

평면 데이터 분포에 영향을 끼치는 점 분포의 부분집합 추출 방법

- 소규모 소매점포의 매출자료를 이용한 상권 및 경쟁력 분석기법을 사례로 -

이정은^{1*}, 사다히로 유키오¹

A Method for the Extraction of a Subset of Points from a Large Set of Points Affecting the Distribution of Surface Data

- A Case Study of Market Area and Competitive Power Analysis by
Sales Data of Micro Scale Retail Stores -

Jung-Eun LEE^{1*}, Yukio SADAHIRO¹

요 약

공간분석에서는 사용되는 공간데이터 종류에 따라, 각각의 특징있는 분석법을 요구하므로, 다양한 접근법을 필요로 한다. 특히, 두 종류의 상이한 공간데이터 집합이 공존하는 경우, 어느 한 공간객체의 분포에 영향을 끼치는 다른 공간객체 집합에 대한 부분집합을 추출하여, 공간객체간의 관계성을 규명할 수가 있다. 지리정보시스템 등의 공간객체 분석기술의 발달로, 시각적으로도 간단히 공간객체간의 관계성을 파악할 수 있으나, 데이터의 양이 방대해지는 추세에서는, 지리정보시스템 (geographical information system)상에서 시각화된 공간객체 정보만으로는 서로 다른 공간객체간의 분포에서 상호관계성을 정량적으로 분석하기 어렵다. 따라서, 본 연구에서는 정량적 기준을 통해, 서로 다른 두 공간객체 간의 관계성 분석을 주목적으로 한다. 모델화된 정량적 기준의 평가를 위해 분석에 사용되는 실제 데이터는, 소규모 소매점포의 매출데이터를 사례로 들기로 한다. 데이터 수집에 제한이 있다는 특수환경으로 인해, 본 연구에서 사용되는 모델은 로짓모델을 기반으로 한 미지파라미터 추정이 가능하도록 구성되어있다. 제안된 모델을 사용하여, 대상점포의 매출분포에 큰 영향을 끼치는 것으로 판단되는 경쟁점포를 일련의 경쟁점포 집합에서 추출하며, 그 결과의 타당성을 검증하도록 한다.

주요어: 공간 객체간 상호관계, 마케팅 지리학, 점포선택 행동모델

2005년 6월 15일 접수 Received on June 15, 2005 / 2006년 3월 24일 심사완료 Accepted on March 24, 2006

1 동경대학교 도시공학전공 Department of Urban Engineering, University of Tokyo

* 연락처 E-mail: nalalee@ua.t.u-tokyo.ac.jp

Abstract

Approaches to spatial analysis differ from the type of spatial objects to be treated. Especially, in here, the case where two spatial data sets coexist is considered. The goal of such case lies on detecting a subset of spatial objects out of a large set that affects the distribution of the other object. However, it is not easy to extract a subset from a large set by visualization just with the help of GIS since huge amount of data are provided nowadays.

In this research, therefore, relationship between two different spatial data are analyzed by quantitative measure in the case study of marketing geography. A purchase history data of a small retail store and the location of its competitors are given as source data for the analysis. The goal of analysis from the aspect of this case study is to extract strong competitors of the store that affects the sales amount of the store among many competitors. With the result, therefore, it is expected that market area pattern and competitive power of stores under micro scale retail environment would be understood by quantitative measure.

KEYWORDS: *Relationship between Spatial Objects, Marketing Geography, Store Choice Behavior*

서론

계통지리의 한 분야인 공간분석에서는 사용되는 공간데이터 종류에 따라, 다양한 분석방법을 필요로 한다. 그 중에서도 상이한 두 종류의 공간데이터 집합이 공존하는 경우, 어느 한 공간객체의 분포에 영향을 끼치는 다른 공간객체의 부분집합을 추출해 냄으로써, 공간객체간 상호관계성을 규명할 수가 있다. 나날이 발달되는 지리정보시스템 등의 공간객체 분석기술로 시각화된 공간객체간의 관계성 파악이 가능하게 되었다. 하지만, 계속적으로 방대해지는 데이터의 추세로 볼 때, 이는 지리정보시스템 상에서 단순중첩기능 (overlay) 등으로 시각화되어 보여지는 공간객체의 정보만으로는 어떤 객체가 다른 공간객체의 분포에 어느 정도 영향을 끼치는지를 정량적으로 구분하기란 쉽지 않음을 보여준다. 이러한 한계를 극복하기 위해서는 시각적 접근뿐만 아니라, 정량적 기준으로도 그 관계성을 나타낼 필요성이 요구된다.

공간객체간의 관계성을 규명하는 하나의 예로, 존 스노우의 런던지도 (Snow, 1855)를 들

수 있다. 여기서 주목적은 콜레라 전염의 주요 원인을 규명하는 것으로, 발병 원인으로 추측되는 수도펌프의 분포 중, 발병에 가장 큰 영향을 미친 펌프들의 분포추출에 있었다. 이 사례에서는 시기적으로 오늘날과 같은 공간정보 해석에 관한 기술이 발달되지 않았기 때문에, 데이터의 분석자체가 정량적 기준을 제시하지는 못하였다.

공간분석이 적용 가능한 분야는 다양하지만, 본 연구에서는 마케팅 지리학의 관점에서 상이한 공간객체간의 관계성을 정량적 기준을 통해 알아보려고 한다. 마케팅 지리학은 고객의 구매행동 및 사고와 점포의 경영 및 마케팅 활동간의 상호관계를 지역 및 공간적 관점에서 관찰하고, 그 변화 및 유의성을 체계화하는 학문으로 (Davis, 1976), 소매시장 상권분석은 소매업의 지리적 공간구조 및 고객의 구매행동을 분석하고 이해하는데 초점을 두고 있다. 이는 소매시장권의 체계가 날이 갈수록 세분화되고 있으며, 경쟁력 또한 치열해지는 상권자체의 특성상 거시적 관점이 아닌 미시적 관점에서의 상권분석이 중요한 요소임을 보여준다. 미시적 관점에서의 소매시장권 분석을

위해서는 특히, 고객의 구매행동 분석이 주목적이 되고 있다. 이러한 추세는 시장 규모가 커짐에 따른 일련의 고객 구매행동에 대한 이해의 기회가 줄어드는 경향을 간과해서는 안 되는 것임을 의미하며 (Kotler, 1991), 또한 소매점포 운영자들의 입장에서는 고객의 구매행동의 이해가 마케팅 전략수립에 중요한 정보가 된다.

