

## 영향력을 고려한 적정입지선정 모델 연구

김병철<sup>1\*</sup>, 오상영<sup>2</sup>, 류근호<sup>3</sup>

### A Study of Optimum allocation model with influence

Byung-Chul Kim<sup>1\*</sup>, Sang-Young Oh<sup>2</sup> and Keun-Ho Ryu<sup>3</sup>

**요약** 본 논문은 영향력을 고려한 클러스터링을 위한 알고리즘인 DBSCAN-I를 제안한다.

DBSCAN-I는 기존의 DBSCAN과 DBSCAN-W를 확장한 것으로 공간데이터 클러스터링 시 여러 비공간 속성을 영향력으로 변환한다. 이는 영향력이 클수록 클러스터에 포함될 확률을 높여주기 위한 알고리즘이다. 또한 제안하는 알고리즘을 적용하여 영향력을 고려하여 효과적으로 적정입지를 선정하는 결과를 보였다.

**Abstract** In this paper, we propose DBSCAN-I that is an algorithm for clustering with influence.

DBSCAN-I that extends traditional DBSCAN and DBSCAN-W converts from non-spatial feature to influence while doing spatial clustering. This is an algorithm that increases probability of allocation to cluster when influence is more higher than other.

And also, we present the result that selects effectively optimum allocation with influence to apply the proposed algorithm.

**Key Words** : DBSCAN, clustering, 클러스터링, influence, 입지선정

## 1. 서론

정부 또는 기업에서 주요시설의 입지선정을 위해 GIS(geographical information system)의 다양한 공간자료(spatial Data)의 통합 방법을 많이 이용하고 있는 추세이다. 초기 입지선정을 위한 고려사항으로 A. Weber는 공업입지에 관한 이론에서 생산비와 관련한 요소들을 주요 항목으로 구분하였으나[1], Bunn & Wright는 많은 요인들을 복합적으로 고려하기 위하여 통계적 선형 모형의 결과를 이용한 인간의 혼합적 판단(Adjusted Judgement)과 통계적 선형모형의 활용 방법과 함께 상호비교가 많이 이루어져 왔다[2]. 실제로 정부의 관공서 및 출장소 또는 쓰레기 매립장 등을 새로 건립하고자 할 때, 또는 통신회사에서 기지국이나 중계기를 설치하고자 할 때, 그리고 기업이나 은행에서 지점이나 영업소를 선정하고자 할 때에 최적의 위치를 선정하기 위하여 이용자의 분포, 교통여건, 토지 및 건물의 가격 등과 같은 많은 항목을 복

합적으로 고려하여 최적의 위치를 선정하게 된다. 이런 문제에서 최적의 위치선정은 경제성, 합리성, 객관성 확보를 위하여 타당한 분석 절차와 결과를 요구한다.

본 논문에서는 이와 같이 적정입지 선정을 위한 공간 분석에서 각 객체가 갖는 영향력을 고려하기 위하여 DBSCAN-I 알고리즘을 제안한다. 제안한 클러스터링 기법은 공공시설의 입지 선정 문제, 상권분석 등에서 각 객체의 속성을 충분히 고려할 수 있고, 분석과정의 객관성과 결과에 대한 합리성을 제공할 수 있다. 논문의 구성은 다음과 같다. 2장에서는 이론적 배경으로 밀도를 기반으로 하는 공간 클러스터링 알고리즘인 DBSCAN과 DBSCAN-W를 소개하고, 3장에서는 DBSCAN-I에서 공간데이터 클러스터링 시 영향력을 고려하는 방법을 제안하며, 4장에서는 제안된 방법에 대한 적용 결과를 보여준다. 마지막으로 5장에서는 결론을 제시한다.

