 내용을 정리하면 첫째, 공간상의 제약을 축소했다. 다수의 분석가들이 한 곳에 모여 의사결정을 해야 하는 공간상의 제약을 줄일 수 있게 되었다. 둘째, 공간분석, 시설물 관리처럼 다양한 수준의 GIS 프로그램에 의한 해결책을 웹 환경에서 분석가에게 제공하고 상담할 수 있게 되었다. 셋째, 사회변화 관점에서 지리정보과학을 포함하는 지식분야산업 자체가 웹 환경을 통한 정보교환으로 옮겨지고 있다.

웹 환경을 통한 GIS상담 시스템에는 지리적 사상과 속성정보를 제공하며 주제에 따라 다양한 색채 표현법과 분류법이 요구된다. 또한 분석가는 의사결정을 위해 각기 다른 수요에 맞도록 지리정보를 수정, 편집할 수 있는 기능을 필요로 한다. 특히 어떤 주제에 관해 상세한 정보를 필요로 하는 분석가는 자료의 정확도나 그 주제를 반영하는 축척과 함께 반복되는 분석과정을 위해 초기자료는 손상이 없고, 잘못된 초기자료는 수정해야 한다. 또한 시간차원으로 인한 지리적 속성변화도 반영해야 한다. 이러한 공간 자료 공급의 유통과 변화는 자료가 생성된 필요과정과 서비스를 포함한 부가적인 정보를 포함해야하며 이와 함께 수시로 변하는 공간 자료와 관련된 속성정보가 변화하는 시점에 대한 정보도 공급되어야 한다.

NSDI<sup>17)</sup>의 Geospatial Clearinghouse는 방대한 양의 공간 자료와 비공간 자료를 다양한 축척과 형식으로 저장하고 웹 환경에서 시간적인 공간 자료의 변형과 분석을 수행할 수 있도록 하였다(Sengupt, 1999). 또

---

17) NSDI(National Spatial Data Infrastructure : 국가 공간자료 하부구조) : 미국에서 처음 사용하였으며 대부분의 국가에서 이 용어를 사용한다.

한, Alexandria Digital Library project에 따르면 실세계의 다양한 공간 자료 범위와 정확성(accuracy), 개별적인 자료변형을 제한점으로 지적하였다. 이러한 제한점은 공간 자료의 중복성과 호환성의 문제로 연결되고 구현된 저장소의 자료가 공유되지 못하는 사례가 발생하였다(황철수, 2001). 이러한 문제점을 해결하기 위한 수치화된 공간 자료에 대해 체계적으로 설명하는 문서가 메타데이터(metadata) 이다.

미국 FGDC의 ‘디지털 지리공간 데이터에 대한 메타데이터 표준’(FGDC Content Standard for Digital Geospatial Metadata)은 서비스에 대한 목적, 소개, 서비스 개발자의 소속집단, 업데이트 시기, 자료 처리 환경, 지도 투영법, 서비스하는 영역의 좌표, 자료 제공자, 자료의 형식, Compile 시기, 기준점, 픽셀정보, 속성 값 등을 담고 있는 기준을 제시한다. 이러한 기준은 공간자료의 적합성, 접근성, 변환과 처리를 위해 활용되고 있다(Nordic Metadata Project, 2000<sup>18)</sup>; Electronic Records Management Guidelines, 2002<sup>19)</sup>).

## 2.4 웹 기반 공간 데이터 웨어하우징의 활용 사례 및 연구

본 연구는 인터넷 환경에서 공간정보의 공유와 활용을 위한 저장 공간에 초점을 맞춘 연구로써 기존의 연구를 검토하여 그 시사점을 찾아내고 실험적으로 구현하려는 공간 데이터 웨어하우징 시스템에 참조 하고자한다. 이 시스템에 관련된 연구로는 의사결정을 위한 공간 데이터 웨어하우징에 관한 연구, 웹 기반 지리정보과학에 관한 연구를 검토하였다.

공간 데이터 웨어하우징의 관리적 측면을 다룬 시스템에는 미국 Nat

---

18) <http://www.lib.helsinki.fi/meta/index.html>

19) <http://www.mnhs.org/preserve/records/metadastandard.html>

ional Geospatial Data Clearinghouse가 웹 환경에서 메타데이터를 이용하여 공간 데이터 웨어하우스의 공간 자료에 접근하도록 하였다. 이 시스템은 Alaska GDC, EDX, ESRI, NOAA CSC, FGDC, NRCS의 분산 서버 환경에서 공간 자료를 제공한다<sup>20)</sup>. 병원에서는 질병에 대한 환자들의 신상 자료를 이용하여 특정 지역에 대한 질병 분포 패턴을 연구했다(ESRI, 1998). Government of Alberta는 토지관리를 위해 “AltaLIS”를 운영하고 있다. “AltaLIS”는 공간 자료를 공유하는 방안으로 웹 환경에서 공간 질의를 통해 공간 데이터 웨어하우스에 접근한다. 또한 자연재해 관리 분야에서는 산불관리를 위한 수치지도와 영상을 제공하는 Los Alamos National Lab을 들 수 있다. Los Alamos National Lab은 공간 자료의 공유를 위해 메타데이터를 제공한다.<sup>21)</sup> 우리나라에서 추진되고 있는 서울시 공간 데이터 웨어하우스는 “서울시 엔터프라이즈 GIS 구축 전략”의 일환으로 공간 자료의 공유와 활용 시 행정 업무<sup>22)</sup>에 초점을 맞추어 추진 중에 있다(박수홍, 1999). USGS와 Microsoft는 지역에 대한 영상을 서비스하기 위해 “TerraServer-USA”<sup>23)</sup>를 개설하고 저장 공간으로 공간 데이터 웨어하우스를 구현하였다. 지역 스키마 설계는 “place”를 사실 테이블로 공간 자료의 타입(feature type), 지역의 좌표(place grid), 주(state)를 차원 스키마로 하여 위성영상과 항공사진 영상을 제공한다. 공간 자료를 사용한 분석적 측면에 초점을 맞춘 사례에는 “Convergent Group”<sup>24)</sup>으로 고객 정보 관리(customer information system)를 위해

---

20) <http://www.fgdc.gov/clearinghouse/clearinghouse.html>

21) [http://www.gislab.lanl.gov/data\\_warehouse.html](http://www.gislab.lanl.gov/data_warehouse.html)

22) 공간 데이터 웨어하우스에 공통된 프레임워크 데이터, GIS DB, 기본도, 비 공간 속성 DB등의 데이터 마트를 구성하고 이기종의 운영 시스템에 세무 종합 관리시스템, 환경관리 시스템, 교통 관리 시스템, 주택 건축관리 시스템의 4개 통합 시스템으로 구분하고 그 하위에 세부적인 단위 업무 시스템으로 분류하여 설계하였다(서울시정개발연구원, 1999)

23) <http://terraserver-usa.com/>

24) <http://www.convergentgroup.com/CIS.htm>

공간 데이터 웨어하우스를 구현하고 고객을 관리하고 투자지역을 분석하기 위해 사용하고 있다. "CubeWerx"는 상담을 주 목적으로 주제별 공간 데이터 웨어하우스를 구현하고 웹 환경에서 서비스 한다<sup>25)</sup>.

## 2.5 기존 연구의 한계와 대안

이 장에서는 정리한 이론적 배경과 개념을 토대로 실제 활용되고 있는 사례를 검토하여 기술적 연구 동향을 살펴보았다. 우선 지리정보과학 영역에서는 공간 자료의 공유와 관리 분야에 초점을 맞추어 공간 데이터 웨어하우스를 설계 및 구현하였다. 이는 관리적 측면만을 고려한 것으로서 공간 자료를 활용한 분석적 측면을 고려하지 않았다. 그리고 대부분의 공간 데이터 웨어하우스를 활용한 연구는 국가나 지방자치단체의 행정업무를 위한 서비스가 대부분이었다.

### 2.5.1 기 개발된 인터넷 기반 GIS 지리정보서비스

웹 환경의 지리정보 서비스를 제공하는 사이트를 검토한 결과 지도서비스는 단순히 위치정보만을 제공하며 고정된 분류법과 획일화된 확대, 축소기능, 고정된 공간 자료 타입의 일방적인 서비스만이 제공되고 있었다.

이러한 기능으로는 공간 자료와 관련된 속성자료의 위치적 특성만을 파악할 수 있다. 이것은 공간적 의사결정을 위한 분석가의 원하는 공간 분석을 수행하기가 어렵다고 판단된다. 또한 시간에 따른 지역의 변화와 특정 주제에 관련된 공간 사상이나 이와 관련된 속성의 변화를 고려하기 어렵다. 그러나 실세계에서는 특정 주제에 따른 시간, 공간적 변화가 일어나고 있다. 그래서 변화된 부분, 혹은 관리자의 잘못된 부분을 발견하더

---

25) <http://www.cubewerx.com/main/services.html>

라도 편집, 수정하기 어렵다. 따라서 본 논문에서는 기존의 고정된 지도학적 사상만을 제공하는 기존방식을 배제하고 확장성이 뛰어난 언어를 채택하였다. 그리고 공간 자료의 유통과 중복된 자료구축을 방지하기 위해 이론적 배경을 바탕으로 공간 메타데이터 문서를 제작하였다.

## 2.5.2 웹 환경 공간 데이터 웨어하우스 스키마 설계

기존의 공간 데이터베이스에서 주로 사용되는 온라인 거래처리(OLTP) 환경에서는 데이터베이스를 설계하는데 객체-관계(Entity-Relation) 다이어그램이나 정규화기법이 사용되었다. 그러나 데이터 웨어하우스의 질의는 분석적이고, 다차원적 범위의 질의가 대부분이다. 이러한 질의를 처리하기 위해 성형(星型) 스키마모델(star schema model)이 주로 사용되고 있었다.

성형스키마모델은 하나의 사실테이블<sup>26)</sup>과 여러 개의 차원테이블<sup>27)</sup>로 구성되는데, 사실테이블은 차원테이블의 주키(primary key)들로 구성된 결합키(composite key)이다. 이러한 속성들의 결합들을 다차원 키로 사용하여 질의에 대한 인덱싱(indexing)을 하고 각각의 차원테이블에서 추출된 키들을 cartesian product<sup>28)</sup>로 순서쌍을 구성하여 적용, 조회하여 조인(join)한다. 성형스키마는 분석가의 질의에 빠른 응답을 하는 장점을 가지고 있지만, 조인의 횟수가 많아지면 확장성과 유연성이 떨어지는 단점을 가지고 있다. 또한 새로운 자료를 추가하거나 종속관계를 변경할 때

26) 사실테이블(fact table): 분석가가 필요한 자료에서 실제 발생하는 자료요소로 분석을 요하는 변수차원의 항목들의 집합으로 모든 사실테이블은 기본 키를 갖는다.

27) 차원테이블(dimension table): 각각 질의의 관점에 대한 자료를 저장해 놓은 테이블로 측정하고자 하는 항목들로 구성되어있다.

28) cartesian product: 각각의 테이블로부터 유도된 행의 모든 결합을 말하는데, 둘 이상의 테이블에서 하나의 테이블과 나머지 테이블의 모든 행을 쌍으로 만든다. 예를 들어 조인(join)은 조인 조건에 위배되는 행을 제외시키는 cartesian product의 부분집합이다.

스키마 변경이 어렵고 논리적 설계개념의 부족으로 인한 정규화가 어려워 실제적인 물리적 스키마설계가 복잡해진다. 이러한 단점은 공간자료와 공간객체, 비 공간자료의 연계에 대한 복잡한 연산의 보다 많은 인덱싱으로 인해 저장 공간이 과대해지고 목표 마케팅 지역설정을 위한 다차원적 분석에 부적합하다.

본 연구의 공간 데이터 웨어하우스 스키마모델은 성형스키마의 사실테이블 구조와 동일하게 유지하면서 차원테이블을 정규화한 눈송이 스키마 모델<sup>29)</sup>을 사용하여 설계하였다. 이 눈송이 스키마모델은 공간 데이터 웨어하우스의 사실테이블인 카드거래와 고객신상자료의 각 차원별 데이터마트와 조인되며, 이후 고객신상자료의 주키(primary key)는 공간 데이터베이스의 공간자료와 객체로 연결하고 관계성을 맺는다. 이러한 관계성은 실제 사례에서 다차원적 공간 분석을 통한 지식탐색과 발견과정에서 높은 유연성과 확장성의 장점을 갖는다.

### 2.5.3 공간의사결정을 위한 다차원 모델링 대안

본 연구는 백화점 카드 거래라는 실제 사례를 통해 데이터 마트를 설계하고 웹 환경에서 공간 메타데이터를 통한 공간 자료 집합체와 연동하여 공간 데이터 웨어하우스를 구현하였다. 그리고 구현된 데이터 웨어하우스를 바탕으로 분석적 측면을 고려한 공간 데이터 웨어하우스의 효율성을 검토하여 공간의사결정과정을 지원하는 것이 주 목적이다.

---

29) 눈송이 스키마모델(snowflake schema model): 사실테이블과 직접 조인되는 차원테이블이 있으며, 이 차원테이블은 또 다른 차원테이블 상의 기본 키로 참조하는 외래 키를 가진다. 이렇게 차원테이블이 정규화 됨으로써 차원테이블의 크기가 줄어들고 각 속성 값이 하나의 저장 공간에 유지됨에 따라 자료 무결성을 높은 수준에서 유지한다. 또한 자료 저장 공간을 최소화함으로써 시스템의 유연성을 증가시킨다. 그러나 많은 테이블 조인으로 인해 분석가 질의의 응답속도에 성형 스타스키마 모델에 비해 느리다.