고객의 구매행동 및 점포선택행동 이해를 위한 기존 연구들에서, 모델의 수립은 크게 두 가지 접근법으로 나뉘어진다. 첫 번째는 관찰 및 표준 등의 가정을 기반으로 한 기술 및 결정적 접근법(descriptive-determinist approach)이다. Applebaum 및 Cohen(1961)은 고객에 대한 직접설문을 바탕으로, 시장권에 대한 지리적 패턴을 경험적 방법으로 파악하였다. 또 다른 방법으로는 Christaller(1935)와 Lösh(1954)가 제안한 중심지이론 (Center Place Theory) 이 있으며, Reilly (1931)는 점포선택 행동에 영향을 주는 요인으로 고객 및 점포간의 거리 외에 점포의 매력도를 포함한 분석방법인 라일리 법칙 (The law of retail gravitation)을 사용하였다.

두 번째 접근법은, 과거에 얻어진 데이터베이스 등의 정보로부터 소비자의 구매기회 선택행동을 경험 및 확률적으로 분석하는 접근법 (explicative-stochastic approach)을 들 수 있다. 이는 기술 및 결정적 접근법과는 달리, 고객의 점포선택행동에 있어, 어느 한 점포만을 선택하는 것이 아닌 몇 개의 점포를 선택할 수도 있다는 확률적 개념을 포함하고 있다.

허프 모델 (Huff, 1963)에서는 Luce (1959)가 제안한 고객의 한 점포에 대한 선택행동 확률은 모든 점포들의 유용성의 합에 대한 해당 점포의 유용성과 동일하다는 확률개념을 도입하였으며, McFadden (1974)과 Fotheringham (1983)은 로짓 모델 (Logit model)을 사용한 분석방법을 제안하였다.

위에 언급된 기존의 마케팅 지리학 연구동

향은 고객의 소비행동 및 점포선택행동을 기반으로 한 시장권 분석을 위한 주요데이터로 직접적인 고객설문조사, 경쟁점포들의 매상 및 마케팅 전략 데이터 등의 상세 정보가 필요하다. 하지만, 이런 대량의 상세 정보들을 수집하는 데는 막대한 비용 및 시간이 소요되는 것이 불가피하고, 소규모 소매점포의 경우는 이러한 데이터의 수집 자체가 사실상 불가능하다. 또한, 경쟁점포에 대해서는 위치정보 이외에는 수집이 불가능하므로, 지리정보시스템에서의 영향권 설정 (multi-ring buffering)을 이용한 상권의 공간분포분석도 현실적으로 어렵다.

그렇기 때문에, 소규모 소매점포가 상세 정보 수집의 제한이 있는 조건 하에서, 수집 가능한 해당 점포의 매출데이터 만으로 어느 정도의 상권분석이 가능한 방법 제안에 초점을 두었다. 따라서, 본 연구에서는 데이터 수집의 제한이 있는 상황 하에서, 마케팅 지리학 분야의 경우를 사례로, 한 공간객체가 다른 객체에 미치는 영향을 정량적 기준을 사용하여 두 공간데이터 간의 관계성을 보다 명확히 하는 방법론을 제시하고, 그에 대한 평가를 하고자 한다. 구체적으로는 대상점포의 매출분포에 영향을 미치는 것으로 판단되는 경쟁점포들의 부분집합을 정량적 기준을 통해 추출해 내는 것이 주목적이다. 사용되는 데이터는 소규모 점포의 주문배달 매상데이터 및 상권 내 경쟁점포의 위치정보이다.

제안된 방법론으로 먼저 상권분석을 위한 점포선택 행동의 모델화 및 가상데이터를 이용한 시뮬레이션을 실시하며, 제안된 모델에서 모델의 성능에 영향을 끼치는 주요 공간적 인자들을 파악한다. 가상데이터를 이용한 모델평가 결과를 기반으로 하여, 그 다음단계에서는 모델을 실제 데이터에 적용함으로써 대상점포의 매출분포에 영향을 끼치는 경쟁점포를 점포 매력도로 정의된 정량적 기준을 통해 추출하고, 그 결과를 바탕으로 제안된 모델의 성능 평가를 하고자 한다.

연구 방법

1. 분석에 사용되는 공간데이터

본 연구에서는, 서로 다른 두 공간 객체 즉, 평면 데이터 및 점 데이터 분포 사이에서, 평면 데이터의 분포 패턴에 영향을 미치는 주요 점 데이터 분포의 부분집합 추출을 목적으로, 마케팅 지리학의 소매 시장 상권분석을 기반으로 한, 소규모 소매점포 환경에서의 데이터를 사례로 상권분석을 하고자 한다.

마케팅 지리학에서는, 단순히 공간객체의 위치를 나타내는 x, y 좌표만으로는 데이터로서의 의미를 갖지 못하므로, 주어진 공간객체의 특성을 충분히 활용한, 미시적 시각에서의 상권 내 지리적 패턴 분석이 가능한 방법론의 제안을 중요시하고 있다 (Openshaw, 1995). 이런 이유를 바탕으로, 대개의 경우, 상권분석을 위해서는 해당 점포 및 경쟁 점포들의 매출데이터, 고객의 취향, 매출과 직결되는 점포의 매력도 등 다방면에서의 상세 정보가 요구된다. 하지만, 이러한 상세정보의 수집은 상권분석을 필요로 하는 소규모 소매점포들의 경우에는 비용, 소요시간 면에서 사실상 불가능한 작업이다. 이러한 소규모 소매점포의 특수한 상황을 고려할 때, 대부분의 소규모 소매점포들이 수집 가능한 데이터라면 해당 점포의 매출데이터 정도이다.