## 2. 이론적 배경

### 2.1 공간데이터(spatial data) 일반화

공간데이터에 대한 연구는 1990년대 이후 활발히 논

<sup>1</sup>남서울대학교 지리정보공학과

<sup>2</sup>청주대학교 경상학부

<sup>3</sup>충북대학교 전자계산학과

\*교신저자: 김병철(bckim@nsu.ac.kr)

의되고 있다. 대표적인 공간 데이터베이스에 관한 소개는 Maguire, Goodchild & Rhind[3], Guting[4], Egenhofer[5]에서 찾아 볼 수 있다. 특히 Lu, Han & Ooi은 속성 기반 유도를 이용한 일반화 기반 공간 데이터 마이닝 방식을 제안했다[6]. Ester, Kriegel & Sander는 지리 공간 데이터 마이닝을 위한 공간 데이터 마이닝 방법에 대한 포괄적인 연구 결과를 발표하였다[7]. 또한 Miller & Han은 지리 데이터 마이닝과 지식 발견에 대한 연구를 하였다[8].

### 2.2 밀도기반(density-based) 알고리즘

밀도 기반 군집화 방식인 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)은 Ester, Kriegel, Sander과 Xu에 의해서 제안되었다[9]. DBSCAN은 데이터 집합이 주어지면 클러스터와 어느 클러스터에도 속하지 않은 데이터 점들인 잡음을 식별한다. 여기서 클러스터들을 인식할 수 있는 중요한 이유는, 각 클러스터 내 데이터 점들의 밀도는 클러스터의 외부에서 보다 상당히 높다는 것이다. 더욱이 잡음 영역에서의 밀도는 어떤 다른 클러스터 내에서보다 밀도가 낮다. 또한 김호숙 외[10]에 의해 기존의 DBSCAN을 확장한 DBSCAN-W(a DBSCAN algorithm using region expressed as Weight)는 DBSCAN이 모든 객체들은 위치 속성만을 갖는 점으로 표현되고 각 점들이 갖는 중요도는 고려되지 못하는 것을 보완하기 위한 것으로 DBSCAN-W는 밀도 기반 공간 클러스터링 알고리즘에 가중치를 고려하기 위해서 각 대상물의 위치 속성 뿐 아니라 응용 시스템의 분석 시 필요한 비공간 속성을 클러스터링 시 고려한다. 그러나 객체가 갖는 가중치가 아무리 커더라도 주변에 있는 이웃의 수가 적으면 소용이 없게 된다. 따라서 본 연구에서는 DBSCAN-I알고리즘을 이용하여 가중값이 큰 객체들의 경우 주변 이웃의 수가 비록 적더라도 클러스터에 포함될 수 있도록 하는 방안을 제시한다.

## 3. 영향력을 고려한 밀도기반의 공간클러스터링 알고리즘의 설계

### 3.1 DBSCAN-I 아이디어

DBSCAN-I(a DBSCAN algorithm using region expressed as influence)는 밀도기반 클러스터링 알고리즘으로 객체가 갖는 여러 속성 들을 영향력 값으로 변환하여, 어느 한 점이 이웃(neighborhood)의 수에 의해서만 중심 객체(core object) 여부가 결정되는 DBSCAN과

DBSCAN-W 알고리즘을 보완하여 이웃 객체의 수는 적더라도 이웃들의 영향력의 합이 인자로 주어지는 최소 영향력 이상이라면 그 점을 중심 객체가 되게 한다. DBSCAN-I에서는 다음의 사항을 정의한다.

**[정의 1]** 모든 객체는 해당 응용시스템에서 그 객체가 갖는 영향력에 따라서 서로 다른 크기로 표현되는 영역을 갖는다.

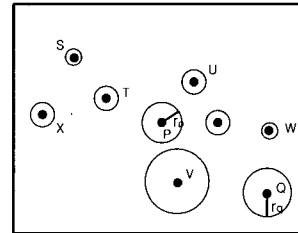


그림 1. 영향력(influence) 반영

이때 원의 중심 좌표값은 공간상의 객체의 위치 좌표이고, 반지름은 영향력 값으로, 하나 또는 여러 속성값에 의해 계산된다.

