

## 비즈니스 GIS에서 공간 데이터마이닝(Spatial Data Mining)기법을 이용한 상권추출

이병길\*

### Defining of Trade Area using Spatial Data Mining Technique in Business GIS

Byoung-Kil Lee\*

#### 요 약

최근 마케팅이나 기업전략 수립분야에서 비즈니스 GIS를 적용하기 위한 다양한 응용시스템이 개발되고 있다. 이 중 대부분이 의사결정지원을 위한 정보로서 지리공간 상에서의 상권이라고 하는 특정 범위에 대한 통계정보의 산출을 요구한다. 기존에 상권범위의 정의는 대부분 개별점포에 대해 경험치를 적용하여 이루어져 왔으며, 특정 업종이나 소비자 상권과 같은 일반적인 의미의 상권을 정의하는 객관적인 방법이 제시되지 않았다. 본 연구에서는 비즈니스 GIS 분야에 축적된 점사상 정보에 공간 데이터마이닝 기법을 적용하여 상권의 범위를 추출할 때, 기법 간의 장단점을 비교분석하고, 상권추출의 타당성을 검증하고자 하였다. 본 연구의 결과 점사상으로 사용된 카드사 가맹점 데이터와 소매체인 회원의 정보를 이용하여 상권의 추출이 가능하며, DENCLUE(DENsity-based CLUstEring) 기법이 적합한 공간 데이터마이닝 기법임을 알 수 있었다.

**주요어** : 비즈니스 GIS, 공간 데이터마이닝, DENCLUE, 상권추출

**ABSTRACT** : Lots of application systems are developed for applying business GIS in marketing or strategic planning of the company, recently. Almost of the systems require statistics for some areas(trade areas or sales areas) as the important information of decision support. As far as now, trade areas are defined for individual stores using know-how of the

---

\* 서울대학교 공학연구소 특별연구원, 주식회사 타스텍 정보기술연구소장

specialists, but there is no well-defined method for defining of trade areas of the specific business domains or trade areas of the customers. In this study, we have applied the spatial data mining methods to the point features in GIS, evaluated the results of each methods, and discussed the feasibility of defining of trade areas. From the results of this study, we have concluded that the defining of trade areas from point features, such as franchisees of credit card company or memberships of retail chain store, and that the DENCLUE(DENSity-based CLUstEring) method is the best suitable spatial data mining algorithm for this purpose.

Keywords : Business GIS, Spatial Data Mining, DENCLUE, Defining of Trade Area

## 1. 서 론

최근의 지리정보의 획득 및 응용은, 지리정보를 활용하는 분야가 많아지고 정보의 활용이 고도화 되어가는 것과 함께, 정보획득을 위한 수단이 다양해지고 이미 생산된 정보에 대한 접근성이 향상되는 것이 상호 작용을 일으키면서 선순환하고 있는 것으로 보여 진다. 즉, 지리정보의 활용이 증가하면서 모든 분야에서 점차 많은 양의 정보가 축적되어가고 있으며, 축적된 정보가 다양해질수록 더 많은 응용분야에서 많은 활용이 일어나고 있다는 것이다.

특히 비즈니스 GIS와 같이 최근에 도입이 활발히 이루어지고 있으면서, 기본도 정보 외의 다수의 동적인 정보를 취급해야하는 분야에서는 다양한 내용의 점사상 정보가 폭발적으로 증가하면서 이를 이용한 다양한 분석 알고리즘이 연구되고, 새로운 응용분야가 생겨나고 있다.

초기 비즈니스 GIS의 활용이 주로 자사 점포의 위치와 기본도 정보 정도를 활용

한 기본적 지역정보의 획득에 머물렀다면, 최근의 비즈니스 GIS의 활용은 특화된 지역마케팅, 점포전략지원 등을 위한 전반적인 상권현황의 파악, 고객의 행동 분석 등 보다 심화된 정보분석차원에 까지 이르고 있다. 이와 같은 심화된 정보 분석을 위해서는 방대한 양의 동적정보(고객정보, 매출정보, 업종별 POI 등)가 필요하며, 이미 상당수의 업체에서는 이러한 정보를 축적하고 있다. 대부분의 경우 이러한 동적정보는 주소 정보를 가지고 있기 때문에, 주소와 지번도의 매칭을 통하여 좌표를 부여할 수 있으며, 부여된 좌표를 점사상으로 변환하여 GIS에서 활용하게 된다.

기존 비즈니스 GIS에서 상권은 주로 특정 점포를 중심으로 점포의 세력이 미치는 범위, 즉 점세권을 말하는 용어로 쓰여 왔으며, 업종에 따라 경험이 많은 현업 실무자들의 의견에 따라 그 범위를 규정해 왔다. 예를 들어, 은행의 경우 각 지점으로부터 도보 5분거리(약 300m) 이내를 1차상권으로, 도보 10분거리(약 500m) 이내를 2차상권으로 정의하고 있으나, 실

제 고객의 위치를 이용하여 상권을 나누는 CST(Customer Spotting) 기법<sup>1)</sup>등을 이용한 과학적이고 실증적인 연구가 이루어지지 않는 것이다.

개별점포를 평가하고 지원하기 위해 활용되던 기존의 응용분야에서 벗어나, 최근의 비즈니스 GIS에는 사업 전체적인 관점에서 신규사업의 기회를 찾거나 점포 네트워크를 최적화하거나 새로운 표적시장을 발굴하거나 하는 등의 응용분야[정보통신부, 2003]가 등장하고 있어, 이들 응용분야의 의사결정을 지원할 수 있는 정보제공이 요청되고 있다. 이상과 같은 응용시스템에서 산출되어야 할 정보를 제공하기 위해서는 상권현황의 파악, 고객 집단의 행동분석과 같은 속성데이터의 마이닝 외에도, 정보를 훼손하지 않고 집계하고 분석할 수 있는 기본단위인 상권<sup>2)</sup>의 설정이 필요하다.