본 연구에서 사용되는 실제 데이터는 한 소규모 소매점포의 주문배달 대상데이터 및 상권 내 경쟁점포들의 위치정보로 제한되어있다. 텍스트 파일 형식으로 제공된 데이터들은 주문배달 대상데이터는 평면 데이터로, 경쟁점포의 위치정보는 점 데이터로 변환되어 지리정보시스템 상에서의 표현 및 상권분석에서 사용된다.

2. 점포선택 행동의 모델화

본 연구의 목적을 소규모 소매점포의 상권분석의 시각에서 다시 기술하자면, 평면 데이터

로 표현되는 한 점포의 매출데이터의 분포를 기반으로, 이 분포에 큰 영향을 미치는 것으로 예상되는 몇몇 경쟁점포를 다수의 경쟁점포들 중에서 추출해 내는 것이다.

이 때, 영향력 있는 몇몇 경쟁점포를 추출하기 위한 정량화된 기준은 각 점포들의 매력도이다. 점포의 매력도는 일반적으로 다변량 분석을 통해 주요 영향인자들에 대한 통계분석이 요구된다. 하지만, 본 연구에서는 분석에 필요한 데이터 수집에 한계가 있는 상황에서 점포 매력도 추정을 목적으로 하고 있다. 그러므로 본 연구에서 일컫는 점포의 매력도라 함은, 소비자의 점포선택 행동 확률과 비례하는 기준치로 정의하기로 한다.

본 연구에서 정의된 점포 매력도 추정방법으로는 Ben-Akiva 와 Lerman (1985)이 제안한 다항 로짓 모델을 기본 개념으로 기술할 수 있으며, 이는 다음의 식과 같다.

$$P_n(i) = \frac{e^{V_{jn}}}{\sum_{j \in C_n} e^{V_{jn}}} \quad (1)$$

식 (1)에서 $P_n(i)$ 는 집합 C_n 의 원소인 i 가 n 의 의사결정자들에 의해 선택될 확률을 의미한다. C_n 는 가능한 선택사항을 원소로 하는 집합, j_n 은 선택사항 집합의 원소개수, 그리고 V_{jn} 은 선택사항의 유용도를 나타낸다.

식 (1)을 점포선택 행동의 확률을 기술하는 요소들로 재구성하면 다음과 같이 나타낼 수 있다.

$$p_i(\mathbf{x}) = \frac{A_i \exp(-\alpha|\mathbf{x} - \mathbf{z}_i|)}{\sum_j A_j \exp(-\alpha|\mathbf{x} - \mathbf{z}_j|)}$$

$$s.t. \sum_{j=1}^N A_j = 1, \forall A_j \geq 0, i \in j \quad (2)$$

$p_i(x)$ 는 소비자가 N 개의 점포들 중에서 점포 i 를 선택할 확률로, 점포 i 의 매력도 A_i 에 비례하며, 위치 x 의 고객과 위치

z_i 의 점포 i 간의 거리 $|x - z_i|$ 와는 반비례한다. α 는 거리에 대한 상수이다.

수식 (2)를 계산하는데 있어서의 문제점은 식 (2)에 기술된 확률은 소규모 소매 점포가 수집 가능한 데이터만으로는 직접적으로 값을 대입하여 확률 자체를 계산해 내기는 불가능하다는 점이다. 점포의 매력도를 추정하는 기존 연구들 중에는 소비자를 대상으로 한 설문조사를 기반으로 제작된 구매 OD(origin destination) 행렬을 사용하는 방법 (矢野 등, 2002) 및 라이프사이클 집단에 따른 고객의 소비성향과 그 비용의 흐름을 이해하기 위하여 분석의 기본 데이터로 복수개의 백화점에 대한 소비자의 구매 데이터를 사용한 연구 (中谷와 花岡, 2003) 등이 있는데, 이들은 모두 식 (2)의 확률을 직접 계산할 수 있는 충분한 상세 데이터를 바탕으로 한 연구들이다.

하지만, 본 연구에서는 분석 데이터의 수집에 있어 제한을 받는 소규모 소매점포의 시각에서 상권분석이 가능한 방법론은 제안하고자 한다. 그러므로 미지 파라미터인 점포 i 의 매력도 A_i 및 거리상수 α 를 추정하는 간접적 접근법으로 계산해야만 한다.

3. 미지파라미터의 추정

앞 절에서 언급했듯, 식 (2)의 확률을 직접적으로 계산하는 것은 불가능하므로, 간접적인 방법을 사용해야 한다. 그러기 위해서, 분석 대상 상권이 격자형으로 구성되어있다고 정의하고, 고객이 한 점포를 선택할 확률은 실제로 그 점포의 매출과 비례한다는 배경아래, 식 (2)의 점포선택 확률을 다시 기술하면 식(3)과 같다.

$$P_i(U_k) \approx Mn_k \frac{A_i \exp(-\alpha |x_k - z_i|)}{\sum_j A_j \exp(-\alpha |x_k - z_j|)}$$

$$s.t. \sum_{j=1}^N A_j = 1, \forall A_j \geq 0, i \in j \quad (3)$$

$P_i(U_k)$ 는 수집한 데이터 중 상권이 격자형의 공간단위 U_k 로 구성되어있다고 하였을 때, 각 공간단위별 소매점포 i 의 매출액을 나타내며, M 은 고객의 평균 구매금액을, n_k 는 공간단위 U_k 내의 인구밀도를 나타낸다. 식 (3)에서 나타나듯, 점포의 매력도 A_i 와 점포의 매출액 $P_i(U_k)$ 은 서로 비례관계에 있으므로, 이 관계를 이용하여 미지파라미터인 A_i 및 α 를 추정하기로 한다. 즉,

$P_i(U_k)$ 값은 수집 데이터에서 계산이 가능하므로, 이 값에 근사치를 보이는 미지파라미터 값의 추정을 하면 된다.