본 연구의 공간의사결정을 위한 실험적 시스템에서 공간적 위치와 위상 관계를 갖는 공간자료는 OLTP 방식의 운영시스템인 공간데이터베이스에 저장된다. OLAP 시스템에 필요한 자료는 공간객체, 시간, 공간큐브 등 OLTP 방식의 운영시스템으로부터 추출되어, 정제 및 변화과정을 거쳐 데이터마트에 저장된다.

공간적 의사결정과정을 지원하기 위해 공간 데이터 웨어하우스 내부에 데이터마트를 구성함으로써 관리자는 공간 데이터베이스의 원 자료를 보호하고 분석가에게 빠른 응답성능을 제공하며, 공간 자료에 거래자료, 고객 신상자료가 통합된 양질의 자료를 제공한다. 또한 각 차원별 공간 자료집합체에 자료를 요약, 일반화된 형태로 논리적 계층수준을 두어 분할함으로써 분석가에게 공간, 시간별 세분화된 거래량과 상품품목을 분석하고 시-공간 차별적인 목표 마케팅 계획이 가능하다.

목표 마케팅 지역설정의 방법론에는 티센 다각형<sup>30)</sup>을 활용하여 공간이웃을 정의하는 방법과 군집기법을 활용했다. 공간이웃과 군집분석은 공간에서 거리를 사용한다는 점에서 유사하나, 공간이웃은 고객의 위치간 거리를 이용하여 고객의 밀도를 확인하는 정리나 예측에 사용하였다. 즉, 공간이웃은 고객의 이웃을 정의할 때 단지 거리개념을 적용하여 고객의 밀집을 예측하려는 감독학습(supervised learning)이다. 반면에 군집기법은 단순거리가 아닌 차원에 따라 변수를 사용하는 무감독 학습(unsupervised learning)으로 특별한 외부 기준 모델을 사용하지 않고 등질 그룹을 설정한다. 이 연구에서는 구매량과 고객 수에 따라 군집분석을 수행하였

---

30) 티센 다각형(thiessen polygon)은 각각의 점집합들이 주변에 영향을 미치는 개개의 영역을 결정하는 방법으로, 점 자료를 연결한 이후 그 연결선의 중간지점을 연결한다. 각 점 자료들에 대하여 중간지점을 연결하여 생성된 면자료를 티센 다각형이라 한다. 불규칙한 다각형은 각각의 속성을 갖고 그 면적을 점자료의 공간 이웃(spatial neighbour)이라한다. 본 연구에서는 고객위치로부터 같은 거리에 놓이는 방법으로 구성한다.

다.

본 연구는 “목표 마케팅 지역 선정”이라는 주제를 가지고 웹 환경에서 공간 자료의 관리적 측면과 분석적 측면을 동시에 고려하는 실험적 시스템을 구축하여 고객의 카드 거래 패턴에 따른 특성을 파악하였다. 그리고 발견된 특성을 공간에 적용하여 공간 군집성을 파악, 군집 패턴과 고객의 구매특성을 조사했다. 마지막으로, 분석을 통한 지식을 바탕으로 목표 마케팅 지역을 선정하였다. 목표지역의 고객들의 구매패턴을 조사하기 위해 고객을 구매량에 따라 세분화하고 분류된 고객의 선호 브랜드와 상품을 연관규칙<sup>31)</sup>을 사용하여 공간적으로 분류된 카드고객별 선호도와 신뢰도를 발견하였다. 이러한 지리정보과학(GIS: geographic information science)과 경영정보시스템(MIS: management information system)<sup>32)</sup> 요소의 적절한 융합은 민간기업을 위한 새로운 서비스 모델로서 차별적인 시공간적 마케팅전략과 거래패턴별 지역적 고객관리 계획수립이 가능하다.

---

31) 자료로부터 특정한 연관규칙을 발견하는 것은 데이터마이닝의 지식발견 방법론 중 하나이다. 자료 내에 내재하는 유형들을 체계적으로 추출 후 유형별로 정확도와 신뢰도를 이용하여 발견한다.

32) 경영정보시스템은 관리자들에게 정보를 제공하며, 조직 내의 운용과 경영 및 관리자의 의사결정기능을 지원하는 종합적인 사용자-기계시스템(man-machine system)으로 정의된다. 경영정보시스템은 하드웨어, 소프트웨어, 수작업 절차, 분석 및 계획모형, 통제와 의사결정 및 데이터베이스, 모델, 정보통신 등을 활용함으로써 그 기능을 수행한다.

## 3장 실험적 시스템의 설계와 구현

### 3.1 실험적 시스템 구현 요소

본 연구의 웹 기반 실험적 시스템은 공간 데이터 웨어하우스, 공간적 의사결정을 위한 지식 탐색 및 발견, 인터넷 지도서비스로 구성되며 각 시스템의 구성요소는 상호 결합되어 있다.

#### 3.1.1 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)<sup>33)</sup>에 미들웨어로써 ArcSDE 8.2<sup>34)</sup>를 사용하여 공간 자료와 관련된 속성자료를 저장하였다. 웹 환경에서의 개발도구는 인터넷 서버인 IIS<sup>35)</sup>, 외부 응용프로그램과 웹 서버와의 상호 연계는 Servlet<sup>36)</sup> 엔진인 Jakarta-Tomcat-3.2.3<sup>37)</sup>, 공간 자료의 로딩(loading)을 위해 ArcIMS 4.0<sup>38)</sup>, 환경은 Java 2 Standard Development Toolkit 1.3.6,

---

33) Oracle은 데이터베이스 관리 시스템 소프트웨어로서 Oracle 그 자체가 데이터베이스는 아니다. 즉, 데이터베이스를 효율적으로 관리할 수 있도록 도와주는 시스템 소프트웨어이다.

34) ArcSDE는 데이터베이스 관리 시스템에서 공간 자료 관리를 용이하게 하는 GIS 미들웨어다.

35) IIS(Internet Information Server) : Microsoft에서 제공하는 웹 서버를 위한 플랫폼이다.

36) servlet은 인터넷 환경의 C/S 구조에서 client의 요청이 있을 때 전통적인 server에서 CGI방식의 프로세서를 생성하는 것이 아닌 작업부하가 적게 걸리고 속도에서 월등한 Java 기반의 thread를 생성한다.

37) Sun사와 Apache 연구 그룹이 공동으로 개발한 웹 서버의 일종으로 주로 servlet엔진으로 사용한다.

38) 인터넷 환경의 지리정보의 배포를 위한 플랫폼이다.

공간 자료 명세서인 메타데이터 문서는 ArcCatalog를 사용하여 XML문서<sup>39)</sup>로 작성하였다. Microsoft SQL-2000의 확장 모듈인 Analysis Services<sup>40)</sup>는 구축된 공간 데이터 웨어하우스에 다차원적 접근을 위한 응용 시스템으로 온라인 분석처리에 사용하였다.

### 3.1.2 백화점 공간 데이터 웨어하우스 구성 방안

공간 데이터 웨어하우스의 벡터타입 점 자료는 고객의 위치를 표현하고, 도로와 철도는 선 자료, 행정동과 구별 행정구역은 면 자료로 구성하였다. 또한 래스터타입의 영상자료는 전 처리과정을 거쳐 공간데이터베이스에 저장하는데, 공간데이터베이스는 Oracle에 미들웨어인 SDE를 통해 저장되어 지도큐브를 형성한다. 그리고 비 공간 자료인 고객신상자료와 거래 자료는 각각의 데이터베이스에서 추출 및 변화과정을 거쳐 정제되어 데이터마트로 구성된다. 각각의 데이터마트와 지도큐브는 공간 데이터 웨어하우스를 구성하고, 결합구조로 연결하였다. 분석가는 고객신상, 거래 데이터마트와 지도큐브에 구조화된 질의 언어를 통해 개별접근이 가능하고, 공간 데이터 웨어하우스에 온라인 분석기법(OLAP)을 통해 다차원적 접근을 시도하였다.

### 3.1.3 지식탐색 및 발견을 위한 방안

실험적 시스템에서 목표 마케팅 지역설정을 위한 지식은 카드고객이 어디에 밀집해 있으며 어느 정도를 구매하였는가? 또한 그들이 구매행위가

---

39) W3C에서 제안한 웹 페이지 소스체계로 기존의 HTML보다 확장성, 구조, 안정성이 뛰어나고 계층적 구조를 포함하며 플랫폼 독립적이다. XML parser는 사용자의 request를 해석하고 수행하며 GIS에서는 지리학적 사상의 공간 객체별 customization을 수행한다.

40) Data Mining과 OLAP을 수행하기 위한 소프트웨어로서 본 논문에서는 온라인 분석처리(OLAP)에 사용한다.

주로 발생하는 시간과 선호하는 브랜드 및 품목은 무엇인가? 라는 질문에 대한 해답일 것이다. 이 질의에 대한 해결책으로 공간과 시간, 선호하는 품목들에 대한 고객을 세분화하고 지식을 탐색하여 목표 마케팅지역을 설정하기 위한 지역과 특정시간을 발견한다. 발견된 지식은 공간적 특징을 갖는 정보로서 특정지역에 마케팅 노력을 집중할 수 있다.

본 연구에서는 공간 데이터 웨어하우스를 구현하고 OLAP 기법으로 거래 자료를 조회 및 다차원 접근을 시도하여 분석가의 필요성에 따라 관련된 공간자료와 고객 신상자료, 구매 품목 및 브랜드별 다차원 모델을 구성하고 지식을 탐색한다. 여기서 고객의 위치와 특정 차원에 따른 공간정보를 분석하여 수치지도로 목표지역을 시각화한다. 그리고 의사결정자와 분석가사이의 통신 수단을 만들어 상담할 수 있는 기능을 삽입했다. 지도 서비스는 특정 차원에 따른 지식을 발견하고 분석가에게 의뢰하고 구매사건이 발생하는 각기 다른 차원들과 대비한 정보를 시각화한다. 여기에는 GIS의 공간분석, 데이터 마이닝의 군집분석이 활용되었다.

기존의 지식발견을 위한 공간통계기법은 지식 탐색을 위한 결과도출의 과정이나 결과 자체의 관점에서 데이터 마이닝과 비슷하다. 그러나 목표 마케팅지역을 설정하는데 공간통계기법이 거래행위의 각 시간 및 공간 등 각 차원별 추측을 하고 검토하는 것이라면, 데이터 마이닝은 추측과 검증 과정을 공간 데이터 웨어하우스에 접근하여 다차원 모델을 생성하고, 자동으로 공간자료와 비 공간자료들을 상호 연계하여 지식을 탐색 발견하는 차이를 갖는다. 즉, 이 연구에서는 기존의 확정적 분석방법의 공간통계기법을 배제하고, 자동화된 마이닝 도구를 사용하여 공간 데이터 웨어하우스의 정제된 자료에서 카드구매에 따른 시공간적, 품목별 의미 있는 패턴이나, 상호연관성을 발견한다. 그리고 발견된 지식은 고객을 세분화하여 차별적인 고객관리 및 마케팅 전략수립을 가능하게 한다.

### 3.1.4 공간적 의사결정을 위한 인터넷 기반 상담 시스템

웹 기반 상담 시스템은 의뢰자와 상담자간의 의사교환 및 일련의 과정을 웹 환경에서 도와주거나 대행하는 일이다. 즉, 직접대면을 통한 상담이 아닌 웹 기술을 기반으로 웹 환경에서 상담이 이루어진다. 본 연구에서 다루게 되는 실험적 시스템은 웹 환경에서 실시간으로 지리 자료를 활용하여 지식을 발견하고 의사결정과정을 지원하는 시스템을 구현하였다.

최근 민간기업에서 지리정보에 대한 관심이 높아지고 있는데, 그 이유는 고객에게 차별적인 서비스를 제공하기 위한 정보의 상당수가 지리정보이기 때문이다. 예를 들어 마케팅 전략가는 소매지점이 포괄하는 상권과 구매 고객들의 밀집지역을 공간적으로 세분화하기 원한다. 이렇게 고객에 대한 자료 수집 및 구매패턴 분석은 상세한 시장분석을 가능하게 한다.

이러한 맥락에서 과거의 산업사회는 재화의 소비생활을 뒷받침하는 소비형 산업이 중요했다. 그러나 미래의 지식 정보화 사회는 새로운 기술과 아이디어에 의한 신제품의 개발과 이를 뒷받침하는 전문적인 지식, 정보서비스 활동이 주된 부가가치의 창출원이 될 것으로 전망하고 있다(주성재, 2001). 더불어 지식 정보 사회는 발전된 인터넷 환경에서 전문적 지식과 정보를 발견하여 의사결정과정에 이용한다. 이러한 사회 현상의 원인으로 기업들은 급변하는 환경변화에 혼돈스럽지만 고객들은 합리적이고, 현명해 지고 있다(Oracle, 2003).

즉, 고객 개개인의 개성으로 자신에 맞는 특화된 상품과 서비스를 원하는데, 이러한 수요를 원하는 고객들은 지역적 집적의 형태로 공간상에 나타난다. 왜냐하면, 그들을 고용한 상호 연관된 기업들은 공간적으로 집적하고 구성 기업간 연계와 네트워크를 통해 상호의존하기 때문이다<sup>41)</sup>.

---

41) 기술과 시장의 수요가 빠르게 변화하는 경제 환경은 기업간 상호의존성 및 상호작용의 강화로

집적된 지역의 고객들을 확보하고 유지하는 일은 마케팅 전략에서 지리정보의 공유, 체계적인 관리, 다양한 활용이 중요시된다. 본 연구에서 구현한 인터넷을 활용한 실험적 시스템은 분석가와 의사결정자의 능동적인 대화가 가능하며, 분석에 필요한 많은 자료를 체계화하고, 자료관리 기능과 공간 분석을 포함한 여러 다차원적 분석을 수행한다.