본 연구에서는 다양하게 축적된 점사상 정보로부터 상권을 추출하고 설정하는 객관적인 방법을 제시하고, 그 성능을 평가하고자 하였다. 이를 위해 A카드사 가맹점의 정보와 B소매체인의 회원 정보를 가공하여 얻어진 점사상을 기본정보로 활용하였으며, 이 점사상을 그 분포형태에 따라 클러스터링하여 상권을 분할하고 추출하는 작업을 수행하였다.

점사상을 클러스터링하는 방법으로는 최근 공간 데이터베이스 분야에서 많은 연구가 이루어지고 있는 공간 데이터마이닝 기법을 이용하였다. 공간 데이터마이닝은 공간 데이터에 암시적으로 존재하는 기하학적 또는 통계적 특성과 상호관계를 규명하는 작업[Han, 2001]으로, 공간 데이터를 클러스터링하는 기법에 대한 많은 연구가 수행되고 있다. 공간 데이터마이닝에서 클러스터란 공간 객체가 명시적으로 군집화 된 것을 의미한다.

본 연구에서는 기존의 공간 데이터마이닝 기법에 필요한 세가지 요구사항<sup>3)</sup> [Han, 2001] 외에, 비즈니스 GIS의 다양한 동적정보를 유연하게 처리할 수 있는지 여부를 중요한 평가요소로 고려하였다.

본 논문에서는 여러 가지 공간 클러스터링 기법 중 k-medoid기법과 비교적 최근에 연구되어 GIS 공간 데이터마이닝에 활발히 적용되고 있는 SMTIN 및 DENCLUE기법을 적용하여 상권추출을 실험하고 그 결과를 비교평가하였다.

본 논문은 다음과 같이 구성되어 있다.

1장에서는 연구내용의 개략을 서술하였으며, 2장에서는 비즈니스 GIS와 공간 데이터마이닝의 기존 연구동향을 개관하였다. 3장에서는 본 연구에 사용된 데이터의 특성과 클러스터링 결과를 제시하고,

1) 지도상에 점포를 중심으로 점포를 이용한 고객의 위치를 표시하여 전체고객의 60%-80%가 포함되는 지역을 1차상권, 90%가 포함되는 지역을 2차상권으로 정의하는 방법[안광호, 1998]  
 2) 여기서의 상권은 앞서의 점세권과는 달리 해당 지역에 시장이 형성되어 있는지 여부를 판단하는 것으로 어떤 업종의 상권이 형성된 지역이 다른 업종에 대해서는 그렇지 않을 수도 있고, 특정 점포가 상권이 형성되지 않은 지역에 위치해 있을 수도 있으며, 상권이 형성된 지역에 점포가 없을 수도 있다. 즉, 현재 소비가 일어나고 있거나, 잠재적으로 소비가 일어날 수 있는 지역을 상권으로 정의한다.  
 3) 첫째, 다루어야 할 공간 데이터의 양이 방대하기 때문에 마이닝 알고리즘의 실행 속도가 빨라야 한다. 둘째, 클러스터에 포함되는 공간객체, 클러스터의 형상과 분포 등에 대한 풍부한 정보제공이 가능해야 한다. 셋째, outlier에 민감하지 않으면서 데이터가 추가되었을 때 기존 outlier도 데이터로 사용 가능해야 한다.

결과를 비교, 평가하였으며, 마지막으로 4장에 간략한 결론과 향후 연구발전 방향에 대해 논의하였다.

## 2. 연구동향

### 2.1 비즈니스 GIS

1950년대에서 1960년대에 이르는 계량지리학의 혁명기에 여러 응용지리학의 한 분야로서 경영지리학(비즈니스 geographics)이 등장하게 되었고, 이 시기에 지금의 비즈니스 GIS에도 널리 사용되는 매출예측의 방법인 유추법(analogy method)이 Applebaum에 의해 정립되었으며[안광호, 1998], 소매점 입지분석을 위한 중력모형인 Huff 모형[Huff, 1963]이 창안되었다. 이러한 이론들은 상당한 양의 수작업 계산을 위한 인력 및 시간을 요구하였기 때문에 실제 현장에 적용한 사례는 극히 제한적이었다[Daniel, 1993].

여러 응용분야에서의 이론을 현실화하기 위해 필요한 많은 양의 수작업이 GIS의 다양한 분석 및 연산기능에 의해 대체되어 실제 활용도를 높이고 있는 것과 동일하게 경영지리학 분야에도 비즈니스 GIS가 도입되며 다양한 이론들이 적용되고 실험되었으며, 현재에도 많은 연구가 이루어지고 있다.

비즈니스 GIS는 전통적으로 지방자치단체, 시설물관리 등에 적용된 정통 GIS라고 하는 응용분야와는 상당히 많은 다른 점을 보인다. 비즈니스 GIS가 가지는 세가지 중요한 특성은 다음과 같다[Daniel, 1993].

첫째, 응용시스템 중심. GIS가 가지는 2000개가 넘는 다양한 기능을 실제 업무 처리의 흐름과 기능에 맞도록 커스터마이징된 응용프로그램으로 대체한다. 비즈니스 GIS에서는 불명확한 모든 목적을 지원하기 위해 모든 잠재 기능을 제공하지 않고, 몇가지 목적에 초점을 맞춘 소수의 특화된 응용프로그램을 제공하여야 한다.

둘째, 데이터 소비. 일반적으로 비즈니스 GIS의 사용자들은 데이터의 생산자가 아니다. 통상의 정통 GIS 사용자에게는 데이터를 생성하여 입력할 책임과 의무가 부여되는 경우가 많다. 반면에 비즈니스 GIS의 사용자는 데이터를 입력하기보다는 기성품(시중에서 구할 수 있는 수치지도 등)을 구매하여 사용하는 경우가 많다.

셋째, 동적 데이터 중심. 정통 GIS의 주요 데이터는 수치지도인 경우가 대부분이다. 반면에 비즈니스 GIS에서 수치지도는 단순히 배경으로서 또는 사용자에게 주변 정황에 대한 정보를 제공하는 목적으로서만 사용되고, 실제 관심의 대상이 되는 데이터는 고객, 판매실적, 시장조사, 인구통계와 같은 동적인 데이터인 경우가 대부분이다.