하지만, 식 (3)의 정의만으로는 정확한 의미에서의 상권분석이 어렵다. 그 이유는, 식 (3)에서 정의된 점포 i 의 상권 내 공간단위별 매출액 $P_i(U_k)$ 은 공간단위별 구매총액을 나타낸 것으로, 고객 1인당 구매총액으로 표준화되지 않은 상태이기 때문이다. 이러한 표준화되지 않은 구매총액으로는 점포의 상권 내 매출분포 및 그에 따른 다른 경쟁점포의 영향력을 제대로 추정하기 힘들다. 그러므로, 식 (4)에 기술된 바와 같이, $P_i(U_k)$ 값을 공간단위별 인구밀도 n_k 로 나눈 값인 $S_i(U_k)$ 을 기준으로 하여, 점포 매출분포를 1인당 구매총액으로 표준화하여, 우변에 있는 미지파라미터 A_i 및 α 를 계산하는 것이 보다 정확한 의미에서의 추정이라고 하겠다.

$$S_i(U_k) = \frac{P_i(U_k)}{n_k} = M \frac{A_i \exp(-\alpha |x_k - z_i|)}{\sum_j A_j \exp(-\alpha |x_k - z_j|)}$$

$$s.t. \sum_{j=1}^N A_j = 1, \forall A_j \geq 0, i \in j \quad (4)$$

식 (4)의 $S_i(U_k)$ 값은 수집된 데이터로 계산이 가능하며, 우변의 수식표현 중에서는 각 공간단위별 고객과 점포간의 거리 $|x - z_i|$ 및 고객의 평균 구매금액 M 은

수집 데이터로부터 계산이 가능하다. 그러므로 미지파라미터인 A_i 및 μ 의 값은 우변의 수식 값이 $S_i(U_k)$ 의 값과 근사치를 갖는 경우의 결과가 추정치가 된다.

본 연구에서 거리상수인 μ 에 대해서는, 일단 매출분포 및 점포와 고객간의 거리에 대한 관계에서 그 값을 추정하여 그 값은 모델 정의에 의해 거리상수로 두고, 경쟁점포의 영향력을 평가하는 것이 주목적이므로 또 다른 미지파라미터인 A_i 의 추정을 중심으로 상권분석에 사용한다.

연구방법의 적용 및 고찰

제안된 연구방법은 크게 두 단계로 나뉘어 적용된다. 첫 번째는 가상 데이터를 이용한 점포선택 행동 시뮬레이션 단계이다. 이 단계에서는 실제 데이터를 이용한 계산과는 달리, 모델 성능에 영향을 끼치는 인자들이 어떤 것인지 파악하는데 있어, 다양한 관점에서의 예비실험이 가능하다. 두 번째 단계로는 실제 데이터를 이용한 모델의 성능평가 단계로서, 이때 모델이 내놓는 결과의 타당성 여부는 실제 데이터가 보여주는 상황과 얼마나 근접한 결과를 계산해 내는지를 기준으로 판단하게 된다.

1. 점포선택 행동 시뮬레이션

제안된 방법론을 먼저, 점포선택 행동모델에 가상데이터를 이용한 시뮬레이션으로 성능평가 하기로 한다. 이 과정에서는 경쟁점포의 개수 및 위치적 요소 (경쟁점포와 대상점포 사이의 거리 및 방향) 등 다양한 인자를 통해, 제안된 모델에서 어떤 인자가 모델의 성능에 영향을 끼치며, 이를 바탕으로 한 소규모 소매점포 환경에서의 상권분석에서 모델이 어떤 성능을 보이는지 살펴보는 것이 주목적이다.

점포선택 행동 시뮬레이션의 흐름을 단계별로 간략히 설명하자면 다음과 같다.

* 단계1

- 로짓 모델을 기반으로 정의된 식(4)에서 점포 i 에 대한 공간단위 U_k 에서의 1인당 구매총액에 해당하는 $S_i(U_k)$ 값을 계산하기 위한 기초데이터를 생성한다. 이 때 생성해야 할 데이터로서는 점포의 개수 및 위치정보, 점포의 매력도 등이 있다. 특히, 점포의 위치결정에 있어서는 다양한 공간적 옵션을 고려할 것이며, 이는 후반부에 추가설명 될 것이다.

* 단계2

- 단계1에서 생성된 기초데이터를 사용하여 가상의 $S_i(U_k)$ 의 값을 계산한다.

* 단계3

- 가상 데이터로부터 계산된 $S_i(U_k)$ 값을 기준으로 미지파라미터인 각 점포의 매력도 추정을 실시한다.

* 단계4

- 추정 결과를 토대로 모델 성능을 평가 및 결과 분석을 한다.

1) 소수의 점포를 대상으로 한 시뮬레이션

제안된 방법론의 평가를 위해 대상점포 및 그의 경쟁점포들간의 2 가지 공간적 특성을 고려한 시뮬레이션 시나리오를 구성하도록 한다. 그 첫 번째는 대상점포 및 경쟁점포간의 상대적 거리이다. 그림 1에서 나타나듯, N (Near), M (Moderate) 그리고 F (Far)로 나누어 상대적 거리 옵션을 두도록 한다. 이 때, N을 기본거리라 할 때, M은 N의 2배, F는 N의 3배가되는 거리로 정의한다. 공간적 특성의 두 번째 요소는 대상점포를 중심으로 한 점포들간의 상대적 배치방향 이다. 그림 1에서 보듯이, 대상점포를 중심으로 하여 상권을 두 부분으로 나누었을 경우, 경쟁점포들이 그 경계선을 기준으로 같은 방향의 상권범위에 있으면 S (Same), 서로 다른 방향의 상권범위에 위치하면 O (Opposite)의 값을 옵션으로 할당한다.

예를 들면, 그림 1에서 거리 옵션 N과 M의 위치에 있는 경쟁점포들 간에는 방향 옵션 S를 가지고, 거리 옵션 M과 F 그리고 N과 F의 경쟁점포들의 경우는 방향 옵션이 O가 된다.