## 3.2 시스템 설계

이 장에서는 앞서 기술한 공간 데이터 웨어하우스의 개념과 활용, 사례 연구에서 정리된 내용을 이용하여 “목표 마케팅 지역 선정”을 위한 분석가의 요구사항을 파악하고 공간 데이터 웨어하우스와 그 응용 시스템을 설계한다.

### 3.2.1 요구사항 분석

본 연구는 공간 데이터 웨어하우징을 활용한 사례로 시스템 설계 및 구축 계획을 수립하는데 있어 다음과 같은 요구사항을 설정하고 연구를 진행하였다.

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

둘째, 자료의 범위는 A백화점의 카드 고객 거래 170만 건 중 서울시 4개 지점에서 카드로 물품을 구매하고 2000년 5월부터 2001년 4월 기간에 거래한 서울시 거주 고객 41,263명의 신상 자료를 대상으로 했다.

셋째, 본 연구에서 고려한 실험적 시스템의 주 대상은 의사결정을 하는

---

이어지는데, 학자들은 이러한 상호작용을 공간적으로 산업클러스터의 형성이라 언급한다. 이 상호작용의 차원들 중 연계와 네트워크를 그 핵심으로 본다(정병순, 신창호, 2002).

분석가로 한정하고 시스템의 보안과 인터페이스의 설계는 이 연구에서 배제하였다.

넷째, 본 연구는 웹 환경을 통한 분석가의 공간적 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.

다섯째, 본 연구에서 사용하는 수치화된 공간 자료와 비 공간 자료는 연구지역인 서울시의 실세계를 추상화(Abstraction), 단순화(Simplification)한 것으로서, 자료의 내용과 형식, 용어와 개념을 정리한 메타데이터 문서를 제공한다. 이 문서는 미국 FGDC의 “디지털 지리 공간 자료에 대한 메타데이터 표준”에 따라 작성했다.

### 3.2.2 시스템 설계

공간 데이터 웨어하우스를 포함해서 응용 시스템의 설계에는 사용자 요구사항에서 도출된 개념을 통합하는 논리적 설계와 어떻게 논리적 설계를 특정 데이터베이스 시스템에 연계하는지 결정하는 물리적 설계로 구분한다.

#### 3.2.2.1 논리적 설계

실험적 시스템은 데이터 웨어하우스, 지리정보학, 정보 시각화의 개념적 영역을 가진다(그림 3.1). 각 영역은 결합관계를 이루고 있는데 데이터 웨어하우스와 지리정보과학의 개념적 결합으로 공간 데이터 웨어하우스, 지리정보과학과 정보 시각화 영역에서 지도학의 공간 자료 시각화, 데이

터 웨어하우스와 정보 시각화 영역으로 데이터 웨어하우스의 스키마 모델을 시각화한다. 마지막으로 이 세 가지 영역의 결합은 공간 자료 집합체로 개념적 모형을 설명한다.

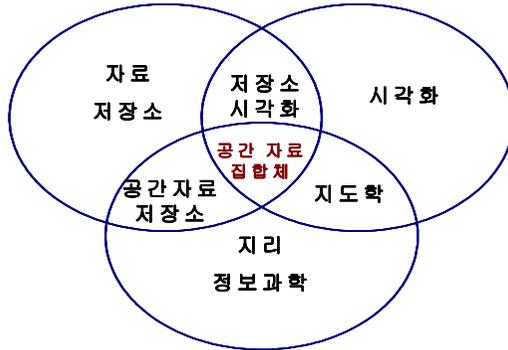


그림 3.1 실험적 시스템의 개념적 영역

앞선 개념적 영역에서 추출된 실험적 시스템의 논리적 체계는 그림 3.2에서 설명하고 있다. 실험적 시스템의 논리적 설계는 지도학, 지리정보과학, 공간 데이터 웨어하우스로 3가지로 구분할 수 있다.

첫째, 공간 데이터 웨어하우스의 영역에서 데이터 웨어하우스는 서울시 통계 자료가 담긴 이기종(heterogenous)의 데이터베이스, 지도큐브의 서울시 수치지도와 영상, 카드 거래 자료를 갖는 데이터 마트와 1:N<sup>42)</sup>의 형식으로 연결된다. 그리고 데이터 마트는 고객의 신상 자료가 있는 고객큐브와, 거래 시간 자료가 있는 시간큐브와 1:1로 연결된다. 그리고 지리적 사상의 시간적인 변동을 담기 위해 지도큐브와 연결된다.

둘째, 지리정보과학 영역에서 벡터 레이어는 공간 객체인 점, 선, 면과 1:N으로 연결되고, 각각의 객체가 담고 있는 사상은 고객위치, 주요도로와 철도, 행정도와 수계 등을 담고 있다. 이때 연구지역의 영상인 래스터 레이어는 벡터 레이어와 중첩하기 위해 지도큐브와 1:N으로 연결한다. 최

42) 본 논문에서 'N', 'M'은 불특정 다수를 표현한다.

종적으로 지리정보과학의 레이어는 데이터 웨어하우스와 N:M형식으로 연결된다.

셋째, 지리정보과학에서 수행하는 공간분석, 레이어 중첩, 가시화는 지도의 구성요소인 주기, 심볼과 적절한 지도학의 시각화 기법을 사용한다.

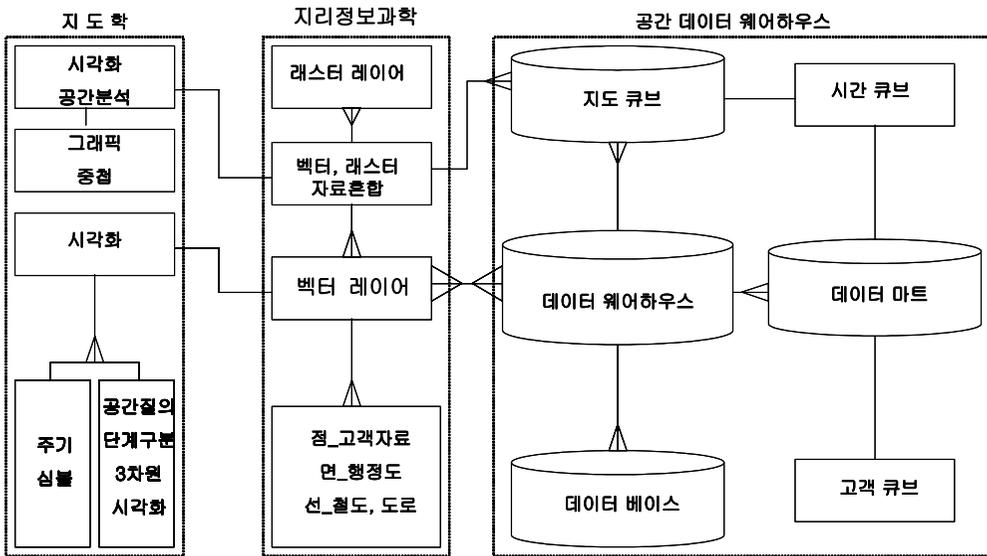


그림 3.10 공간 데이터 웨어하우징 논리적 체계

백화점 자료를 담고 있는 데이터 마트의 논리적 관계에서 고객과 상품은 거래라는 관계성을 맺는다. 고객과 상품은 N:M형식으로 연결되고, 거래시간과 지점과는 1:1로 연결된다. 그리고 고객과 거주공간은 거주라는 관계성을 맺으면서 1:1로 연결된다(그림 3.3).

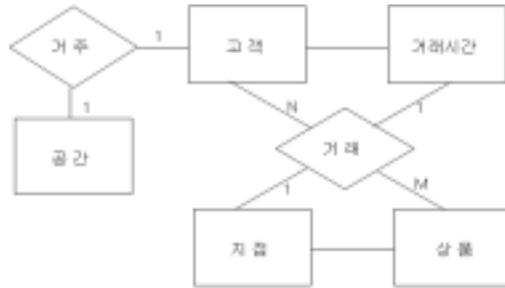


그림 3.3 백화점 자료 논리적 설계

### 3.2.2.2 물리적 설계

실험적 시스템의 물리적 설계는 공간 자료 저장소와 비 공간 자료인 백화점 자료로 구분하여 설계했다.

백화점 자료의 물리적 설계는 백화점 거래량을 사실 테이블로 데이터 마트를 설계하였다(그림 3.4). 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 스키마모델(snowflake schema model)로 설계했다. 그리고 주소차원의 고객ID는 공간 데이터베이스와 연결하여 지리정보과학에서 제공하는 공간 분석을 수행한다.

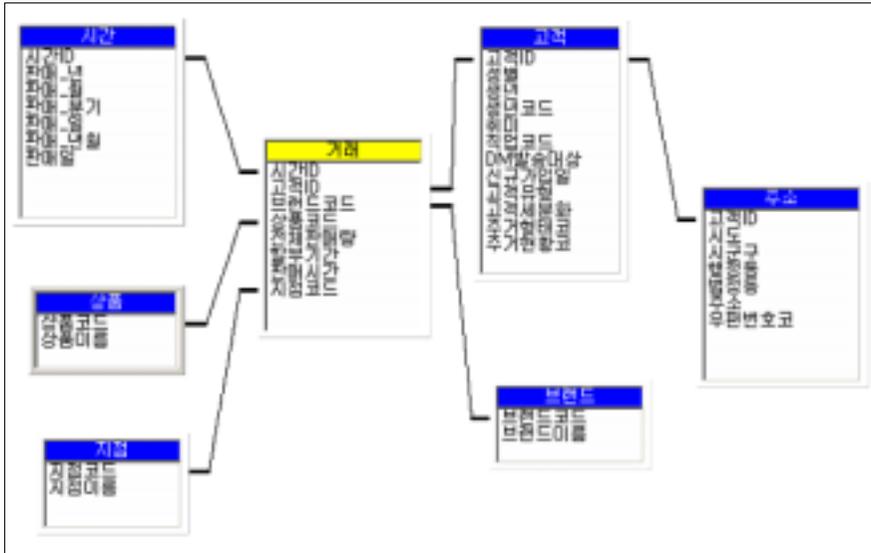


그림 3.4 백화점 자료 눈송이 스키마 모델(snowflake schema model)

본 연구에서 사실 테이블의 고객ID는 고객 신상자료의 질의과정에서 웹 기반 지도서비스에서 공간객체인 점 사상으로 표현하기 위해 지도큐브와 연결했다. 이 지도큐브는 공간 데이터베이스의 공간객체들이 저장되었다. 이러한 구조는 고객 신상자료가 저장된 데이터마트와 공간 데이터베이스의 점 자료가 지도에서 사상으로 나타내기 위해 1:1구조 cartesian products로 순서쌍을 이루어야한다. 그래서 백화점 거래 자료와 각 차원별 자료를 눈송이 스키마모델(snowflake schema model)로 설계하였다. 이 모델은 스타 스키마모델(star schema model)을 보완한 모델로 저장 공간을 최소화하고 유연성을 개선한 것이다. 그러나 성형 스키마모델보다 많은 조인(join)연산으로 인해 상대적인 느린 응답을 하는 단점이 있다.

### 3.3 자료 생성 및 변환과 시스템 구현

#### 3.3.1 전처리 과정

본 연구는 2000년 5월부터 2001년 4월까지 A백화점 서울지역의 4개 지점의 자료를 GIS의 다양한 단위(레이어)로 변형, 구현하였다. GIS의 레이어는 공간 자료와 관련 속성정보 구조로 된 여러 공간 객체를 복합적으로 구성하고 있다.

전처리 과정을 통하여 고객 자료는 공간 객체로 구현하고 문자로 된 백화점 자료와 구현된 GIS의 공간 자료의 관련 속성정보를 온라인 분석기법을 활용하여 다중 테이블 조인을 수행한다.

A백화점 자료는 카드를 이용한 카드 고객 자료와 거래 자료로 구성된다. 거래 자료는 거래별 거래 정보 자료, 거래시점, 상품 브랜드, 그리고 품목이 있으며 고객 자료는 성별, 생년, 고객유형, 주거형태 등 고객 신상에 관한 자료와 주소정보가 우편번호 코드체계로 구성되어 있다. 고객 자료는 50,000여건에 달하는 테이블과 거래, 지점, 브랜드, 상품, 직업을 설명하는 6개 테이블로 구성되어 있다. 그러나 GIS의 공간자료로 구성하기 위해 고객 자료의 우편번호 코드체계를 행정동 체계로 변형하고 고객 요약화와 세분화가 가능하도록 구별, 동별로 점, 면 사상을 이용하여 생성하였다. 또한 공간 분석을 수행하기 위해 사례지역을 서울시로 국한하여 서울시에 거주하는 41,263명의 고객 신상 자료를 대상으로 추출하여 공간 데이터 웨어하우스를 구현하였다.

본 연구는 구현된 데이터 웨어하우스에 공간 차원을 고려한 공간 데이터 웨어하우스로서 시간, 공간차원에 따라 다양한 수준에서 분석을 수행한다. 따라서 시간차원에서 년, 분기, 월별로 자료를 변형하고 공간 차원

은 구별, 행정동으로 변형하였다. 데이터 마트와 고객ID키로 연결된 주소 차원은 GIS의 공간 데이터베이스의 점사상의 속성 테이블로 연결하였다.

공간 데이터베이스의 고객 점 사상 위치는 고객의 위치와 위상관계를 담고 있으며 고객 데이터베이스와 연관되어 공간의사결정의 도구로 사용된다.