정통 GIS와 비교한 비즈니스 GIS의 특성 중, 특히 세번째의 특성은 축척에 독립적인 속성정보이며, 고정된 좌표를 가지지 않으며, 각 데이터가 가지는 차원, 척도 및 공간상의 범위가 상이하다는 등의 동적 데이터 자체가 가지는 독특한 특성이 있기 때문에, 실제 적용에 많은 어려움이 발생하게 된다. 일례로 고객은 주소로 통해 개별 지번상에 위치하게 되고, 판매실적은 지사 또는 지점의 상권 단위

(관리영역)로 집계되며, 인구통계는 행정동 단위로 집계되기 때문에 이들을 통합하여 분석하기에는 많은 가정과 추정이 필요하게 된다. 이러한 현상은 특히 우리나라와 같이 통계구역<sup>4)</sup>이 대규모인 경우에 두드러지게 된다.

현재 국내에서 비즈니스 GIS를 도입한 많은 사례에서 나타나고 있는 현상으로, 일정 점포나 일정 지역의 상권에 대한 특성 또는 정보를 추출하는 상권분석에서 상권에 대한 특성과 정보는 거의 행정동 단위 통계를 기반으로 추정하여 추출하고 있다. 그런데, 우리나라의 상권은 역세권을 중심으로 형성되는 경우가 많기 때문에 일정 지역의 상권과 행정구역의 경계가 일치하지 않는 경우가 대부분이며, 상권의 성장이 주변 교통 및 배후지역 형성과 연관되기 때문에 상권의 형상은 매우 불규칙하다. 예를 들어, 흔히 강남역 상권이라고 일컬어지는 강남역 인근 지역은 서초구와 강남구의 2개구에 속하는 서초 2,4동, 역삼1동 등의 3개 동에 걸친 상권을 형성하고 있으며, 확대된 강남역 상권은 북으로는 강남대로를 따라 신사역까지, 동으로는 테헤란로를 따라 역삼역까지 “L”자 형태의 광역상권을 형성하고 있다.

이와 같은 일반적 의미의 상권을 정의하는 기존의 방법은, 전문가가 경험에 의해 범위와 형상을 결정하기 때문에, 일관된 성과를 얻을 수 없다는 문제점이 있으

며, 정의된 상권에 대한 정보도 표본에 의한 현장조사에 의존해 왔다. 상권의 속성, 형성, 발전과정 등을 연구하기 위해서는 지역이나 시간에 따른 차이가 없는 객관성 있고 재현성 있는 상권추출 방법이 필요하다. 최근 비즈니스 GIS분야에서 축적된 다양한 동적 정보(매출을 포함한 점포정보, 고객정보 등)를 활용한 다양한 분석이 제안되고 있으며, 본 연구에서는 경험에 의한 가정에 의존하지 않고 실제로 축적된 데이터와 GIS의 분석기법을 활용하여 공간상에서 일어나는 경제 현상을 객관적으로 설명할 수 있는 상권추출 기법을 제안하였다.

## 2.2 공간 데이터마이닝

공간 데이터마이닝을 위한 클러스터링 기법은 크게 분할기법(Partitioning Method), 계층적기법(Hierarchical Method), 밀도기반기법(Density Based Method), 격자기반기법(Grid Based Method)으로 분류할 수 있으며, 각각에 대해 여러 클러스터링 기법에 제약조건을 추가하는 분석기법에 대한 연구가 이루어지고 있다.

### 가. 분할기법(Partitioning methods)

분할기법은 데이터마이닝의 등장 이전부터 널리 사용되어온 클러스터링 기법이다. n개의 객체를 가진 d차원 공간의 집

4) 통계구역이란 통계청에서 센서스 결과를 집계하여 공표하는 지역의 최소 단위를 말한다. 도시지역에 대해, 미국에서는 3,000명~5,000명의 상주인구가 포함되도록 통계구역을 설계하고, 영국에서도 5,000명 내외로 설계된다. 우리나라의 통계구역인 행정동에는 약 20,000명 내외의 인구가 상주(도시지역의 경우)하여 통계구역의 크기도 상대적으로 크며, 그 범위도 최소 100여명(서울시 성북구 월곡4동)에서 최대10만명(부산시 해운대구 좌동)으로 직접적인 통계수치의 비교에 어려움이 있다.

합  $D$ 와 입력 매개변수  $k$ 가 주어졌을 때, 분할 알고리즘은 클러스터 중심이나 클러스터 분포로부터 각 객체의 편차가 최소가 되도록 모든 객체를  $k$ 개의 클러스터로 편제하는 것이다. 편차는 유사도 함수(similarity function)를 이용하여 계산되며 각각의 알고리즘에 따라 측정방법은 상이하다.

일반적으로 많이 사용되는 분할기법 알고리즘으로는  $k$ -평균( $k$ -means) 알고리즘, 기대값 극대화(Expectation Maximization) 알고리즘과  $k$ -medoid 알고리즘 등이 있다. 분할기법에 속하는 알고리즘 들은 전체 공간을 최적으로 분할하는  $k$ 개의 중심 또는 분포를 알아내는 것이 중요하다. 최적의 중심  $k$ 개 또는 분포를 찾게 되면,  $n$ 개의 객체는  $k$ 개의 클러스터에 자동적으로 할당된다. 그러나, 중심  $k$ 개 또는 분포의 전역 최적화(global optimization)는 상당히 어렵기 때문에 분할기법에서는 반복적 위치변경기법(iterative relocation technique)을 사용하여 국지적 최적화(local optimization)를 시도한다.

#### 나. 계층적기법(Hierarchical methods)

계층적기법은 주어진 객체 집합을 계층적으로 분해하여, 데이터베이스를 재귀적으로 작은 부분집합으로 분할하는 나무인, 덴드로그램(dendrogram)을 만들어 낸다. 덴드로그램은 상향식과 하향식의 두 가지로 나뉘는데, 상향식은 “응집화(agglomerative)” 접근법이라고 하여, 모든 객체가 하나의 그룹이라고 가정하고 출발한다. 여기서는 두 그룹 중심 간의

거리와 같은 측정치에 따라 객체나 그룹을 지속적으로 통합시키는 과정을 모든 그룹이 하나의 그룹으로 통합되거나 정해진 종료 조건에 도달할 때까지 반복한다. 하향식은 “분열화(divisive)” 접근법이라고 하여, 모든 객체를 하나의 클러스터로 간주하고 시작한다. 각 반복 단계에서, 각 객체가 하나의 클러스터가 되거나 종료 조건에 도달할 때까지 클러스터를 측정값에 따라 분리하는 과정을 거친다.