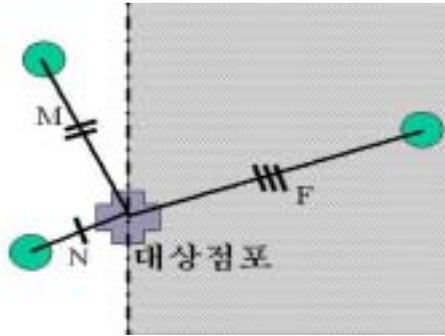


FIGURE 1. 경쟁점포들의 상대적 거리 및 방향

제안된 모델의 성능평가기준 3가지로서는, 1) 모든 점포의 순위정확도, 2) 영향력이 큰 경쟁점포, 3) 제곱오차의 평균을 제안하며, 아래에 각각에 대해 설명하도록 한다.

첫 번째로, 모든 점포의 순위정확도는 추정된 점포의 매력도를 기준으로 했을 경우, 추정된 값에 의한 점포의 순위가 가상데이터 생성시 할당된 원래 순위와 같은 경우를 말한다. 예를 들어, 원래 세 점포들의 순위가 1-2-3이라면, 추정된 점포 매력도의 순위가 1-2-3인 경우만이 이 기준을 만족시키는 결과에 해당된다.

두 번째로, 영향력이 큰 경쟁점포는 대상점포의 순위를 제외한 경쟁점포들만의 상대적 순위를 고려하는 것이다. 가상데이터에서 주어진 점포들의 순위가 1-2-3이라고 하자. 이 때, 순위 1은 대상점포를 나타내고, 순위 2, 3의 점포가 경쟁점포라고 할 때, 대상점포를 제외한 경쟁점포들에 대한 추정된 순위가 2-3, 1-3, 1-2의 경우들이 허용된다.

세 번째 모델 성능평가 기준은, 제곱오차의 평균 즉, mean square error (MSE)라고 하며,

앞서 언급한 두 가지의 성능평가기준이 추정된 순위를 기준으로 했다면, 이것은 미지파라미터의 실제 추정치가 원래 값에 얼마나 근사하게 계산되었는지를 정량적으로 알아보는 것이다. 여기서 MSE 값은 각 시나리오별 복수개의 추정된 파라미터들의 MSE 평균치를 취한다. 모델의 성능은 MSE 값이 작을수록, 추정의 오차가 작음을 의미하므로, 정확한 추정치라고 판단할 수 있다.

표 1은 상대거리 및 방향에 의한 시나리오별 추정계산결과를 정리한 것이다. 주어진 점포의 매력도에 의한 각 점포의 순위는 1-2-3이 된다. 여기서 순위가 1인 점포가 대상점포, 나머지가 경쟁점포들로 정의된다. 시나리오 ID는 경쟁점포의 수와 거리 및 위치 옵션에 따라 구분되며, 예를 들면, 시나리오 ID 2-1은 경쟁점포의 수는 2이며, 1이 거리 및 위치 옵션으로 생성된 계산의 경우 중 첫 번째 시나리오임을 나타낸다.

계산횟수는 파라미터의 초기치 등의 조건을 달리하여 총 1,000회를 실시하였으며, 그 중 무의미한 결과를 제외한 총 689개의 분석이 가능한 해를 얻었다. 689개의 가능해 중에서, 196개의 결과가 영향력이 큰 경쟁점포의 순위가 맞게 추정된 것으로 나타났다. 그리고, 모든 점포의 순위정확도가 맞게 추정된 가능해의 수는 196개로 나타났다.

점포간의 상대적 위치가 추정결과에 미치는 영향을 알아보기 위해 경쟁점포 위치가 같은 방향인 경우, 반대방향인 경우로 나누어 점포들 순위가 정확히 추정된 계산 수를 정리해보았다. 방향옵션이 S인 경우가 190개, O인 경우가 202개로, 이 때의 추정정확도를 나타내는 MSE 값은 방향옵션이 S인 경우가 0.153, O인 경우가 0.130이었다. 이로 볼 때 추정결과와 경쟁점포들의 위치가 대상점포를 중심으로 반대방향에 위치해 있는 경우, 보다 정확한 추정결과를 보였다.

다음으로, 대상점포 및 경쟁점포들간의 상대

적 거리 측면에서의 분석결과를 보기 위해 시나리오 ID 2-1 부터 2-6 까지를 그룹 1 (상대거리가 N 혹은 M이 포함된 경우), 그리고 2-7 부터 2-10 까지를 그룹 2 (상대거리가 M 혹은 F가 포함된 경우)로 나누어 각각의 MSE 값을 비교하였다. 그 결과, 그룹 1의 경우 평균 MSE값은 0.188, 그룹 2의 평균 MSE값은 0.071로 나타났으며, 이는 모든 경쟁점포가 M 혹은 F의 거리옵션인 경우가, 보다 정확한 추정 가능성이 가능했음을 보여준다.

2) 다수의 점포를 대상으로 한 시뮬레이션

1.1절에서는 소수의 점포를 대상으로 한 점포선택 행동의 시뮬레이션을 실시하였다. 하지만, 제안된 방법론의 확장을 위해서는, 실제 상황과 유사한 경쟁점포가 다수 존재하는 경우도 고려해야 할 것이다. 그러므로, 본 절에서는 점포의 수가 증가된 경우에 대한 시뮬레이션을 행하도록 한다.

여기서는 앞 절과는 달리, 시뮬레이션의 시나리오 수를 줄이도록 한다. 가장 큰 이유는, 앞 절에서 언급된 상대 거리 및 위치의 옵션을 모두 적용하게 되면 점포의 수가 늘어남에 따라 생성되는 시나리오 수도 기하급수적으로 늘어나기 때문이다. 그렇게 되면, 이는 모델의 평가에 있어 그다지 효율적이지 못하므로, 앞

절에서 언급된 옵션들 중, 모델 성능에 영향을 끼치는 요소로 판단된, 상대거리의 M 및 F 옵션만을 고려하기로 한다.

계산의 결과를 표 2에 정리하였다. 우선, 주어진 점포들의 매력도는 대상점포가 가장 큰 값을 갖도록 정의하고, 경쟁점포들에 대해서는 점차 작아지도록 설정하였다. 여기서의 시나리오 ID는 2개의 숫자 및 알파벳으로 구성되어 있다. 예를 들면, 시나리오 ID 9M-1은 경쟁점포의 수가 9, 거리옵션이 M, 그리고 초기치 설정의 변화를 구분하기 위해 1 혹은 2를 말미에 추가하였다.