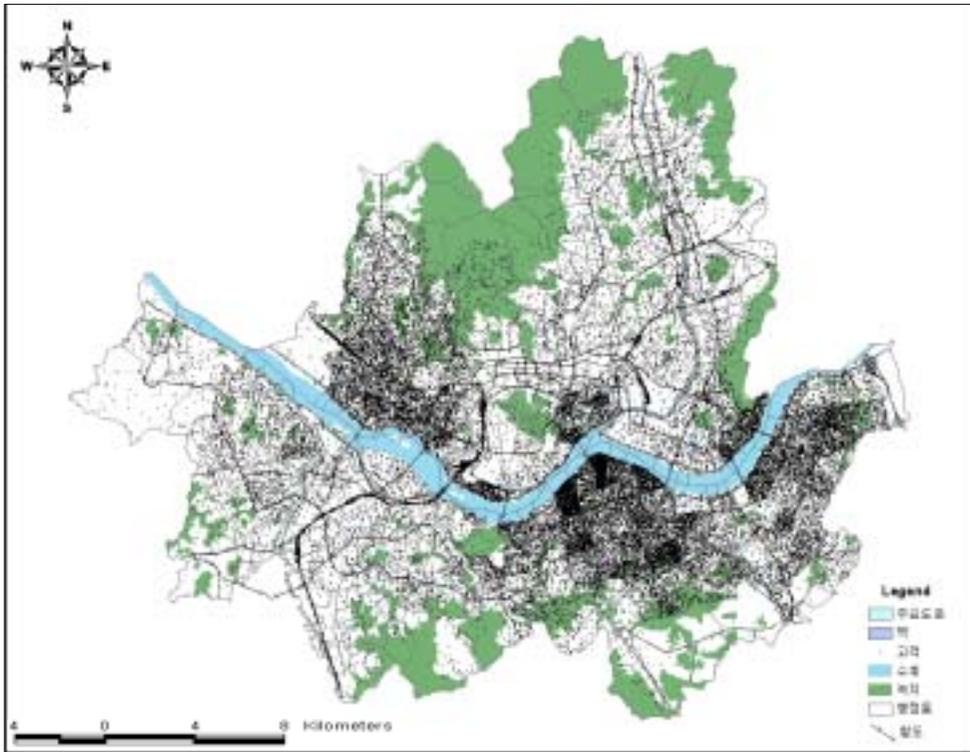


그림 3.5 연구지역의 A백화점 카드 고객 위치

그림 3.5는 연구지역에 고객의 위치를 점 사상으로 표현한 그림이다. 우선 서울시에 거주하는 고객만을 추출하여 GIS의 점 사상으로 표현하고 x, y좌표 값을 이용하여 동별로 구분하였다. 또한 면사상과 그에 연관된 속성 값으로 공간통계에 사용될 자료를 정량화하였다. 연구에 부수적으로 필요한 토지이용, 주요도로, 수계 등 다양한 GIS레이어는 수치지도와 공

간격체를 생성, 변형하여 구현하였다.



그림 3.6 영상 자료 전처리 과정

실험적 시스템에 사용된 래스터 자료는 위성영상을 사용하여 분석가에게 해당지역을 시각화한다. 본 연구에서는 LANDSAT TM<sup>43)</sup> 영상을 사용하였다. 일반적으로 인공위성의 관측센서에 의해 수집된 영상 자료는 그 밝기 영역이 컴퓨터의 시각화 영역과 일치하지 않고 기하학적 왜곡이 발생한다. 그래서 본 연구에서 사용할 영상자료는 시각적 판별력을 위해 영상의 밝기 특성과 포함된 시각적 분리도를 향상하고(그림 3.6 상단), 수치화된 서울시 지번도를 이용해서 지상기준점(Ground Control Point)을 선정하여 기하보정(geometric correction)을 하였다(그림 3.6 하단).

43) 미국 항공우주국(NASA)에서 자원관리와 지도화를 위해 운영하는 시스템으로, 한 영상은 남북 170km, 동서 185km의 지상면적을 래스터 자료형태로 기록한다.

### 3.3.2 메타데이터 작성

본 연구는 실험적으로 구현된 시스템에 사용된 수치화된 공간 자료, 비공간 자료의 명세서를 작성하였다. 그 문서는 구현된 목적, 제공되는 정보, 시-공간 범위, 자료형식, 웹상의 자료위치, 시스템 접근 권한, 문서 작성자, 좌표체계, 사용된 자료의 좌표범위, 공간 객체타입을 “디지털 지리공간 데이터에 대한 메타데이터 표준”에 따라 작성하였다(그림 3.7).



그림 3.7 메타데이터 문서

### 3.3.4 웹 기반의 시스템 구현 방안

본 연구에서 분석가는 구현한 실험적 시스템에 웹 환경에서 접근한다. 시스템의 H/W 및 S/W는 크게 세 가지로 분류된다. 공간 데이터 웨어하우스는 공간 자료와 비공간 자료를 저장, 관리하는 저장소로의 역할을 하며, 웹 환경 지도서버는 저장된 자료를 질의, 조회, 처리, 분석의 시각화 기능을 수행한다. 그리고 서버에는 Servlet 엔진을 지원함으로써 마케팅 의사결정자의 클라이언트 접근, 시각화하는 채널을 제공한다. 마지막으로 다차원 분석을 위한 OLAP엔진과 마이닝 도구는 분석가 클라이언트의 소프트웨어에서 서버모듈로 연결하여 저장되며 의사결정자에게 다차원분석과 데이터 마이닝 결과를 시각화하여 상담한다.

의사결정자의 요구에 따른 조회는 OLAP엔진에서 구조화된 질의 언어로 변경되고 공간 데이터 웨어하우스에 전달된다. 전달된 요구에 의하여 분석가의 분석결과는 ArcIMS의 서버모듈에서 Servlet엔진인 Jakarta-Tomcat에 전달되어 의사결정자에게 공간적 위치와 사상을 제공한다.

또한 경영정보시스템(MIS: management information system)의 거래 및 고객자료가 저장된 데이터 마트는 백터 레이어의 점 사상 공간객체에 각 차원별 주키(primary key)로 연결되어 요구에 따라 차원별 세분 및 요약되어 분석된다(그림 3.8).

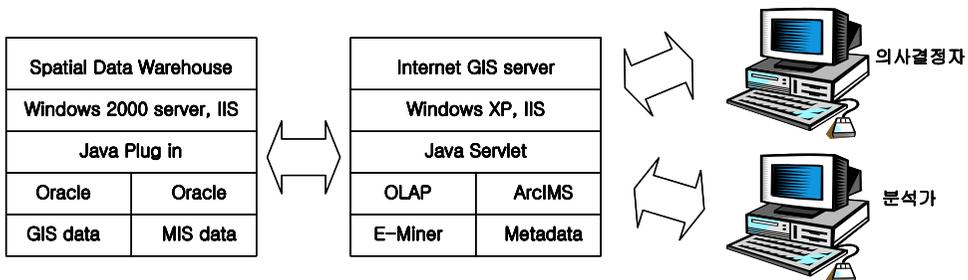


그림 3.16 실험적 시스템 소프트웨어 흐름도

웹 환경에서 구현된 실험적 시스템에 접근하기 위해 사용자 첫 화면을 구성하였다(그림 3.9 상단). 그림 3.9에서 A에 해당하는 부분은 메타데이터와 연결되며 B는 분석된 자료를 공간 자료의 형태로 표현하기 위한 수치지도로 연결된다. 또한 C는 비 공간 자료의 분석을 위한 데이터 마이닝 결과 페이지로 연결되고 D는 비 공간 자료를 테이블 형식으로 제공한다.

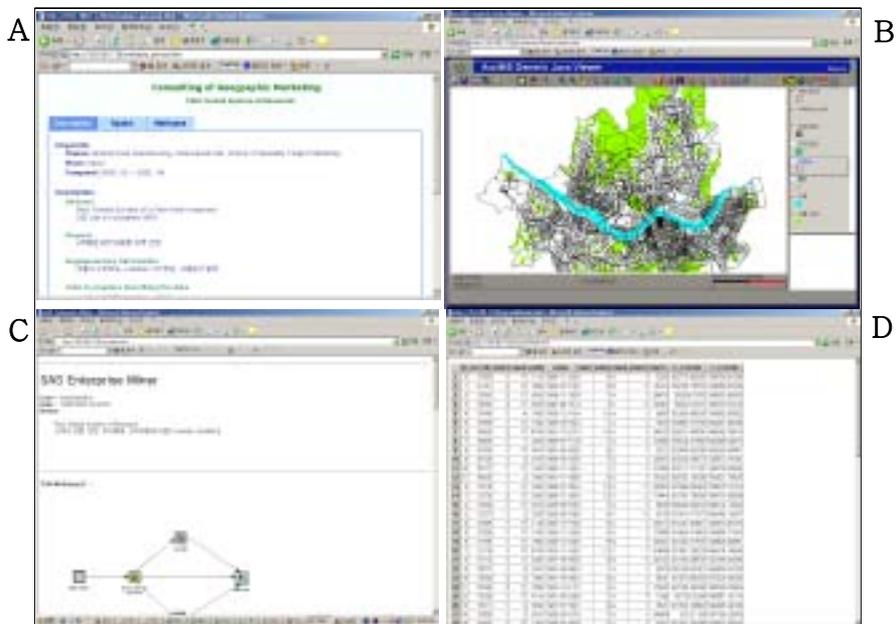
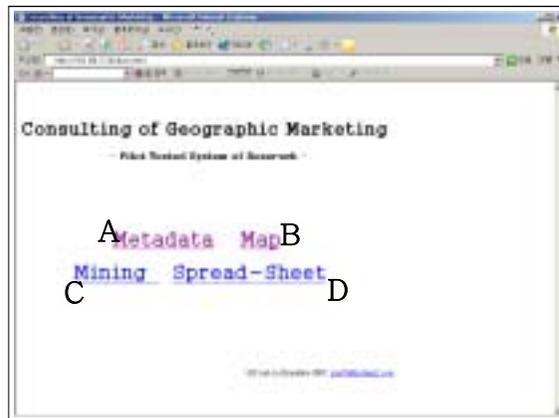


그림 3.9 사용자 접근 화면

## 4장 사례 분석

이번 장에서는 구현된 실험적 시스템에 실증사례를 적용하였다. 본 연구에서 구현된 공간 데이터 웨어하우스를 실제 백화점 마케팅 지역 선정에 적용함으로써 효율성을 검증하였다. 사례 분석에서 다루게 될 실험적 GIS상당 시스템은 실제 소매환경에서 분석가와 의사결정자가 웹 환경에서 의사결정을 위한 양방향 통신을 수행한다.

### 4.1 카드 구매 고객의 공간적 분포특성

#### 4.1.1 전역적 공간분포패턴

데이터 웨어하우스에 저장된 고객 신상정보는 구, 행정동과 같은 공간 단위를 기준으로 저장되어 있다. 이 고객의 위치와 관련된 속성정보를 활용하여 분석하기 위해 고객ID와 주키(primary key)로 연결된 공간 데이터베이스에 저장된 공간 객체를 사용하였다.

A백화점 고객의 공간적 분포 패턴을 파악하기 전에 고객의 밀집도를 대략적으로 살펴보기 위하여 단계구분을 하였다. 그림 4.1 좌측 상단은 행정동에 포함된 고객 수의 기술적 통계 정보를 제시한다. 서울시 522개동에 고객수의 최대값은 1246명으로 잠원동이며 한명의 고객도 없는 동은 삼선2동, 월곡4동, 동선2동, 가리봉2동, 강일동이다. 각 동에 대한 평균 고객수는 79명이고 표준편차는 121이다. 즉, 최대의 고객이 밀집한 지역은 잠원동, 압구정1동 등 주로 강남구와 서초구에 고객들이며 밀집해 있다. 지도위에 있는 팝업(popup)창은 수치지도에 관련하여 기술한 설명이다. 이 기능은 지식을 발견하기 위한 의사결정자와 분석가 사이에서 양방향 통신을 수행한다.

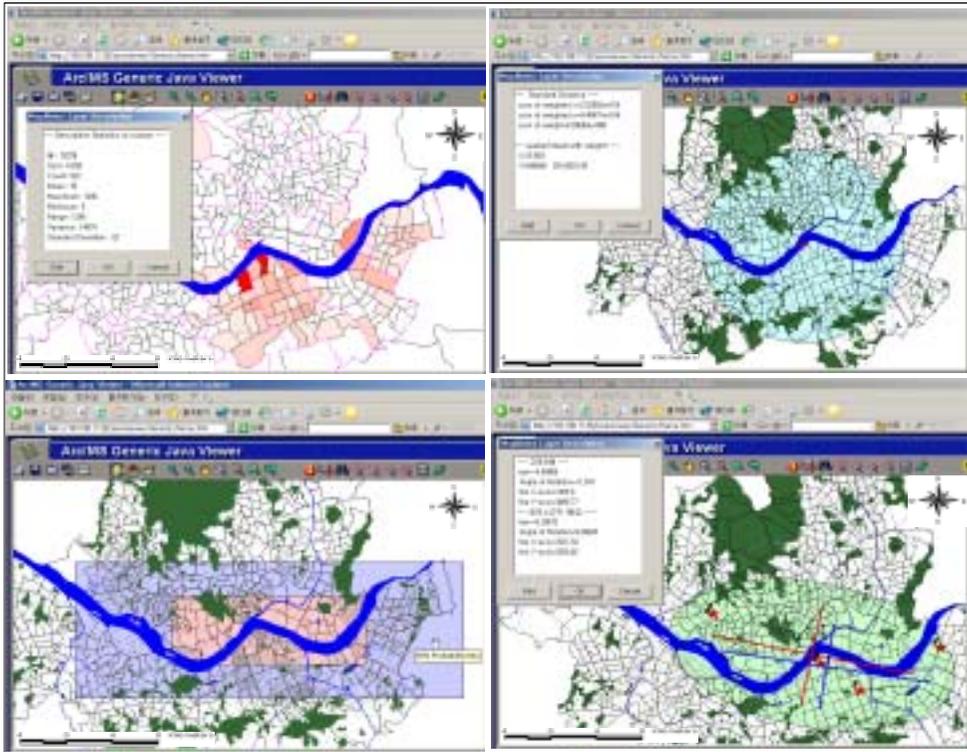


그림 4.22 고객의 공간적 분포 패턴

고객의 구매량에 따른 공간적 분포의 특성을 밝히기 위해서 점 분포의 중심 경향을 살펴보았다. 구매량에 따른 중심점(mean center)<sup>44)</sup>의 위치는 TM 좌표로 (313621, 548066)이며 신사동에 위치한다. 총 구매량에 가중치를 적용하여 생성한 표준거리(1)는 중심점을 중심으로 반경 8523.45 m이다(그림 4.1 우측 상단). 표준거리 내에 고객들은 구매량에 따른 분포를 표현한다.