초기의 계층적기법에는 AGNES(AGglomerative NESTing)과 DIANA(Divisia ANALYSIS)가 있었으나 클러스터의 분할이나 통합에 사용되는 측정치를 과도하게 단순화하였고, 수행된 단계를 되돌려 다시 할 수 없다는 문제가 있어 클러스터링 결과에 착오가 발생하는 경우가 많았다.

계층적기법의 효율성을 강화하기 위한 최근의 알고리즘으로는 CURE(Clustering Using REpresentatives) 또는 CHAMELEON 과 같이 분할과 통합에 사용되는 방법을 개선한 것이 있고 BIRCH(Balanced Iterative Reducing and Clustering using Hierarchies)와 같이 계층적 응집화 알고리즘에서 산출된 초기 결과를 반복적 재배치를 통해 클러스터를 정제하는 것이 있다.

#### 다. 밀도기반기법(Density-based methods)

대부분의 분할기법에서는 객체간의 거리를 기준으로 객체를 클러스터링한다. 이러한 기법은 타원형 클러스터만을 찾아낼 수 있으며, 임의의 형상을 가진 클러스터를 찾아내기는 어렵다. 밀도의 관점에 기초하여 만들어진 알고리즘에서는 데이터 공간에서 객체가 조밀하게 밀집된 지

역을 클러스터로 간주하여 낮은 밀도(잡음을 나타냄)의 지역으로부터 분리해 낸다. 밀도기반기법은 잡음(또는 outlier)를 제거하거나, 임의의 형상의 클러스터를 찾아내는 데 사용된다.

가장 처음 시도된 밀도기반 알고리즘은 DBSCAN(Density Based Spatial Clustering of Applications with Noise)으로 객체로부터 거리 이내에 존재하는 데이터 점이 기준인 MinPts보다 크다면 객체 주변의 밀도가 충분히 조밀하다고 판단하는 것이다. 찾아지는 클러스터는 매개변수인  $\epsilon$  과 MinPts에 따라 달라지므로 DBSCAN의 성과는 사용자의 매개변수 선택결과에 의존하게 된다. 이러한 문제를 극복하기 위해 OPTIC(Ordering Points to Identify the Clustering Structure) 알고리즘이 제안되었다. 여기서는 클러스터 집합을 고정된 매개변수를 이용하여 명시적으로 만들어내지 않고, 후처리가 가능한 클러스터 배열(cluster ordering)을 만들어 낸다.

DBSCAN이나 OPTIC은 효율적 처리를 위한 공간 인덱스가 필요하기 때문에 다차원 데이터를 효과적으로 처리하기 어렵다. 다차원 데이터를 효율적으로 처리하기 위한 DENCLUE(DENsity-based CLUstEring)는 데이터 점의 영향력 함수의 합으로 객체 주변의 전체적인 밀도를 해석적으로 구해낸다. 영향력 함수의 합을 효율적으로 계산하기 위해서 격자형 구조를 이용한다. 실험에 의하면 DENCLUE는 DBSCAN에 비해 45배정도 빠른 처리 속도를 보이지만 DBSCAN과 유사하게 격자의 크기, 영향력 함수의 선택 등에 의해 클러스터링 결과는 크게 달라지게 된다.

## 라. 기타

최근에는 이상의 알고리즘 외에도 공간상의 제약조건(도로, 철도나 하천과 같은)을 추가적으로 거리의 계산이나 클러스터의 분할, 통합에 고려하는 보다 정교한 알고리즘과 함께, SMTIN(Spatial data Mining method by Delaunay Triangulation)이라는 점사상의 기하학적 분포를 고려한 클러스터링 기법이 제안되었다. SMTIN은 분할기법이나 계층적기법의 약점인 복잡한 형상의 분포 패턴을 안정적으로 찾아내기 위해 제안된 알고리즘이다. 점사상으로부터 TIN을 구성하고 TIN의 변 가운데 일정 임계치  $T$ 를 벗어나는 변을 삭제하고 남은 폴리곤을 이용하여 convex hull을 추출하는 과정을 거쳐 임의의 형상의 클러스터를 분리해 낼 수 있다. 밀도기반기법과 유사하게  $T$  값이 작아질수록 핵심지역만이 남게 되므로  $T$  값을 조정하여 클러스터 간의 다층적 구조(nested structure)를 해석할 수 있다는 점 외에도, 데이터에 대한 사전지식 없이도 분석이 가능하다는 장점이 있다[Kang, 1997].

## 3. 사용데이터 및 자료처리

### 3.1 대상 데이터 및 사용 소프트웨어

본 논문에서는 데이터 획득의 가능성, 대상지역 내의 상권분포 등을 고려하여, 실험연구의 연구대상지역으로 서초구를 선정하였다. 서초구는 강남고속버스터미널, 방배동 카페골목, 교대역에서 강남역

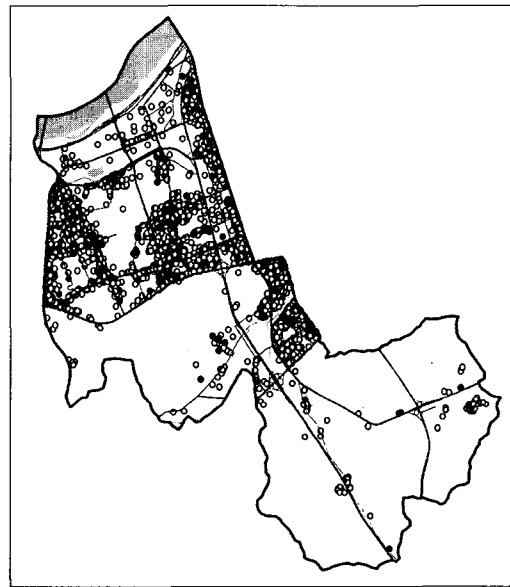
에 이르는 역세권 상권, 강남대로를 중심으로한 사무실 지역 등 경제활동의 다양한 변수들이 존재하며, 연구를 위한 데이터 역시 가장 충실하게 얻을 수 있는 지역이었다.