모델의 성능평가를 위해 2가지의 기준을 설정하였으며, 그 첫 번째는 주어진 매력도 기준으로 상위 3순위에 드는 세 점포의 상대적 추정순위 순서가 정확히 계산된 경우, 두 번째는 그 상위 3순위의 세 점포의 추정된 순위 값이 1, 2, 3이 포함된 경우이다.

제시된 모델평가 기준에서 알 수 있듯, 다수의 경쟁점포의 경우는 모든 경쟁점포에 대한 순위추정 평가가 아닌, 상위 3개 점포에 대한 순위추정이 고려되었다. 이는, 본 연구에서 분석의 초점이기도 한, 평면 데이터 분포에 큰 영향을 미치는 점 데이터 분포의 부분집합을 추출하는 것이기 때문에, 다수의 경쟁점포들 중에서 대상점포의 매출분포에 끼치는 영향이

TABLE 1. 상대거리 및 방향에 의한 추정계산 결과

시나리오ID	상대적 거리 및 방향	추정순위 정확도			평균 MSE
		모든 점포 대상	영향력이 큰 경쟁점포	총 가능해 수	
2-1	N/N/S	30	9	55	0.089
2-2	N/N/O	10	16	52	0.246
2-3	N/M/S	13	9	35	0.246
2-4	N/M/O	21	14	53	0.129
2-5	N/F/S	12	12	32	0.289
2-6	N/F/O	23	17	67	0.131
2-7	M/M/S	23	28	97	0.070
2-8	M/M/O	22	30	100	0.075
2-9	F/F/S	22	32	99	0.070
2-10	F/F/O	20	29	99	0.069

* N: Near, M: moderate, F: Far, O: Opposite

TABLE 2. 다수 점포를 대상으로 한 추정결과

시나리오ID	추정순위 정확도			평균 MSE
	(1)	(2)	총 가능해 수	
9M-1	4	1	19	0.106
9M-2	4	2	19	0.098
9F-1	16	9	32	0.026
9F-2	11	7	42	0.033

*M: Moderate, F: Far

- (1) 추정된 매력도 기준으로 상위 3순위 점포들간의 상대순위추정이 정확히 된 경우
- (2) 상위 3순위 점포들의 추정된 순위값에 1, 2, 3이 포함된 경우

크다고 판단되는 상위 3개 점포만을 고려하기로 한 것이다.

표 2에 나타난 평균 MSE 값이 보여주듯, 다수 점포를 대상으로 한 계산에서도 경쟁점포가 대상점포로부터 멀리 떨어진 거리에 위치해 있을수록 MSE 값은 작게 나타났으며, 이는 곧 추정정확도가 높음을 보여주는 것이다. 다시 말해, 본 연구에서 제안된 방법으로 대상 점포의 매출분포에 대한 영향력 있는 경쟁점포의 추정이 가능함을 알 수 있다.

2. 실제 데이터에의 적용

이 절에서는 제안된 방법론을 실제 데이터인 도쿄의 한 소규모 소매점포의 주문배달 대상 데이터를 이용해 경쟁점포들에 대한 매력도 추정에 적용하고자 한다.

그림 2는 연구대상지역으로 대상 점포의 상권을 나타내며, 지도의 중심에 별표시가 있는 지점이 대상 점포의 위치이다.

사용되는 주문배달 대상데이터에는 고객의 위치정보, 주문내역 및 주문시각 등의 정보가 상세히 기술되어 있다. 이 데이터를 이해하기 위한 한 방법으로 일일 매출 총액의 분포를 그래프로 작성하면, 그림 3과 같다.

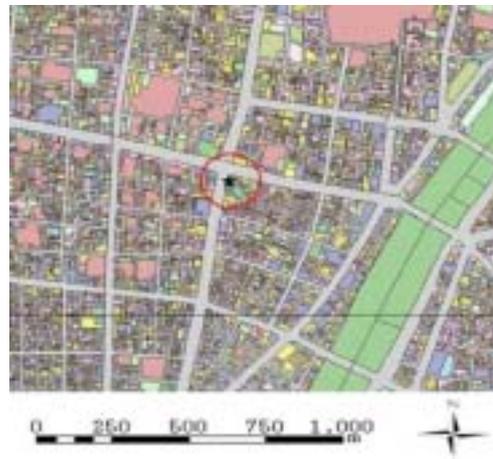


FIGURE 3. 연구 대상지역의 지도

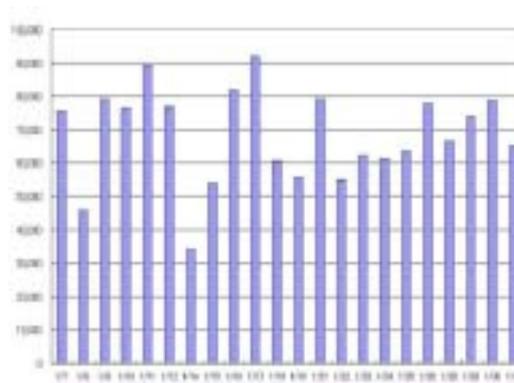


FIGURE 3. 일일 매출 총액(단위: 일본 엔)

그림 3에서 보여진 일일 매출 총액의 정보 이외에도 주어진 데이터로부터 추출해 낼 수 있는 정보가 더 남아있기 때문에, 데이터를 분석 과정에서 심분 활용하기 위해서는 텍스트로 주어진 데이터를 공간데이터 형태로 변환시킬 필요가 있다. 그 결과, 그림 3의 막대그래프 형식으로 표현된 데이터를 공간데이터 형태로 변환한 형식이 그림 4이다.