44) 중심점(mean center)은 자료의 분포를 대표하는 값으로 분포의 중심(center of the distribution)이 어디에 있는가를 말한다. 공간 통계에서는 공간적 분포의 중심경향을 알아보기 위해 속성값에 가중치를 둔 가중치 중심점(the weighted mean center)을 이용한다(Machael, 1997).

$$SD = \sqrt{\frac{\sum_{i=1}^n f_i (x_i - x_{mc})^2 + \sum_{i=1}^n f_i (y_i - y_{mc})^2}{\sum_{i=1}^n f_i}} \quad \dots\dots(1)$$

SD = 가중화된 표준거리,  $f_i$  = 전체 구매량에 따른 가중치

확률지역(probability area)과 표준타원체는(standard deviation ellipsoid) 연구지역내의 점 사상의 공간적 밀집도를 효과적으로 시각화한다. 백화점 카드고객은 확률지역(probability area)통해 고객의 68%가 중심점을 중심으로 동서간격 약 6.5km 남북 약 2.3km지역에 위치하며 한강이 관통한다. 또한 고객의 95%가 서울시의 동서방향으로 늘어진 공간분포패턴을 보인다. 표준타원체를 통하여 고객 위치의 밀도 뿐 아니라 방향성도 알 수 있었다(그림 4.1 좌측하단).

그림 4.1의 우측하단의 전체 고객(붉은색)과 고객을 세분화 하여 최우수 고객 1등급(파랑색)의 밀집 패턴과 방향성을 비교 검토했다(표 4.1). 표준편차 타원체(standard deviational ellipse)의 특징은 점 패턴의 자료의 밀집도와 길이가 서로 다른 두개의 축(x, y)으로 방향성을 알 수 있다. 각 축은 점 자료의 밀집의 최대, 최소의 방향을 지정하며 'tan'의 수치는 양의 값이면 반 시계 방향으로, 음의 값이면 시계방향으로 기운다.

전체 고객의 방향성은 북서에서 남동 방향으로 밀집해 있고 신사동이 고객 분포의 중심지로 표현되었다. 전체 판매량에 최우수 1등급 고객의 분포형태는 장 방향 축의 길이가 전체고객이 축보다 길고 단 방향 축이 짧은 형태가 나타났다. 이러한 결과는 최우수 고객1등급의 고객이 동서 방향으로 더 길게 분포 했음을 알 수 있다. 또한 'tan'의 수치와 회전 값으로 남서 방향으로 치우쳐 분포 하고 있다.

표 4.1 고객 세분화에 따른 표준편차 타원체

	고객 전체	최우수 고객 1등급
가중치 중심점	신사동	압구정 1동
tan	-4.89458	-6.26672
회전	11.547	9.06643
장반경 길이	7425.6 m	7627.54 m
단반경 길이	3889.7 m	3250. 82 m

#### 4.1.2 공간적 고객 세분화

군집분석(cluster analysis)은 서로 연관성이 있는 자료들을 몇몇의 집단으로 묶어 각 집단의 성격을 파악하는 기법이다. 군집분석의 장점은 대량의 자료에서 주어진 내부구조의 사전 정보 없이 의미 있는 자료구조를 발견하고 모형화, 시각화하여 지식을 발견한다. 공간 군집분석은 공간자료들의 속성이나 특성에 대비하여 연관성이 있는 속성을 군집화한 후 지역적으로 분할하여 시각화하는 특성을 갖고 있다.

본 연구는 고객의 총 구매량에 따라 가중치가 적용된 티센 다각형(thiessen polygon)을 통해 공간 이웃을 정의하고 고밀도(high density) 군집과 저밀도(low density)군집패턴을 확인하였다. 또한 최 근린 계층 군집과(nearest neighbor hierarchical clustering) k-평균 군집(k-means clustering)을 통해 고객의 분포패턴을 조사하였다.

k-평균 군집은 분석가 임의의 그룹의 수를 정의하여 그룹의 중심에서 가까운 거리에 있는 군집에 할당하는 방식이다. 최 근린 계층군집(nearest neighbor hierarchical clustering)은 두개 이상의 사건을 군집으로 형성하고 이

후 처음 군집을 이용하여 두 번째 군집을 만드는 계층적인 방법으로 관련 속성의 집중된 지역의 고차원 군집을 생성한다.

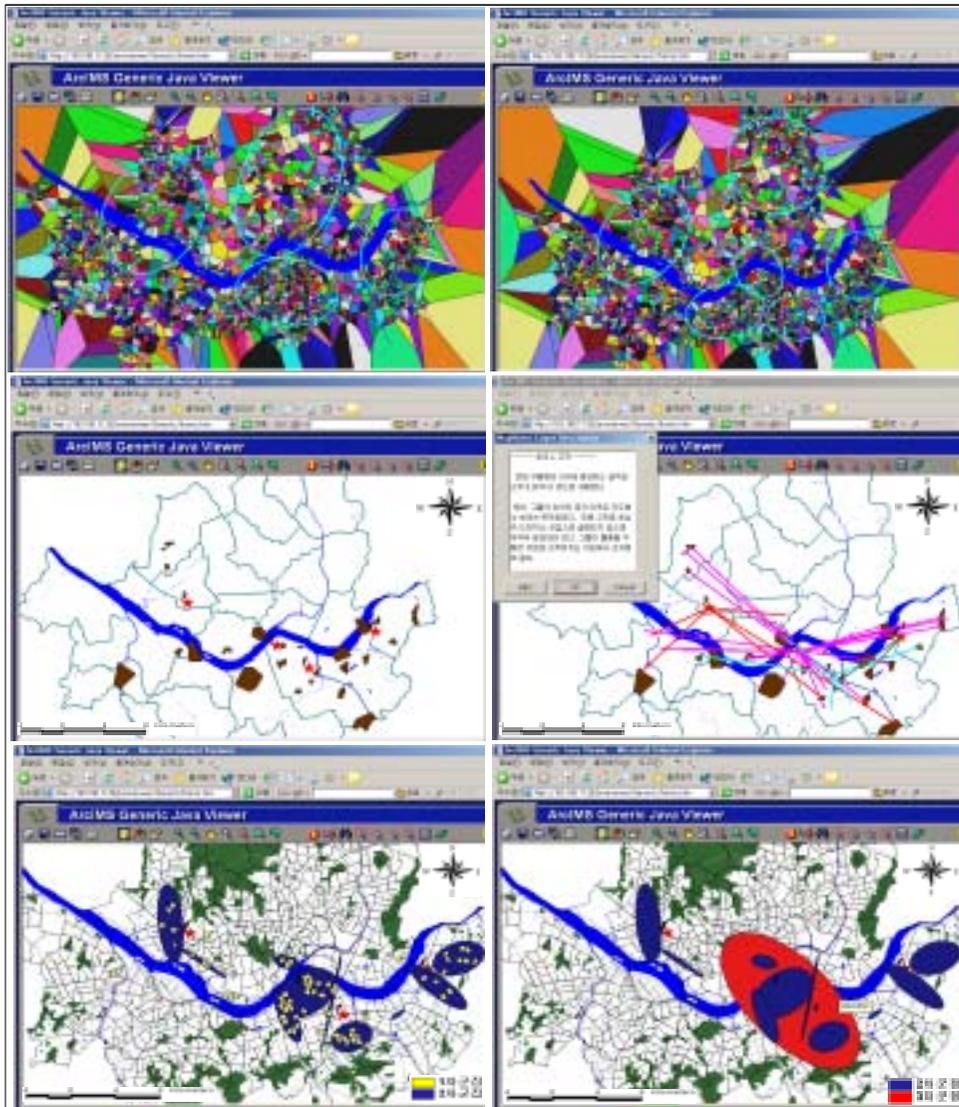


그림 4.2 공간이웃과 군집분석(티센 다각형과 4개의 K-평균군집 중첩(상좌), 티센 다각형과 6개의 K-평균군집 중첩(상우), 초우량고객위치(좌), 초우량고객의 카드거래 지점(우), 1, 2차 최 근린 계층군집(하좌), 2, 3차 최 근린 계층군집(하우))

고객의 전체 구매량에 따른 군집패턴의 결과로서 티센 다각형과 k-평균 군집을 4개(좌측), 6개(우측)로 군집분석을 수행하였다. 그리고 티센 다각형으로부터 얻은 면 자료를 이용하여 전체 구매량 10%의 고객의 분포 패턴을 조사한 결과 불규칙한 형태로 흩어져 있었다. 전체 구매량 10%를 구매한 고객의 전체 고객수의 0.26%이며 법 정동으로 평촌, 역촌, 응암, 구로, 여의도, 용산, 한남, 반포, 신사, 논현, 청담, 잠실, 신천, 천호, 세곡, 방이동에 위치하고 있다.

군집의 계층을 파악하기 위해 최 근린 계층군집을 수행하였다. 계층적 군집분석은 작은 군집부터 큰 군집으로 계층을 생성해간다. 이러한 원인은 군집분석이 감독학습(supervised learning)으로 절대적 기준 모델이 없기 때문이다. 그리고 계층적 군집을 이용하면 바람직한 군집의 수를 취할 수 있다. 이 연구에서는 공간적 근접성과 구매량, 구매회수에 3번에 걸친 최 근린 계층 군집분석과 k-평균 군집결과 강남구와 서초구에 고밀도로 밀집해 있음을 확인 할 수 있었다.

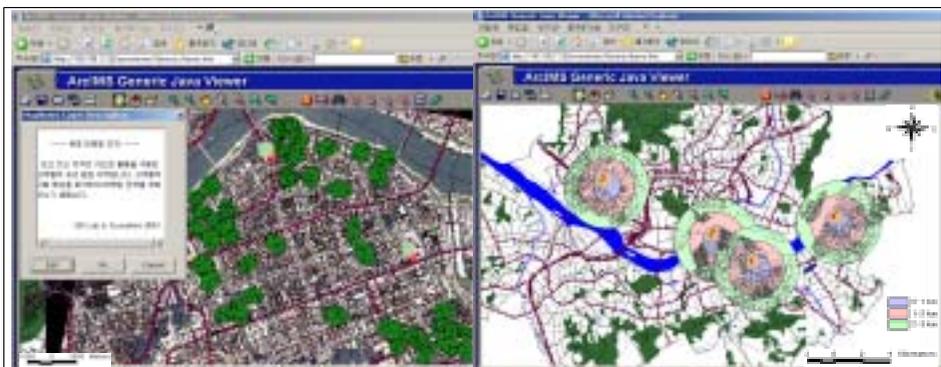


그림 4.3 목표 마케팅 지역과 각 지점 거리별 버퍼존

그림 4.3의 좌측은 고객이 최고 밀집지역으로 발견된 강남구, 서초구지역

으로 래스터 자료인 위성영상과 벡터자료를 중첩한 것이다. 기업은 각 지점의 위치적 특성 뿐 아니라 목표마케팅 지역을 시각화하기를 원한다. 이 연구의 실험적 시스템은 래스터 레이어를 통해 영상을 시각화하고, 비 공간 자료와 연계한 벡터자료를 중첩하여 상담한다. 그림 4.3의 우측은 41,236의 고객 신상 자료에서 거래 내역이 있는 고객자료를 활용하여 지점별 반경 1km 씩 3km의 버퍼존(buffer zone)을 형성하였다.

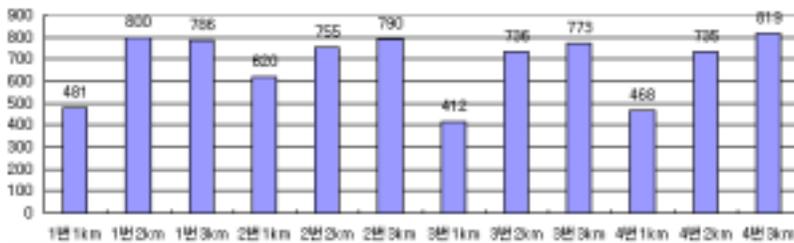


그림 4.4 각 지점과 거리별 고객 수

전체고객 중 각 지점의 3km 반경이내에 분포하는 고객은 42.4%인 6,351명이 밀집하여 분포 패턴을 보이고 있다. 각 지점의 반경 1km의 버퍼존 고객 수는 다음 그림과 같다. 각 지점에서 도보로 통행이 용이한 반경 1km 지역보다 2km, 3km지역에서 더 많은 고객이 구매빈도가 높은 것을 확인할 수 있었다(그림 4.4).

## 4.2 카드 고객의 구매 특성

### 4.2.1 카드 고객의 일반 현황

실험적 시스템은 A백화점의 고객 신상자료를 활용하여 거래 고객의 성향을 분석한다. 41,264명중 거래 자료가 있는 약 15,000건의 자료를 대상으로 구매 특성을 조사했다.