본 연구에서 상권추출을 위해 사용된 데이터는 서초구에 위치하는 A카드사 가맹점 정보와 B소매체인 회원 정보이며, 데이터의 처리에는 통계 패키지인 S-plus와 범용 GIS 소프트웨어인 ArcGIS8.3과 ArcInfo8.1.2를 이용하였고, 소프트웨어에서 지원되지 않는 기능은 Visual Basic을 이용하여 프로그래밍하였다.

사용된 데이터의 제원은 <표 1>과 같다.

카드사 가맹점의 경우 업종별로 카드사용율에 대한 차이와 카드사에 따라 가맹점의 차이는 존재하지만 대체적으로 선매품<sup>5)</sup>이나 고가 소비재에 대한 소비가 일어나는 지역을 상권으로 추출하는 데 대표성이 있다고 할 수 있으며, 소매체인 회원의 정보를 이용하면 소비자의 위치로부터 근거리에 형성되는 일용품<sup>6)</sup>에 대한 상권을 추출할 수 있을 것으로 예상하였다.

획득된 데이터의 주소는 상당한 오차를 포함하고 있어 이를 수정, 정제하였으며, 정제된 주소를 지번 데이터와 매칭하여 좌표를 부여하였다. 좌표가 부여된 데이터를 ArcGIS 8.3으로 입력하고 점사상 shape 파일로 변환([그림 1], [그림 2] 참조)하여 상권추출을 위한 이후 분석작업을 수행하였다.



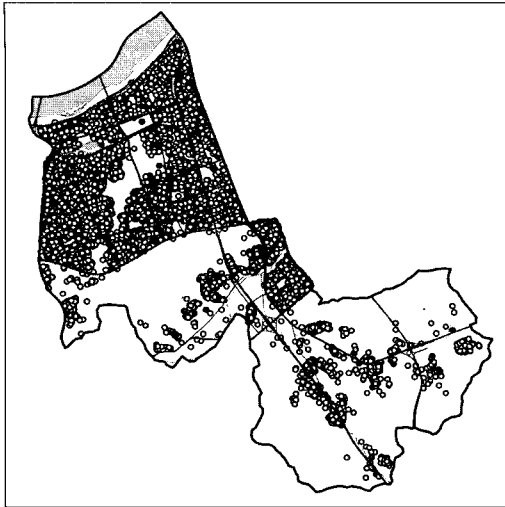
[그림 1] A카드 가맹점 분포도

<표 1> 사용 데이터 제원

종 류	수 량	속성 항목
A카드사 가맹점포	5,923업소	주소, 업종, 가입연월, 2000년 월평균 매출
B소매체인 회원	79,212명	주소, 가입연월

- 5) 선매품이란 소비자가 여러 제품을 비교, 평가하여 구매하고자 하는 상품으로 내구재로서 가격이 비싼 가전제품, 가구, 의류 등이 이 범주에 포함된다. 이러한 물품의 구매 시에는 소비자가 원하는 물건을 취급하는 상점이 밀집한 지역에서 여러 가지 조건을 비교, 평가하고 구매하는 행동양식을 보이는 경우가 많다.
- 6) 일용품이란 소비자가 일상 생활에서 반복 재구매를 통하여 지속적으로 소비하는 상품으로 일반적인 구멍가게나 슈퍼에서 취급되는 상품이 이 범주에 속한다. 이러한 물품의 구매시에 소비자는 편의를 고려하여 현재 위치로부터 가까운 점포에서 구매하는 행동양식을 보이는 경우가 많다.





[그림 2] B소매체인 회원 분포도

### 3.2 공간 클러스터링을 이용한 상권추출

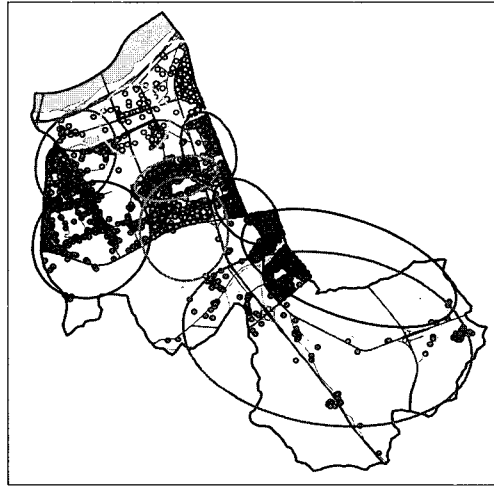
#### 가. k-medoid기법에 의한 상권추출

[그림 3]은 k-medoid기법에 의한 상권추출 결과이다. A카드 가맹점의 좌표를 S-plus에 입력하고 S-plus의 “partitioning around medoids” 메뉴를 이용하여 전체 가맹점을 10개의 클러스터로 분류하였다. 분류시 일치도의 측정값으로는 유클리드 거리를 이용하였다. 10개의 클러스터는 10개의 상권을 나타낸다고 할 수 있다.