그림 4는 대상점포를 중심으로 한 상권 내의 매출총액을 인구밀도로 나눈 값의 분포를 격자모양의 공간단위 별로 나타내고 있다. 여기서 공간단위를 구성하는 한 격자의 크기는 한 변이 50 m이다. 그림 4를 통해, 대상점포의 매출액 분포에 대한 공간적 시각에서의 설명이 가능해진다. 즉, 상권 내의 매출분포가 대상점포의 부근에서는 그 값이 높으나, 거리가 멀어짐에 따라, 점진적으로 낮아지는 경향을 보이고 있어, 주어진 데이터의 공간적 특성과 파악이 가능하다. 하지만, 매출의 공간적 분포가 매끄러운 곡면을 형성하고 있지 않고, 상권 곳곳에 드문드문 매출액이 희박한 지점들이 눈에 띄는데, 그에 대한 가장 큰 이유로는 경쟁점포들의 유무 및 그 영향력을 들 수 있겠다.

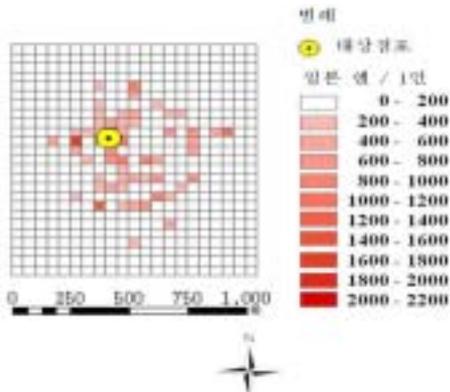


FIGURE 4. 대상점포를 중심으로 한 상권 내의 매출총액을 인구밀도로 나눈 값의 공간단위별 분포

그림 5는 그림 4에 대상점포 상권을 중심으로 한 대상점포의 매출분포 및 그에 대한 경

쟁점포들의 위치정보를 이용하여 그 분포를 표시한 것이다. 앞서 설명한 대상점포의 매출 분포가 매끄러운 곡면을 형성하지 않고 부분적으로 대상점포의 매출분포가 급격히 낮아지는 지점들이 눈에 띄는데, 그 위에 경쟁점포들의 위치정보를 실제로 표현하고 보니, 실제로 경쟁점포들이 위치한 지점에서 그러한 경향이 두드러지는 것으로 확인되었다.

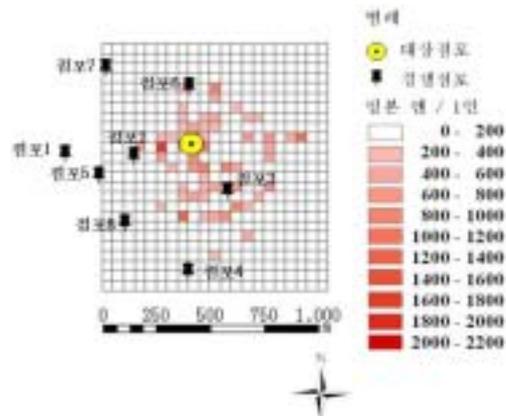


FIGURE 5. 상권 내 대상점포에 대한 경쟁점포들의 위치 분포

그림 6은 본 연구에서 제안한 방법을 통해, 추정된 경쟁점포들의 매력도를 도식화한 것이다. 여기서 추정된 각 점포의 매력도는 기존의 마케팅 분야에서 통용되는 각 점포의 규모, 매출, 고객성향 등의 방대한 데이터를 기반으로 계산된 것이 아니며, 소규모 소매점포라는 특수한 상황 하에서 제한된 데이터를 기준으로 계산된 것이기 때문에, 점포들간의 영향력을 비교하는 상대적인 기준치로 그 개념을 정의한다.

대상점포에 대한 경쟁점포들의 매력도는 그림 6에서 각 점포의 위치에 표시된 삼각형의 크기를 기준으로 표현되었으며, 그 크기가 큰 점포일수록 점포의 매력도가 크음을 나타낸다. 또한 이는 매력도가 큰 경쟁점포일수록 대상점포의 매출분포에 큰 영향을 미침을 의미한다. 이러한 결과로 미루어보아, 제안된 방법을

사용한 추정결과가 실제 데이터의 상황을 근접하게 설명한다는 것을 알아낼 수 있다.

가상데이터로 모델성능을 평가하면서, 모델 성능에 영향을 끼치는 것으로 판단되는 공간적 요소가 실제 데이터에서도 영향을 미치는지의 여부를 알아보도록 한다. 그림 6에서 영향력이 큰 경쟁점포들의 위치를 보면, 대상점포로부터 200~300m 혹은 500m 이상 떨어진 거리에 있음을 알 수 있다. 상권의 규모상 상대적 거리옵션의 기준거리 N을 150m라 할 때, 실제 데이터에서는 영향력이 큰 경쟁점포들이 거리옵션 M 혹은 F의 범주에 분포하고 있으며, 이는, 제안된 방법이 거리옵션이 M 혹은 F인 경우 추정결과에서 높은 정확도를 보였다는 것과 일치한다. 또한, 경쟁점포들간의 상대적 위치에 대해서는 상권을 대상점포 기준으로 어떤 방향으로 나누는가에 따라 다른 해석이 나올 가능성이 있으며, 그림 6의 실제 데이터를 볼 때, 영향력이 큰 경쟁점포들의 위치 역시 상권을 어떤 방향에서 둘로 나누느냐에 따라 해석이 달라질 수 있다. 따라서, 상대적 위치에 대해서는 모델성능의 평가가 어려움을 알 수 있다.

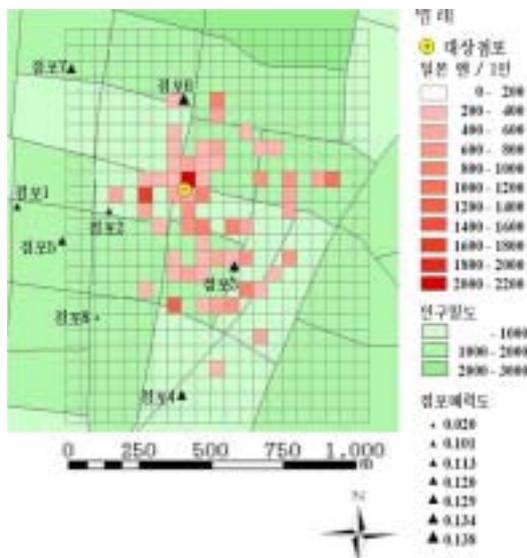


FIGURE 6. 경쟁점포들의 추정된 매력도

결론

본 연구에서는 상이한 두 공간데이터 집합이 공존하는 경우, 공간데이터 상호간의 관계성에 영향을 미치는 공간객체 집합의 부분집합 추출에 분석의 초점을 맞춰, 소규모 소매점포라는 특수한 상황을 사례로 들어 영향력 있는 공간객체의 부분집합을 추출해 내는 유효한 방법론을 제안하였다.