고객의 일반적인 특징은 여성고객 20대 중반에서 30대 초반이 가장 큰 구매력을 보였고, 연령이 증가할수록 구매량은 감소한다. 주거형태는 아파트가 단독 주택과 빌라에 거주하는 고객보다 높았다(그림 4.5). A백화점 카드 구매 고객의 성별 특징은 다음과 같다. 전체 카드고객 중 여성이 69.2%로 남성 고객보다 38.4%정도 많았다. 평균 연령은 남성과 여성이 43.2세, 38.4세로 남성고객이 여성고객보다 5세 가량 높게 분포한다. 또한 전체 판매량에서 남성고객의 29.16% (506,363,870원)에 비해 여성고객은 70.84% (1,230,057,790원)로 여성 고객이 남성보다 41.68% 더 많이 구매하였다.

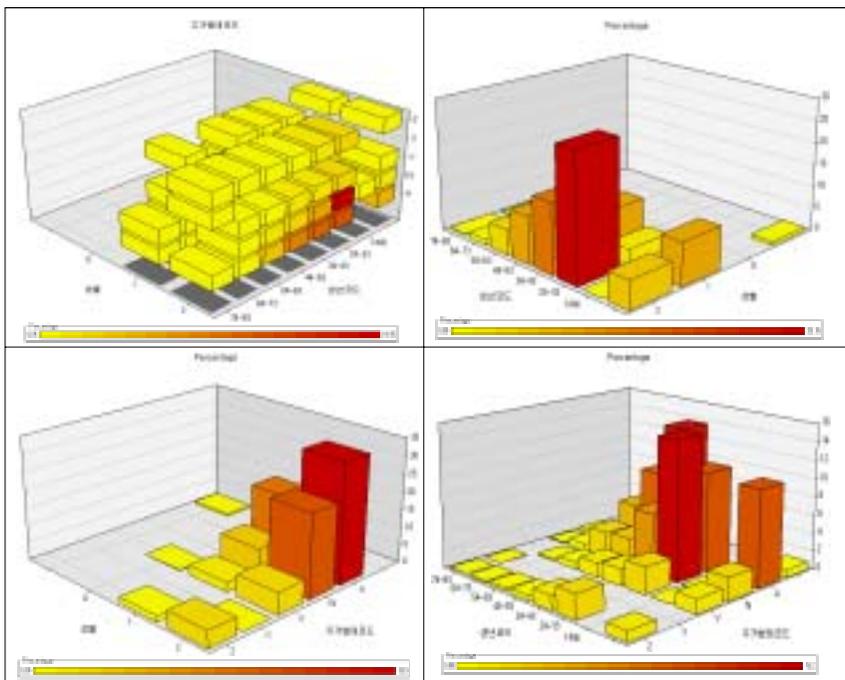


그림 4.5 구매 고객의 일반 현황(주거형태, 성, 연령별(상좌), 성, 연령(상우), 주거형태, 성별(하좌), 주거, 형태별(하우))

표 4.2 고객의 성별 자료 분석

성 별	고객 수	평균 연령	연령 표준편차	총 구매액(원)
남	12,498	43.2	11.02	506,363,870
여	28,764	38.4	10.75	1,230,057,790
계	41,262			1,736,870,660

표 4.3 주거형태 자료 분석

주거 형태	수	평균 거래량(원)	총 구매량(원)
아파트	8,094	120,402	974,538,510
빌딩	1	99,000	99,000
병원	1	79,600	79,600
단독주택	4,069	108,462	499,902,850
빌라	1,388	80,081	152,793,290
기타	882	124,111	109,466,410
계	14,435		1,736,879,660

#### 4.2.2 시공간적 거래패턴

이번 장에서는 본 연구의 실험적 시스템을 활용하여 고객의 거래패턴을 살펴본다. 그리고 분석가는 공간차원에서 지역을 세분화하여 발견된 목표 마케팅 지역의 고객의 구매패턴을 검토하고, 선호하는 상품정보를 발견한다.

전체 고객의 지점별 판매액은 서울시 4개지점 중 2번 지점, 시간차원을 계층적 세분화한 수준으로 분기별 판매횟수는 2분기에 가장 높았다(표 4.4). 그러나 총 판매량은 연말이 포함된 10~12월에 높은 거래량을 나타낸다(표4.5).

표 4.4 지점별 거래량

지 점	수	전체 거래량(원)
1	3,235	314,194,450
2	7,963	1,036,847,930
3	1,499	187,519,800
4	2,278	198,317,480
계	14,975	1,736,879,660

표 4.5 분기별 거래량

지 점	수	전체 거래량(원)
1분기	3,514	401,895,620
2분기	4,098	404,664,670
3분기	3,400	396,928,830
4분기	3,963	533,390,540
계	14,975	1,736,879,660

표 4.6 일별 거래량

일	수	전체 거래량(원)	평균 거래량(원)	판매량 표준편차
1-10	6,205	758,293,300.00	1,214,668.50	5,622,251.68
11-20	4,382	516,099,720.00	1,180,478.35	3,975,964.06
21-31	4,388	462,486,640.00	1,174,363.61	6,004,319.23
계	14,975	1,736,879,660	3,569,510.46	15,602,534.97

시간 차원에서 일별 거래량은 1-10일 사이에 가장 많은 구매현상이 발생하고(표 4.6), 시간에 따른 거래량은 15시에서 19시 사이에 가장 활발한 구매현상이 발생한다(그림 4.6).

표 4.7 시간에 따른 거래량

시 간	수	전체 거래량(원)
9시-12시	739	17,421,080
13시-15시	2,363	21,590,390
15시-19시	4,522	174,624,180
19시이후	321	3,835,190
계	14,975	1,736,879,660

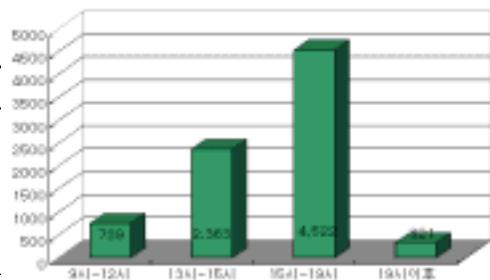


그림 4.6 시간대별 거래 고객 수

### 4.3.3 목표 마케팅 지역의 다차원적 분석

앞의 장에서 군집분석의 결과로 고객이 고밀도로 밀집된 지역을 강남구와 서초구로 밝혀내고 목표 마케팅 지역을 선정했다. 목표 마케팅 지역의 고객들의 구매 특성은 강남, 서초구의 5,103명중 52%인 2,697명이 2번 지점에서 상품을 구매하였다. 목표 마케팅 지역의 의사결정을 위한 분석가는 우수 고객의 상품의 항목을 발견하여 마케팅에 이용한다. 강남, 서초구의 OLAP 기법을 활용하여 공간차원의 동별, 시간차원의 2000년 월별, 상품과 거래량 자료를 활용하여 고객을 세분화하였다.



그림 4.7 목표 마케팅지역의 다차원적 거래특성

그림 4.7은 공간 데이터 웨어하우스에 구현된 A백화점 각 데이터 마트에 다차원 접근을 시도하여 초우량등급 고객의 구매 특성을 조사한 것이다. 이러한 고객 세분화는 고객의 선호품목과 구매량에 따라 행정동 수준의 공간 차별적인 마케팅 전략을 수립한다. 이 전략은 서로 다른 행정동에 거주하는 고객들의 브랜드별 구매습관을 시계열적으로 관찰하고, 기업의 이익증대를 위한 차별적인 고객관리를 가능하게 한다.

```

/*****
강남, 서초구의 고객 중 2번 지점을 대상으로 2000년 12월에 구매한 고객을 검색하라. 그리고
그 고객이 구매한 상품명과 전체 판매량을 행정동으로 오름차순 정렬하라. - GIS Lab in KHU -
*****/

SELECT AL2.행정동, Count (AL2.행정동), AL5.상품명, AL3.전체판매량
FROM 고객 AL1, 주소 AL2, 거래 AL3, 시간 AL4, 상품 AL5
WHERE (AL1.고객ID=AL3.고객ID AND AL5.상품코드=AL3.상품코드
AND AL4.시간ID=AL3.시간ID AND AL2.고객ID=AL3.고객ID
AND AL2.고객ID=AL1.고객ID) AND (AL2.시군구='강남구'
OR AL2.시군구='서초구' AND AL1.성별=2 AND AL4.판매_월='12'
AND AL1.고객유형='A' AND AL1.고객세분화=1)
GROUP BY AL2.행정동, AL5.상품명, AL3.전체판매량
ORDER BY 4 DESC, 2

```

그림 4.8 목표 마케팅지역 다차원적 분석을 위한 구조화된 질의 언어

초우량 1등급 고객은 2번 지점에서 가장 많은 거래량을 기록했다. 2번 지점에서 고객은 1년동안 849,399,250원을 구매했다. 그리고 구매 패턴을 세분화하여 구매한 상품 정보를 발견하였는데, 주로 유명 디자이너 부띠끄 품목과 수입명품이 높은 판매실적을 기록했다. 이러한 초우량 1등급 고객은 고객 타입을 세분화한 분류에서 최상의 판매실적을 기록한 고객들이며 전체 판매량의 10%를 구매한 고객들이다(그림 4.9).

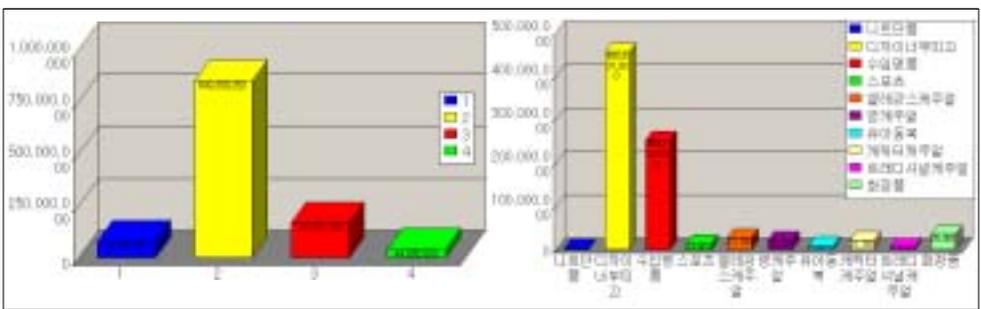


그림 4.9 초우량 1등급 고객 구매특성

그림 4.9는 다차원 접근을 수행하여 공간 데이터 웨어하우스의 자료로부터 정보를 추출한 그림이다. 이들의 구매 특성을 발견하기 위한 구조화된 질의 언어는 그림 4.10과 같다.

```

*/*****
서울시 4개 지점의 고객 중 초우량 1등급 고객이 구매한 브랜드 이름과 판매량, 상품명을
행정동으로 오름차순 정렬하라. - GIS Lab in KHU -
*****/

SELECT AL6.행정동, Count (AL6.행정동), AL1.전체판매량, AL3.브랜드이름, AL4.상품명
FROM 거래 AL1, 고객 AL2, 브랜드 AL3, 상품 AL4, 주소 AL6, 주소 AL7
WHERE (AL2.고객ID=AL1.고객ID AND AL3.브랜드코드=AL1.브랜드코드
AND AL4.상품코드=AL1.상품코드 AND AL6.고객ID=AL1.고객ID
AND AL6.고객ID=AL2.고객ID AND AL7.고객ID=AL1.고객ID
AND AL7.고객ID=AL2.고객ID) AND (AL2.고객세분화=1
AND AL1.전체판매량>=500000)
GROUP BY AL6.행정동, AL1.전체판매량, AL3.브랜드이름, AL4.상품명
ORDER BY 1, 3

```

그림 4.10 초우량 1등급 고객구매 분석을 위한 다차원 접근

거래된 상품은 브랜드별 연관규칙이 발생하는데, 그림 4.11은 목표 마케팅 지역에서 구매한 상품의 연관규칙을 설명한 것이다. 이러한 연관성 분석은 거래에 포함된 두 사건 이상의 상호관련성을 발견하는 것으로 평가 기준으로는 지지도<sup>45)</sup>, 신뢰도가<sup>46)</sup> 있다. 목표 마케팅 지역에서 가장 높은 구매 빈도는 “지오다노 ==> 샤넬” 브랜드이며 지지도는 50.31%이며 신뢰도는 71.93%이다. 그 반대의 구매경우는 지지도는 같고 신뢰도는 75.23%로 3.3%의 가능성이 더 크다. 또한 ‘폴로 & 샤넬’을 구매한 고객이 ‘지오다노’ 브랜드를 구매할 가능성은 80.28%로써 매우 높은 신뢰도를 나타내고 있다.

45) 지지도(support)는 두 사건이나 거래가 일어날 확률로 본 연구에서는 두 상품의 동시 구매빈도를 측정하는데 사용되며 상호 대칭적이다.

46) 신뢰도(confidence)는 첫 번째 상품이 구매되었을 때 또 다른 상품이 구매되는 조건부 확률의 지표로써 상호 대칭적이지 않다.