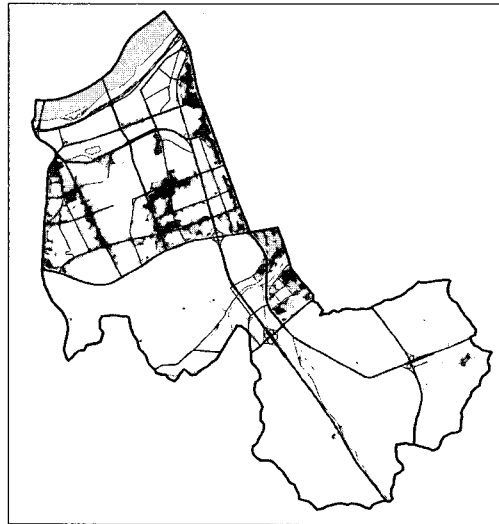
#### 나. SMTIN기법에 의한 상권추출

[그림 4]는 SMTIN기법에 의한 상권추출 결과이다. 먼저, 가맹점의 데이터로부터 매출액이 높고 값인 TIN을 작성하고, TIN을 line 커버리지(coverage)로 변환하였다. 변환된 결과로부터 임계치 T로 200m, 100m, 75m, 50m를 적용하여 T보다 긴 line

을 제거한 커버리지를 4개 작성하고 최종적으로 4개의 커버리지를 하나의 커버리지로 통합하였다. 가장 큰 임계치를 200m로 설정한 것은 임의 지점에서 도보에 의해 영향을 주는 1차적 한계거리를 200m로 가정하였기 때문이다.



[그림 3] k-medoid기법에 의한 가맹점 상권추출 (10개 상권으로 분류)



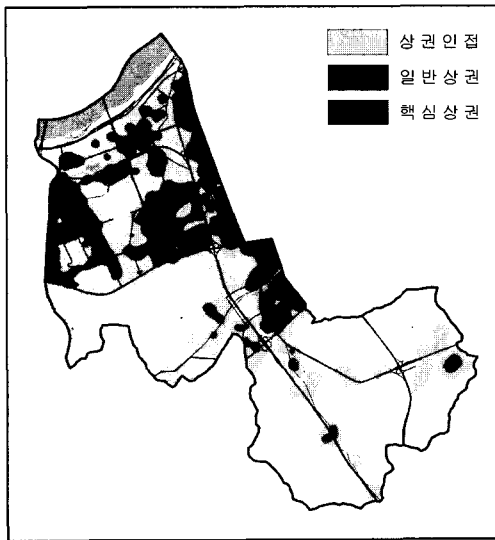
[그림 4] SMTIN기법에 의한 가맹점 상권추출

**다. DENCLUE기법에 의한 상권추출**

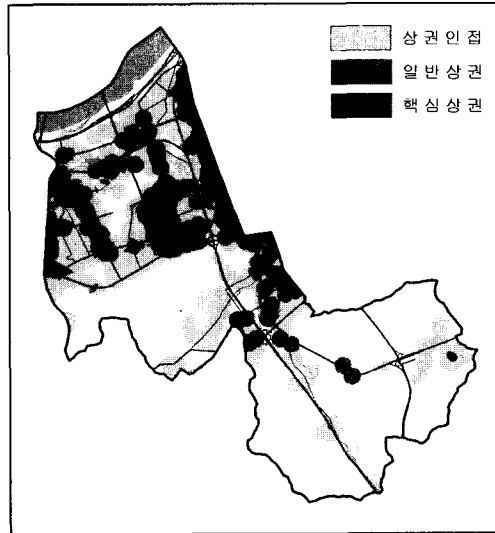
[그림 5]는 DENCLUE기법을 이용하여 가맹점의 분포에 의한 상권을 추출한 결과이다. 밀도는 각 격자위치로부터 반경 200m내에 포함된 점포의 수를 면적으로 나누어 계산되었다. 계산된 밀도로부터 4분위수를 구하여 다음과 같이 결과를 분할하였다. 1사분위수 이하의 지역은 상권이 형성되지 않은 지역, 2사분위수 이하의 상권에 인접한 지역, 3사분위수 이하의 일반상권 지역, 3사분위수 이상은 핵심상권 지역으로 분할하였다.

[그림 6]은 단순한 가맹점 수가 아니라 가맹점에서 나타나는 월평균 매출에 DENCLUE 기법을 적용하여 상권추출한 결과이다. 처리과정과 기준은 [그림 5]와 동일하다.

밀도의 계산은 일반적으로 square-wave function에 의해 일정 범위 내에 포함되는 것은 모두 동일한 가중치로 계산에 산입하는 것이 일반적이나, DENCLUE기법에서는 Gaussian function, Inverse Distance Weight 등과 같은 거리체감함수를 사용하기도 한다. [그림 7]은 가맹점의 월평균 매출에 DENCLUE기법을 적용할 때 단순

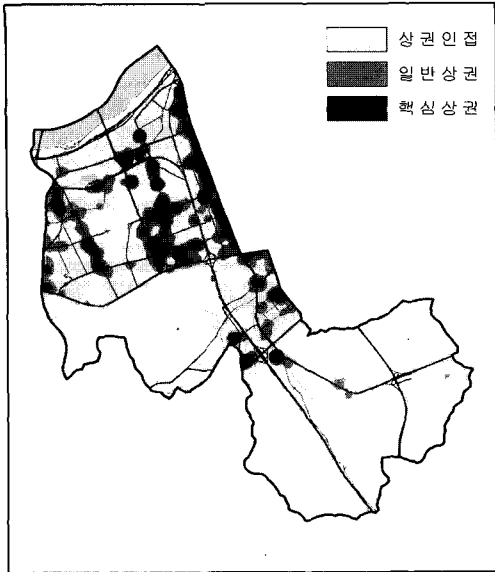


[그림 5] DENCLUE기법에 의한 가맹점 상권 추출 (단순 가맹점 밀도)

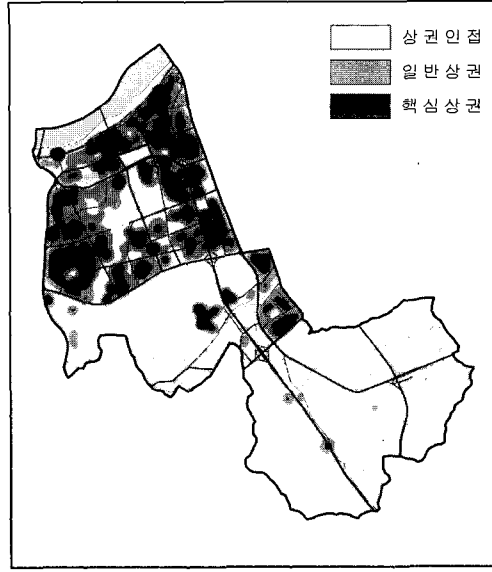


[그림 6] DENCLUE기법에 의한 가맹점 상권추출 (단순 가맹점 매출 밀도)