구체적으로, 본 연구에서는 복수개의 경쟁점포들 중에서 대상점포의 매출분포에 영향을 미치는 경쟁점포를 매력도로 정의된 정량화된 기준을 사용하여 대상점포의 매출분포와 경쟁점포간의 관계성을 보다 명확히 하는 방법론을 제시하고, 그에 대한 평가를 하였다.

먼저, 제안된 방법론의 특성 및 성능을 이해하기 위해 가상의 데이터를 생성하여, 점포선택행동 시뮬레이션을 실시하였다. 첫 번째로, 소수의 경쟁점포를 대상으로 한 경우, 대상점포 및 경쟁점포간의 상대적 거리 및 대상점포를 중심으로 한 점포들간의 상대적 방향이 미치는 영향을 알아보기 위하여 3가지 평가지표를 설정하였다. 모든 점포의 순위정확도, 영향력이 큰 경쟁점포, 제곱오차의 평균(MSE)이 그 지표들이다. 얻어진 가능해의 수를 기준으로 할 때, 모든 점포의 순위정확도는 평균 31.1%, 영향력이 큰 경쟁점포 측면에서는 평균 28.3%의 추정 정확도를 보였다. 공간적 특성을 고려한 상대적 거리 및 방향에 대해서는, MSE를 기준으로 삼았을 경우, 경쟁점포들이 대상점포로부터 M 혹은 F의 거리만큼 떨어져 있을 때가, 그리고, 대상점포를 중심으로 경쟁점포들이 반대방향에 분포되어 있을 때가 추정의 정확도가 높았음을 보였다.

두 번째로, 다수의 경쟁점포를 대상으로 한 경우, 대상점포의 매출 분포에 보다 큰 영향력을 끼치는 경쟁점포만을 추려내기 위해 상위 3순위의 경쟁점포들만의 추정결과를 고려하기로 하였다. 그 결과, 경쟁점포들이 대상점포와의 거리옵션이 F인 경우, 추정 정확도가 높았다.

다음으로, 제안된 방법론을 실제 데이터에 적용하였다. 사용된 데이터는 일본 도쿄의 한 소규모 소매점포의 주문배달 매상데이터 및 상권 내 경쟁점포의 위치정보이다. 계산 결과, 다수의 경쟁점포들 중에서 대상점포의 매출분포에 큰 영향을 끼치는 경쟁점포의 추출이 가능하였다. 또한, 추정 결과치를 실제 데이터와 비교해본 결과, 추정결과로 추출된 경쟁점포가 대상점포의 매출분포가 거의 없는 곳에 위치함이 드러나, 상권의 실제상황을 잘 반영하고 있음이 나타났다. 그리고, 가상데이터를 이용한 시뮬레이션에서 모델 성능에 영향을 미치는 것으로 드러난 공간적 요소들 중, 상대거리는 실제 데이터의 계산에서도 경쟁점포가 멀리 떨어져 있음이 확인되었다. 하지만, 경쟁점포 간의 상대위치에 대한 영향은 상권구분기준에 따라 해석이 달라질 수 있기 때문에 설득력이 약한 것으로 나타났다.

감사의 글

본 논문을 심사해주신 심사자 분들께 본 지면을 빌어 감사의 뜻을 전합니다. 

참고 문헌

- 中谷 友樹, 花岡 和聖. 2003. 空間的マイクロシミュレーションを用いた小賣モデリング. 草津市消費者購買行動に関する調査研究. 第2回調査 2001-2003年度「GeoComputationによる施設配置計畫の支援システムの構築」. 研究報告書第3章:35-56.
- 矢野 桂司, 古賀 慎二, 中谷 友樹. 2002. 詳細な消費者購買行動データを用いたインパクト豫測モデルの構築. 共同報告. 2002年度人文地理學會大會. お茶の水女子大學.
- Applebaum, W. and S. B. Cohen. 1961. The Dynamics of Store Trading Areas and Market Equilibrium, *Annals of the Association of the American Geographers* 51:73-101.
- Ben-Akiva, M. and S. R. Lerman. 1985. *Discrete Choice Analysis*. The MIT Press:82-84, 103-108.
- Christaller, W. 1935. *Die Aentralen Orte in Sudlentschland*, Jena, E. Germany: G. Fischer.
- Davis, R. L. 1976. *Marketing Geography: with special reference to retailing*, Methuen & Co. Ltd.
- Fotheringham, A. S. 1983. A New Set of Spatial Interaction Models: The Theory of Competing Destinations, *Environmental and Planning a* 15:5-31.
- Huff, D. L. 1963. A Probabilistic Analysis of Consumer Spatial Behaviour, ed. Decker, W. S. *Emerging Concepts in Marketing*, Chicago: American Marketing Association:443-361.
- Kotler, P. 1991. *Marketing Management: Analysis, Planning, Implementation and Control*, 7th Edition, Prentice Hall.
- Lösch, A. 1954. *The Economics of Location*, New Haven, Conn.: Yale University Press.
- Luce, R. 1959. *Individual Choice Behavior*, New York: John Wiley & Sons.
- McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behaviour, *Frontiers in Econometrics*, Zarembka, P. eds. Academic Press, New York.
- Openshaw, S. 1995. *Marketing Spatial Analysis: A Review of Prospects and Technologies Relevant to Marketing*. Longley, P. and Clarke, G. eds. *GIS for Business and Service Planning*, *GeoInformation International*:150-165.
- Reilly, W. J. 1931. *The Law of Retail Gravitation*, New York, W. J. Reilly Inc.
- Snow, J. 1855. *On the Mode of Communication of Cholera*, London: John Churchill, New Burlington Street, England. 