Reliability (R)	Success (%)	Confidence (C)	Transaction	Date
1	100	71.91	75.00	75.00
2	100	48.15	75.00	75.00
3	100	48.15	75.00	75.00
4	100	48.15	75.00	75.00
5	100	48.15	75.00	75.00
6	100	48.15	75.00	75.00
7	100	48.15	75.00	75.00
8	100	48.15	75.00	75.00
9	100	48.15	75.00	75.00
10	100	48.15	75.00	75.00
11	100	48.15	75.00	75.00
12	100	48.15	75.00	75.00
13	100	48.15	75.00	75.00
14	100	48.15	75.00	75.00
15	100	48.15	75.00	75.00

그림 4.11 구매 브랜드의 연관 규칙

이 절에서 목표 마케팅 지역에서 다차원분석을 통해 각 차원별 고객을 세분화하여 조사하였다. 공간데이터 웨어하우스의 사실자료인 거래현황에 각 차원별 세분화, 요약화는 특정한 가정이나 추측이 아닌 저장된 자료에서 상이한 분석결과를 얻을 수 있었다. 발견된 지식은 고객ID와 연결된 공간자료로 표현가능하다. 또한 의사결정자와 분석가는 인터넷 기반 수치지도의 "MapNote"를 통해 상담이 가능하며, 기업의 각 지점별 마케팅 전략을 수립할 때 시공간 차별화가 가능하다. 그리고 구매 고객의 세분화를 통해 미래의 구매행위를 예측하고 대비할 수 있다.

#### 4.2.4 평가

이상에서 실험적 시스템을 활용하여 사례분석을 통한 결과는 시공간적 특징을 고려하여 다차원적 분석을 통해 민간기업인 백화점에서 발생한 거래자료를 기반으로 고객을 세분화하였다. 의사결정자는 공간적 세분화를 통해 서로 다른 형태의 집단에 속한 고객을 관찰하고 시간에 따른 거래특성과 변화습관을 예측할 수 있다. 이러한 예측은 시공간차원의 고객 수익성을 발견하여 차별적 마케팅 전략수립이 가능한 새로운 서비스 모델을 제시하였다.

## 5장 결론

이 연구는 인터넷 환경에서의 상호 운용성을 통해 대용량의 지리공간자료를 활용하여 효과적인 공간적 의사결정을 위한 지식발견에 초점을 맞추었다. 또한 실세계에서 발생하는 현상을 더욱 사실적으로 분석하고 표현하기 위해 지리정보과학의 지리정보시스템과 고객관계관리의 요소를 삽입하여 공간 데이터 웨어하우스를 설계 및 구현하였다. 본 연구의 실험적 시스템을 활용한 사례분석은 민간부분에서 활용을 위한 서비스모델에 목표를 두고 있다. 지금까지 인터넷 환경에서 공간 데이터 웨어하우스에 관한 연구는 대부분 공공부분의 관리적 측면에 국한되어 왔고 분석적 측면에 대한 연구는 미비한 실정이다.

본 연구에서는 인터넷 환경에서 공간자료를 활용하기 위한 방법의 유형을 다음의 세 가지로 구분하고 그 특징과 한계를 분석하여 대안을 제시하였다. 첫째, 기존의 인터넷 기반 지리정보서비스는 의사결정을 위해 공간자료를 활용하여 분석하는 측면이 취약하다. 그 원인은 단순 위치정보와 관련 속성만을 제공하는 일방적인 서비스에 기인하며, 무엇보다도 공간자료의 고정된 타입의 자료 분류법이 적용되기 때문이다. 둘째, 공간자료와 해당속성자료를 제공하는 데이터베이스의 물리적 스키마설계는 실세계의 개념적 모형을 설명하기 불충분하다. 즉, 특정한 목적을 갖는 시스템에서 각 차원별 해상도에 따른 유연한 분류, 분석방법론을 지원하지 못하며, 공간에서 발생하는 다차원적 사상(feature)을 표현하기 어렵다. 셋째, 최근 민간부분은 다양하고 급격하게 증가된 고객의 욕구를 지리자료에서 공간적 측면의 해결책을 요구하고 있다. 그러나 이러한 맥락에서 다양하고 세분화된 고객의 욕구를 수치화된 공간자료만으로 해결하기에는 한계를 갖고 있으며, 그 한계는 연속적인 시간적 특성을 이산적인 표본으로 측정

하는 방법, 특정 주제에서 시간에 따른 공간 해상도를 고려하지 않는 문제 등 한계를 갖는다.

기존의 인터넷 환경에서 공간자료의 활용에 관한 연구와 사례분석을 통해, 본 연구에서는 민간기업의 활용성에 초점을 맞추어 실제 서울시 백화점 카드거래 자료를 활용하여 인터넷 환경에서 공간 데이터 웨어하우스의 다차원적 모델링을 구현하고, 기존 활용사례의 노출된 한계를 극복하기 위해 실험적 시스템의 설계 방향을 설정하였다.

첫째, 네트워크상에서 전송되는 지리자료의 초기자료를 벡터타입의 공간객체로 표현하고, 관련 속성자료를 연결하여 저장하였다. 분석가는 실험적 응용시스템에서 지리적 사상을 갖는 벡터자료를 Java Servlet Connector로 연결하고, 개방형 GIS 컨소시엄(OGC: open GIS consortium)에서 논의된 확장된 언어의 지도형식인 axi(ArcXML)형식으로 변형하여, 초기자료에는 손상이 없으면서 지식발견을 위한 다양한 분류와 분석이 수행가능 하도록 구성하였다. 둘째, 시공간적 차원과 더불어 특정주제에 맞는 차원을 고려하기 위해 각 차원별 데이터베이스에서 추출 및 변화 등의 정제과정을 거쳐 데이터 마트를 구성하여 원 자료를 보호하고, 각각의 데이터 마트는 통합화과정을 거쳐 공간 데이터 웨어하우스에서 모델링하였다. 물리적 스키마 설계는 일반적인 데이터 웨어하우스에서 사용되는 확장성이 떨어지는 스타스키마모델을 배제하고 공간객체가 저장된 공간자료 집합체와 연결하기 위해 눈송이스키마모델을 사용하여 저장 공간을 최소화하여, 유연성을 극대화하였다. 셋째, 분석가는 OLAP기법을 통하여 공간 데이터 웨어하우스에 다차원적 접근을 시도하며, 각 차원별 계층적 수준을 두어 공간적 지식발견을 위해 의사결정자의 정의에 따른 차원변경, 세분화, 요약화가 가능하도록 구현하였다.

본 연구에서는 이와 같은 설계 방향에 따라 공간 데이터 웨어하우스의 관리적 측면으로 실험적 시스템 구축목적, 제공되는 새로운 서비스의 모델, 사용되는 자료를 설명하기 위해 공간 메타데이터 문서를 제공하였다. 또한 분석적 측면에서 백화점 카드고객의 “목표 마케팅 지역설정”이라는 목적에 따라 공간적 의사결정과정을 위해 인터넷 상담시스템을 제안하였다. 이 인터넷 상담시스템은 공간적 의사결정을 위한 지식발견을 목적으로 의사결정자와 분석가 사이의 양방향 통신이 가능하며, 설정된 목표를 위해 지리정보과학의 공간분석기법과 확정적 자료 분석기법인 공간통계를 활용하였다. 더불어 비공간 자료의 거래, 고객신상자료를 활용하여 경영정보시스템의 고객관리기법 중 발견적 자료기법인 데이터 마이닝을 활용하였다. 다시 말하면 지리정보과학의 GIS와, 경영학의 MIS 분석방법융합은 백화점 카드고객자료를 활용하여 마케팅을 위한 새로운 서비스 모델을 제시하였다. 새로운 서비스 모델은 공간 데이터 웨어하우스를 활용한 분석적 측면을 설명하며, 이를 중심으로 연구내용을 요약하고 그 결과를 종합하였다.

첫째, 백화점 카드 거래분석에서 구매 고객의 공간적 분포패턴을 분석하기 위해 우편번호코드체계로 된 거주지 정보를 공간 객체의 점자료로 생성하고 도로, 철도, 등의 자료를 변형 생성하여 위상관계를 부여하였다. 이를 통해 카드구매 고객의 공간자료가 추상화단계를 거쳐 정제된 후 데이터마트에 저장된다. 공간객체인 점자료는 고객ID를 주키로 고객신상자료와 연결된다. 이러한 공간자료를 통해 기술적 통계로 대략적인 행정동별 고객수를 단계구분도로 시각화하며, 집중 경향치와 고객 분포의 중심점을 설정하였다. 여기에 구매량에 따른 고객을 세분화하기 위해 구매량에 가중치를 두어 tan의 회전각을 통해 구매고객 분포방향을 설정하고 표준편차타원체의 장반경, 단반경 축의 길이를 통해 분포의 분산, 밀집정도를 시각화하였다. 또한 각 급간에 속해있는 세분화된 고객들의 분포지역

과 밀집지역을 조사할 수 있다. 이 활용방법은 웹 기반 지도서비스에서 의사소통을 함으로써 의사결정자가 원하는 지식과 정보를 설명한다. 또한 공간자료를 활용하여 분석함으로써 백화점 카드 고객을 구매량에 따라 공간적으로 세분화하여 목표마케팅을 위한 새로운 대안을 제시한다.

둘째, 데이터 마이닝의 군집분석을 적용하여 고객충성도가 높은 지역을 세분화하여 시각화하고, 차별화된 고객 수익성이 높은 지역을 예측할 수 있다. 공간통계를 포함한 통계기법은 지식발견기법 중 분석가의 예측을 검증하거나 오류를 지적하는 확정적 분석방법이며, 거래행위에 따른 각 차원별 추측을 검토, 예측에 국한되어 있다. 본 연구에서 사용한 티센 다각형을 이용한 근접이웃과 K평균군집, 계층적 군집방법을 통해 공간 데이터 웨어하우스에 접근하여 다차원모델 혹은 공간, 비공간 자료들의 연계에서 의미 있는 분포패턴이나 구매량에 따른 연관성 있는 군집을 형성하고, 계층적 군집분석을 통하여 특정 변수에 따른 최고 밀집지역을 발견하였다. 그리고 래스터 자료인 위성영상을 벡터 자료와 중첩하여 제공함으로써 기업의 입지에 따른 주변지역을 시각화한다. 그리고 각 지점으로부터 거리에 따른 버퍼존을 형성하여 고객수를 분석하였다. 여기서 도보로 통행이 가능한 지점으로부터 반경 1km지역보다 반경 2km, 3km에서 많은 고객이 구매할 것을 발견할 수 있었다.

셋째, 이 연구에서는 “목표마케팅 지역설정”이라는 목표를 정하고 사례 분석을 실시하였다. 앞서 지리정보과학의 영역의 분석에서 고객 세분화에 따른 공간적 분포패턴과 고객이 밀집된 지역을 선정하였다면, 목표지역의 세분화된 고객들과 상호작용을 촉진시키기 위해 선호하는 브랜드와 품목을 시간대별로 세분화하여 분석하여야한다. 여기에는 OLAP기법을 통해 공간, 시간 등 다차원적 분석을 통해 분석하고 개인화에 따른 고객의 욕구를 위한 연관규칙을 분석하였다. 거래에 포함된 연관성분석은 두 브랜

드별, 품목별 상호관련성을 지지도와 신뢰도를 통해 목표마케팅지역의 세분화된 고객들의 요구에 유연하게 대처할 수 있는 환경을 제공한다.

본 연구의 향후 발전 방향은 다음과 같다.

우선 공간 데이터 웨어하우스 스키마설계에서 실제사례에 적용하여 계속적인 서비스 모델을 제시하여야 한다. 공간 데이터 웨어하우스는 특정한 목적을 갖고 설계되므로 실세계에서 발생한 관련된 차원들을 더욱 구체화하고, 시스템에 적용해야한다. 그리고 시공간적 특성을 명확하게 분석하며, 해당하는 측정방법과 의미 있는 지식을 발견할 수 있는 방법론을 개선시켜야 한다. 또한 본 연구에서는 공간자료의 한 부분인 래스터 자료의 시각적인 측면만을 고려하였다. 이 연구에서 고려한 공간 데이터 웨어하우스의 눈송이스키마설계는 확장성이 뛰어나며, 위성영상분석과 연계할 경우 더욱 실세계를 구체적으로 묘사할 수 있을 것이다. 이것은 구체적인 활용측면에서 공간의사결정을 위한 시스템에 기여할 수 있을 것으로 전망한다.

## 참고문헌

### 국내문헌

- 강영옥, 박수홍, 1999, “서울시 지리정보시스템 구축에 관한 연구”, 「서울시 정개발연구원」, 시정연 99-R-29.
- 강현철, 한상태, 최종후, 1999, “SAS Enterprise Miner 4.0을 이용한 데이터 마이닝: 방법론 및 활용”, 「자유아카데미」.
- 김진욱, 2000, 「데이터 웨어하우징의 효익과 손실 및 산업 추세」, 서울대학교 석사학위논문.
- 임은선, 2001, 「생활 폐기물 관리를 위한 공간적 의사결정지원시스템: 폐기물 수거경로계획과 소각시설 입지선정을 사례로」, 건국대학교 박사학위논문.
- 오충원, 2002, 「지리정보시스템의 시간 요소에 관한 연구: 안양시의 지가 변동 분석을 사례로」, 서울대학교 박사학위논문.
- 정병순, 신창호, 2002, “수도권 정보통신산업 클러스터의 지역거버넌스”, 「지역연구」, Vol.29 pp.29-53.
- 주성재, 2001, “산업집적지 조성의 필요성과 과제”, 「지리학총」, pp.1-10.
- 채승경, 2001, 「데이터마이닝을 이용한 웹 데이터 분석」, 고려대학교 석사학위논문.
- 최경희, 2003, 「GIS와 공간 데이터 마이닝을 활용한 은행 연구 분석」, 경희대학교 석사학위논문.
- 황만익, 2002, “디지털 국토통계지도 제작에 관한 연구”, 「국립지리원」.
- 황철수, 2001, “GIS메타데이터 구축의 필요성과 기본적 요구 분석”, 「지리학총」, Vol. 29, pp.11-24.
- 황철수, 1998, 「분산형 수치지도의 설계와 구현」, 서울대학교 박사학위논문.