1) 이상적인 Gaussian function은  $K(s,t) = \exp(-(s^2 + t^2)/(2C^2))/(2\pi C^2)$  이다. 여기서, (s,t)는 중심점으로부터의 상대좌표이므로 위 식을 중심점으로부터의 거리 r과 방위각 theta를 이용한 극좌표로 고치면,  $K(r,theta) = \exp(-r^2/2C^2)/(2\pi C^2)$ 과 같이 나타난다. C는 함수 K가 영향을 미치는 범위로 3\*C 이상의 거리에서 K는 거의 0이 된다. 한편,  $q(r) = (1-(r/3)^2)^2$ 은 r = 0에서 1이고, r=3에서함수값이 0이면서도 도함수값도 0인 특성을 가지고 있어 K(r,theta)와 유사한 특성을 갖는다. 따라서, Gaussian function을 다음의 4차식으로 근사 시킬 수 있다.  $K(r,theta) = (1-(r/3C)^2)^2$



[그림 7] DENCLUE기법에 의한 가맹점 상권추출 (거리체감함수를 적용한 가맹점 매출 밀도)



[그림 8] DENCLUE기법에 의한 소비자 회원 상권추출 (거리체감함수를 적용한 회원 밀도)

밀도가 아닌 거리체감함수에 의한 밀도를 적용한 상권추출 결과이다. ArcGIS8.3에서는 Gaussian function의 4차 근사식에 의한 거리체감함수<sup>7)</sup>를 제공한다.

[그림 8]은 B소매체인 회원의 수로부터 거리체감함수를 이용한 밀도를 계산하여 DENCLUE기법을 적용한 결과이다.

### 3.3 평가

4.2에서 여러가지 기법을 이용하여 추출된 상권을 두 가지 측면에서 비교, 평가하고자 한다. 하나는 상권추출에 사용된 기법 간의 성과를 비교, 평가하는 것이고, 다른 하나는 A카드 가맹점 정보를 이용하여 추출된 상권과 B소매체인 회원 정보를 이용하여 추출된 상권을 비교, 평

가하는 것이다.

먼저, 상권추출에 사용된 기법 간의 비교, 평가이다.

k-medoid기법은 알려진 바와 같이 공간상의 구조(위계)나, 형상과 같은 GIS에서의 활용을 위한 정보는 거의 밝혀내지 못하고 있어 상권추출에 사용하기에는 적합하지 않은 것으로 보인다.

SMTIN기법은 k-medoid기법에 비해 공간상의 위계나 형상에 대해 더 많은 정보를 찾아낼 수 있는 방법이기도 하지만, 크게 두 가지 약점을 가지고 있음을 알 수 있다.

첫째는 임계치의 설정에 관한 문제로, 시행착오를 거치지 않으면 적절한 임계치를 설정하기 어렵다는 점이다. 본 연구에서 사용된 임계치 200m는 DENCLUE 결

과와 비교할 때 상권이 지나치게 과소하게 나타나고 있다. 이러한 문제는 지역이 달라지거나, 전체 업종이 아니라 부분부분의 업종을 따로 추출하여 처리하는 경우, 즉 대상 데이터의 평균적인 산포가 달라지는 경우에는 분석의 일관성을 유지하기 어렵게 된다.

둘째는 본 연구에 사용된 데이터와의 적합성 문제이다. 본 연구에서뿐만 아니라 대부분의 비즈니스 GIS 응용시스템에서는 입력되는 다양한 정보에 주소와 지번을 매개로 하여 좌표를 부여한다. 즉, 고속터미널과 같은 복합상가에 입점한 여러 개의 점포가 동일한 지번을 사용하고, 주상복합과 같은 건물에 거주하는 회원이 동일한 지번을 사용하게 되므로 하나의 좌표에 여러 개의 점사상이 중복하여 위치하게 된다. 이 경우 TIN구성 시 중복된 점사상 중 하나의 사상만을 사용하게 되므로 거리를 임계치로 사용하는 단순한 클러스터링 과정에서 누락되는 부분이 발생하게 된다. 예를 들어 본 연구에서는 고속터미널의 양대 복합 상가가 상권에서 누락되었다.

마지막으로 DENCLUE기법은 공간상의 위계와 형상에 대해 많은 정보를 산출해주는 것과 동시에 다양한 속성정보를 다양하게 사용할 수 있는 유연성을 제공해 줌으로써 비즈니스 GIS와 같이 동적인 다양한 정보를 사용해야 하는 경우에 적합한 방법임을 확인할 수 있었으며, 특히 다음과 같은 두가지 점에서 상권추출에 장점을 가지고 있음을 알 수 있었다.

첫째는 상권추출을 위한 전처리로 밀도를 구하는 과정에서 매출액과 같은 다양

한 정보의 활용과 거리체감함수를 사용한 영향력의 조정이 가능하다는 점이다. 본 연구에서도 단순한 점포수 밀도, 점포의 평균 매출액 밀도를 사용하여 상권을 추출하였으며, 단순밀도, 거리체감함수에 의한 밀도를 사용하여 상권을 추출하였다. 이외에도 업종별 가중치, 영업 연수별 가중치 등 다양한 변수를 동시에 고려하여 밀도를 계산하는 것이 가능하기 때문에 사용목적에 따른 상권추출이 가능하다.

둘째는 상권추출의 일관성을 유지할 수 있다는 점이다. 계산된 밀도의 4분위수를 구함으로써 3사분위수 이상을 핵심상권, 2사분위수에서 3사분위수의 범위를 일반상권 등과 같이 분할하는 것은 지역과 업종, 데이터의 평균 산포와 무관하게 일관되게 적용할 수 있고, 이해할 수 있는 지표이다. 물론, 지역과 업종에 따라서 또는 사용목적에 따라서 몇 분위수를 구하여 사용할 것인 가는 또 다른 연구의 대상이지만, 본 연구에서와 같이 일단 하나의 기준을 설정하면 다른 데이터와 다른 속성에서 구해진 결과에 대해서도 동일하게 적용하고 해석할 수 있게 된다.

다음으로 카드 가맹점 정보를 이용하여 추출된 상권과 소매체인 회원 정보를 이용하여 추출된 상권을 비교, 평가하였다.

A카드 가맹점의 위치와 매출로부터 추출된 상권은 주로 강남대로, 고속터미널, 교대역, 방배역 등의 역세권이나 접근성이 우월한 지역에 상권이 형성되어 있으며, 특히, 대규모 쇼핑공간이 존재하는 지역(예, 양재 I.C. 인근 쇼핑센터)에서는 지역적 범위의 크고 작음 또는 점포수에 의한 핵심상권 여부와 관계 없이 높은 매출

이 일어나고 있어 그 자체가 하나의 상권이 되고 있음을 보여주고 있다.

B소매체인 회원의 위치로부터 추출된 상권은 일반적으로 아파트가 많은 지역이 핵심상권으로 일반주거 밀집지역이 일반상권으로 추출된 결과를 알 수 있다.

그림 5와 그림 8을 비교하면 방배동, 양재동, 포이동, 서초로 인근 지역에서 상권의 중복이 일어나고 있음을 알 수 있다. 반면에 그림 7과 그림 8을 비교하면 속칭 방배동 카페골목으로 알려진 지역을 제외하면 대부분의 지역이 중복되지 않음을 알 수 있다. 이는 앞서 설명한 바와 같이 카드 가맹점의 경우 카드매출이 주로 선매점에서 많이 나타나기 때문에 그림 5에 비해 그림 7에서 핵심상권으로 나타나는 지역이 선매점의 점포가 위치하거나 매출이 큰 대규모 쇼핑센터가 입지해 있기 때문에 나타나는 현상이다. 반면에 그림 5에서만 핵심상권으로 나타나는 지역은 방배동, 포이동, 서초로 인근 지역과 같이 주로 음식점이 입지해 있는 지역으로 카드에 의한 매출이 상대적으로 작으면서 점포의 수는 많은 업종이 주거지와 인접하여 입점해 있음을 알 수 있다.

이와 같이 두 종류의 상권 비교를 통해 향후 점포의 입지 선정이나 상권의 발전 방향 등을 찾아 내는 데 도움이 될 수 있으며, 특정 업종 점포가 입점한 상권의 일반적인 특성을 파악하는데 유용한 결과를 얻을 수 있음을 증명하였다.

이상의 연구를 통해 공간 데이터마이닝 기법이 비즈니스 GIS에 적용되어 경제적 현상을 설명하는 데에도 유용함을 확인할 수 있었으며, 비즈니스 GIS에 활용되는

다양한 동적 데이터를 활용하여 일용품 상권, 선매점 상권과 같이 서로 다른 배경을 가지는 상권을 동일한 알고리즘에 의해 추출하여 동등하게 비교분석할 수 있다는 가능성과 향후 도시발전 모델링, 소비자행태분석 등 다양한 분야의 연구를 위한 기초정보 제작이 가능함을 확인할 수 있었다.

#### 4. 결 론

본 연구에서는 비즈니스 GIS의 동향을 분석하여 상권추출의 필요성을 제기하고, 상권추출을 위해 사용되는 여러가지 공간 클러스터링 기법을 알아보고 그 중 k-medoid기법, SMTIN기법 및 DENCLUE기법을 실제 데이터에 적용하여 그 유용성을 평가하였다.

본 연구의 결과 다음과 같은 결론을 얻을 수 있었다.

첫째, 상권추출을 위한 공간 클러스터링 기법으로는 DENCLUE기법이 유연성이나 일관성에서 우월한 결과를 제공한다.

둘째, 카드업체의 가맹점으로부터 추출되는 정보를 이용하여 선매점의 상권을 추출할 수 있다.

셋째, 소매체인 회원의 정보를 이용하여 일용품의 상권을 추출할 수 있다.

넷째, 추출된 상권을 비교, 분석하여 특정 업종의 점포가 위치한 지역의 일반적 특성을 파악할 수 있다.

본 연구가 가지는 한계로는 첫째, 추출된 상권에 대한 정량적 평가가 부족했으며, 둘째, 연구대상 지역이 서초구로 한

정되어 있어 제안된 알고리즘을 일반화하기에는 논리의 비약이 있을 수 있다는 점이다.

이와 같은 문제의 극복을 위해서는 첫째, 상권에서 추출된 특성정보와 상권에서 발생하는 매출 등을 이용하여 추출된 상권의 적정성을 정량적으로 평가할 수 있는 기법의 개발, 둘째, 제안된 방법을 더 많은 지역의 다양한 조건에 적용하여 일반화하기 위한 연구와 같은 추가적인 연구가 필요할 것이다.

향후에는 도시지역의 경년 변화를 예측하는 것과 유사하게 시계열로 추출된 상권을 이용하여 도시지역 내에서 소지역에서의 상권의 발전을 모델링하는 것과, 업종별 상권의 시계열적 분석을 통해 선발업종과 후발업종을 찾아 상권의 라이프사이클을 모델링하는 것 등 다양한 응용분야에서 본 연구에서 제시된 상권추출 알고리즘이 적용될 수 있을 것이다.

## 참고문헌

- Daniel, L., 1993, Enhanced modeling helps position business geographics in the retail industry, *business Geographics*, vol. 1, no. 5, pp.37-39.
- Han, H., et al., 2001., Spatial clustering methods in data mining A survey, *Geographic Data Mining and Knowledge Discovery*, Taylor & Francis, pp. 188-217.
- Huff, D.L., 1963, A probabilistic analysis of shopping center trade areas, *Land Economics*, vol. 39, pp.81-90.
- Kang, I., et. al., 1997, A Spatial Data Mining Method by Delaunay Triangulation, *Proceedings of the fifth ACM international workshop on Advances in geographic information systems*, pp.35-39.
- 안광호, 임영균, 1998, 유통경로관리, 문음사, pp.152-154.
- 정보통신부, 2003, 국가GIS 전문인력양성사업(교재개발 부문)-GIS 부문, [www.e-gis.or.kr](http://www.e-gis.or.kr)