## 국외문헌

- Agrawal, R., Imielinski, T., and Swami, A., 1993, Mining Association Rules between sets of items in Large Database, *ACM SIGMOD*, pp.207-216.
- Agrawal, R., 1998, Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications, *ACM SIGMOD International Conference on Management of Data*, p.7.
- Bedard, Y., 1999, "Visualization modelling of spatial databases towards spatial extensions and uml", *Geomatica Tutorial*, URL:<http://www.pcigeomatics.com/freeware/freeware.html>
- Barclay, T. et. al., 1998, "The Microsoft TerraServer", Technical Report MSR-TR-98-17, United States Geological Survey and Microsoft research WA98052, URL:<http://terraserver-usa.com>.
- Berry, M.J.A., and Linoff, G., 1997, "Data Mining Techniques: For Marketing, Sales, and Customer Support", John Wiley & Sons, Inc.
- Boyles, D., 2002, "GIS means Business: Marketing Communication, Targeting Marketing, Market Research, Internet Map Publishing", ESRI Press.
- Chawla, S., Shekhar, S., and Ozesmi, U., 2001, "Modeling Spatial Dependences for Mining Geospatial data: An introduction", In Miller, H.J. and Han, J., (eds) *Geographic Data Mining and Knowledge Discovery, London: Taylor and Francis*, pp.4-15.
- Codd, E.F, 1995, "Twelve rules for on-line analytic processing", Pushing Zahang Technical Report TR 01-015, Univ of Minnesota, URL:<http://citeseer.nj.nec.com/344983.html>.
- Crompvoets, J., and Bregt, A., 2001, "World Status of National Spatial Data Clearinghouses", *URISA Journal Crompvotes, Bregt*, Vol.

- 15(1), pp.43-49.
- Ding, Q., 2002, "Association Rule on Remotely Sensed Imagery using P-tree", *A Dissertation Submitted to the Graduate Faculty of the North Dakota State University of Agriculture and Applied Science*.
- Dodge, G., and Gorman, T., 2000, "Essential Oracle 8i Data warehousing", John Wiley & Sons, Inc.
- Do-Heon, L., 1998, "Data Mining Concepts and Reserach Trends", *KIS SIGDB Tutorial*.
- ESRI, 1997, "Spatial Data Warehouse", in ESRI White Paper, URL:<http://www.esri.com>.
- ESRI, 1998, "Spatial Data Warehousing for Hospital Organization", An ESRI White Paper, URL:<http://www.esri.com>.
- ESRI, 1999, "Power Your Business with ESRI GIS", An ESRI White Paper, URL:<http://www.esri.com>.
- Estivill-Castro, V., and Lee, I., 2000, Amoeba: Hierarchical Clustering Based on Spatial Proximity using Delaunaty Diagram, pp 1-16.
- Ester, M., Hans-Peter, K., and Sander J., 1997, Spatial Data Mining: A Database Approach, *Fifth Symposium on Large Spatial Database*.
- FGDC Technical Report, 2000, *Contents standard for digital geospatial metadata Workbook*, Federal Geographic Data Committee, FGDC (FGDC-STD-001-1998), pp.32-98, URL:<http://www.fgdc.gov>.
- Gahegan, M., Yuan, M., and Miller, H., 2003, Geospatial Data Mining and Knowledge Discovery, *A UCGIS White Paper on Emergent Research Themes*, pp.1-17.
- Gobel, S., and Jasnoch, U., 2001, Visualization techniques in metadata information systems for geographical data, *Fraunhofer Institute for Computer Graphics, Advanced in Environmental Reserach*,

- pp.415-424. URL:<http://www.elsevier.com/locate/aer>.
- Government of Alberta, 2003, "Metadata Resources Guide", Information Management Branch in Alberta Government Services, URL:<http://www.im.gov.ab.ca>.
- Guo, D, 2002, Spatial Cluster Ordering and Encoding for High-Dimensional Geographic Knowledge Discovery, *GeoVISTA Center and Department of Geography, Pennsylvania State University*, pp.5-13.
- Gupta, K.A., and Sy, B.K., 2001, "Information-statistical approach for spatial-temporal data with application", *Engineering Applications of Artificial Intelligence*, Vol.15, pp.177-191, URL:<http://www.elsevier.com/locate/engappai>.
- Han, J., Koperski K., and Stefanovic, N., 2000, GeoMiner: A System Prototype for Spatial Data Mining, *GeoMiner Reserach Group*, pp. 6-11.
- Han, J., Kamber, M., and Tung, A.K.H, 2002, "Spatial Clustering Methods in Data Mining: A Survey", URL:<http://citeseer.ist.psu.edu/436843.html>.
- Han, J., 2003, "Data Mining in Spatial Database: A Multi-Disciplinary Promise", Database Systems Research Lab, URL:<http://www.cs.uiuc.edu/~hanj>.
- Han, J., 1997, "Spatial Data Mining and Spatial Data Warehousing", Symposium on Spatial Databases(SSD97) Conference Tutorial.
- Harms, S., Li, D., and Tadesse, T., "Efficient Rule Discovery in a Geo-Spatial Decision Support System", NSF Digital Government Grant No. EIA-0091530.
- Huang, Z., 1998, Extension to the K-means Algorithm for Clustering Large Data Sets with Categorical Values, *Water Resources Management, Kluwer Academic Publishers*, pp.283-304.

- Inmon, W.H., and Welch, J.D., 1997, "Managing the Data Warehouse", John Wiley & Sons, Inc.
- Kamber, M., and Han, J., 2001, "Data Mining: Concepts and Techniques", Academic Press.
- Kafatos, M., 1999, Data Access, Querying, Analysis, and Data Mining in a Distributed Framework for Earth System Science support, *Center for Earth Observing and Space Research, Institute for Computational Science and Informatics*, p.17.
- Kimball, R., 2000, "The Data Warehouse Toolkit: Building the Web-enabled Data Warehouse", John Wiley & Sons, Inc.
- Koperski, K., and Han, J., 2001, "Discovery of Spatial Association Rules in Geographic Information Databases", The White Paper of Natural Sciences and Engineering Research Council in Canada.
- Kuba, P., 2001, "Data Structure for Spatial Data Mining", FIMU Report Series, FIMU-RS-2001-05, pp.7-19, URL:<ftp://ftp.fi.muni.cz>.
- Miller, H. J. and Han, J., 1999, "Discovering Geographic Knowledge in Data Rich Environments", *Reporting of a Special Meeting held under the auspices of the Varenus Project, NCGIA*, Vol. 1 (2) pp.10-42, <http://www.ncgia.ucsb.edu/varenus/discovering/description.html>.
- Miller, H.J and Han, J., 2001, Geographic Data Mining and Knowledge Discovery: An overview, In Miller, H.J. (eds) *Geographic Representation in Spatial Analysis, Journal of Geographical System*, pp.4-17.
- Noon, C., E., and Hankins, C., T., 2001, "Spatial Data Visualization in Healthcare: Supporting a Facility Location Decision via GIS-based Market Analysis", *In Proceedings of the 34th Hawaii Interna*

- tional Conference on System Sciences, IEEE 0-7695-0981-9/01.*
- Oracle, 1999, "Oracle 8i for Data Warehousing: Fast and Simple for More Data and More Users", An Oracle Technical White Paper, URL:<http://www.oracle.com>.
- Oracle, 2001, "Oracle9i Data Mining", An Oracle Technical White Paper, URL:<http://www.oracle.com>.
- Oracle, 2003, "Slowly Changing Dimension with Oracle 9i Warehouse Builder", An Oracle White Paper, [http://www.jusungyang.com/DWf\\_older/DataWarehousing\\_Home.html](http://www.jusungyang.com/DWf_older/DataWarehousing_Home.html).
- Papadias, D., Kalnis, P., and Zhang, J., 2001, "Efficient OLAP Operations in Spatial Data Warehouses", Technical Report HKUST-CS01-01 in Hongkong University of Science and Technology, URL: <http://www.cs.ust.hk/~dimitris>.
- Paul, L., and Clarke, G., 1995, "GIS for Business and Service Planning", GeoInformation International.
- Peterson, K., 1998, Development of Spatial Decision Support System for Residential Real Estate, *Journal of Housing research* Vol. 9 (1), pp.135-155.
- Phillips, A., Williamson, I., and Ezigbalike, C., 2002, Spatial Data Infrastructure Concepts, *Contents Page in the Australian Surveyor*, Vol. 44(1), pp.20-28. URL:<http://www.geom.unimelb.edu.au/research/publications/IPW/SDIDefinitionsAusSurv.html>.
- Ravan, S., 2002, Spatial Decision Support System for Biodiversity Conservation, URL:<http://www.cdacindia.com/html/pdf/geom1.pdf>.
- Rawlings, J., and Kucera, H., 1997, Trials and tribulations of implementing a spatial data warehouse, *In Proceedings of the 11th Annual Symposium on Geographical Information System*, Vancouver, pp.510-513.

- Roddick, J.F., Hornsby, K., Spiliopoulou, 2003, Paradigms for Spatial and Spatio-temporal Data Mining, *Department of Resource Management and Environmental Science, Australian National University*, URL:<http://citeseer.nj.nec.com/update/444197>.
- Sadahiro, Y., 2003, "Cluster detection in uncertain point distributions: a comparison of four methods", *Environment and Urban System* in Elsevier Science Ltd, Vol.27, pp.33-52.
- Shekhar, S., and Chawla, R., and Vatsavai, R., 2001, "Map cube: A Visualization tool for Spatial Data Warehouses", John Wiley & Sons, Inc.
- Stefanovic, N., and Koperski, K., 2000, Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes, *in the Proceedings of the IEEE Transactions on Knowledge and Data Engineering*, Vol. 12(6), pp.1-19.
- Stefanovic, N., Han, J., and Koperski, K., 2000, "An Efficient Two-Step Method for Classification of Spatial Data", URL:<http://http://citeseer.nj.nec.com/context/873591/183046>.
- Sui, Z.D., and Hugill, J.P., 2002, A GIS-based Spatial Analysis on Neighborhood Effects and Voter turn-out: a case study in college station, *Political Geography 21*, Vol.21, pp.160-173, URL:<http://www.politicalgeography.com>.
- Sun Microsystems, 2001, "Distributed Programming with Java Technology", Sun Microsystems Inc.
- Turner, A., 2000, Density Data Generation for Spatial Data Mining Applications, *Center for Computational Geography, University of Leeds*, pp.27-30, <http://www.geog.leeds.ac.uk/papers/21stC.html>
- Tung, A., Hou, J., and Han, J., 2001, Spatial Clustering in the Presence of Obstacles, *In the 17th International Conference on Data Engi*

- neering*, pp.8-20.
- Tung, A., 2002, "Mining Spatial Datasets: A New Frontier for Data Mining", Technical Report in IBM Canada and Intelligent Database Systems Research Lab.
- Wang, W., Yang, J., and Muntz, R., 2000, An Approach to Active Spatial Data Mining Based on Statistical Information, *IEEE Transactions on knowledge and data engineering*, vol. 12(5) pp.715-725.
- Wachowicz, M., 2003, Uncovering Spatio-Temporal Patterns in Environmental Data, *Water Resources Management, Kluwer Academic Publishers*, pp.469-487.
- Wachowicz, M., 1998, "Integrating GVis, GIS and KDD for Exploring Spatio-Temporal Data", PLECOM Project, URL:<http://wwlmu.jrc.it/Workshops/5ec-gis/presentations/tuesday/wachowicz.ppt>.
- Witkowski, M.S., Paul, M.R., Riggs, T.L., 2002, "Spatial Data Warehouse Design", in Environmental Dynamics and Spatial Analysis Group, Los Alamos National Laboratory.
- Wong, D.W.S., and Lee, J., 2001, "Statistical Analysis with ArcView GIS", John Wiley & Sons, Inc.
- Yuan, M., 2003, "Geospatial Data Mining and Knowledge Discovery", *A Proposed Emergent Research Topic in GIScience*.
- Zhou, X., Truffet, and Han, J., 2001, Efficient Polygon Amalgamation Methods for Spatial OLAP and Spatial Data Mining, *6th International Symposium on Spatial Databases(SSD01), LNCS Vol.1651*, pp.167-187.
- , 1999, "Cyber-Spatial Analysis: Modelling Web Information Flows", URL:<http://http://www.geovista.psu.edu/sites/geocomp99/Gc99/098/abs99-098.htm>.

## Abstract

# A Design and Practical Use of Spatial Data Warehouse for Spatial Decision Making

Park, Ji-Man

Department of Geography, Kyunghee University

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.

The construction of spatial data warehouse which involves data cleaning and integration, can be viewed as an important preprocessing step for geographic knowledge discovery. Moreover, the pilot tested system of this research provides on-line analytical processing(OLAP)

tools for interactive analysis of multi-dimensional data of geographically various granularities, which facilitate effective spatial data mining.

Especially, the importance of this study is in the use of snowflake schema model capabilities for a GIS framework. This model enables us to maintain and save storage space easily, because a spatial dimension table can become enormous when the performance may be adversely impacted. With the rapid progress of web technology and increasing popularity of e-business, it is expected that there will be increasing requirements for integrating spatial data warehouses and spatial data mining with the web technology. It is likely that the future generations of spatial data warehouses and spatial data mining will be web-based. XML technology will be used generally for future spatial data and information exchange.

Therefore, it is important to investigate how to develop web-based spatial data warehouses and geographic knowledge discovery for the future generations of spatial information systems. And this case study will be useful for marketers who want to find their target marketing area in automatic way.

**Key Words : Spatial Data Warehouse, Multi-dimensional Analysis, GIS, Snowflake Schema Model, OLAP, Internet Consulting System, Targeted Marketing.**