**[정의 2]** MinPts는 최소이웃수(minimum number of points)를 의미하고, MinWt(minimum sum of Weights)는 최소이웃의 영향력의 합을 의미한다. 이때 한 점의 Eps'-neighborhood의 수가 MinPts 이상인 경우이거나, 최소이웃의 영향력의 합이 MinWt 이상인 경우 이점을 core object라 한다. 여기에 자신의 영향력 값도 포함한다.

[그림 2]에서 (a)는 점 p를 기준으로 Eps 반경내의 이웃 객체의 수가 4이고, 이웃들의 영향력의 합은 자신을 포함하여 18이다. 그리고 (b)의 경우는 이웃 객체의 수가 3이고, 이웃들의 영향력의 합은 자신을 포함하여 32이다.

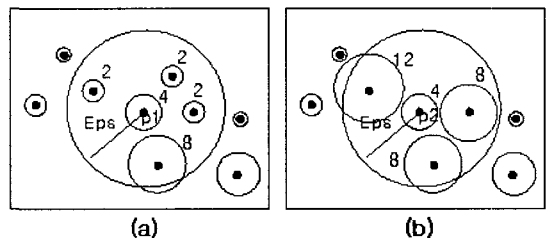


그림 2. DBSCAN-I에서의 core object

여기에서 중심 객체의 조건으로 이웃의 수가 4이상이거나 이웃의 영향력 값의 합이 30 이상인 경우로 설정하

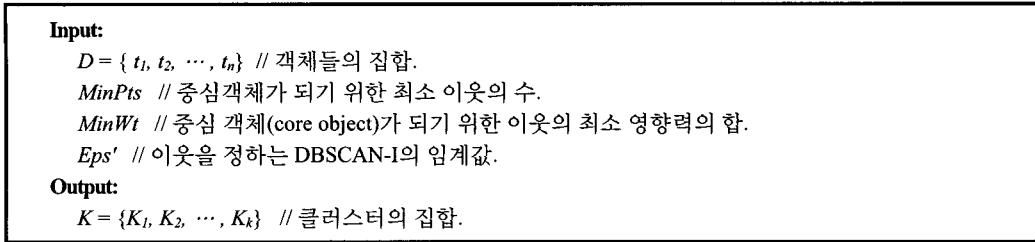


그림 3. DBSCAN-I 입출력 파라미터

게 되면 [그림 2]의 (b)도 중심객체가 된다. 이는 주변에 객체의 수는 비록 적더라도, 객체들이 비중 있는(영향력 값이 큰) 경우라면 중심객체가 되게 하여 클러스터에 포함되도록 하는 효과가 있다.

### 3.2 DBSCAN-I 알고리즘

DBSCAN-I의 클러스터링을 위한 알고리즘의 입출력 파라미터는 [그림 3]과 같다. DBSCAN-I 알고리즘에서는 입력 파라미터가 기존의 DBSCAN과 DBSCAN-W보다  $MinWt$  하나가 더 추가된다.

## 4. 적정입지선정 이론의 적용

### 4.1 알고리즘 적용을 위한 데이터의 전처리

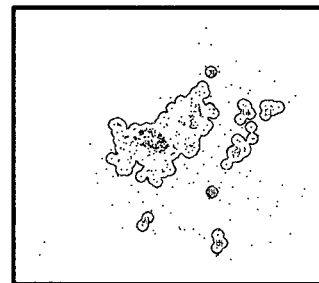
제안된 클러스터링 알고리즘인 DBSCAN-I를 이용하여 청주시에 소재한 제조업체 412개 기업에 대한 데이터를 분석하였다. 연구를 위하여 청주시내 제조업체의 종업원을 위한 스포츠 시설을 제공하기 위하여 적합한 위치를 선정하고자 하는 상황을 가정한다. 그러나 많은 기업이 공단 주변에 밀집되어 있어 이 지역은 제외하고, 그 외의 지역에서 수혜를 받을 수 있는 제조업체 종업원이 많은 지역 5곳을 선정하고자 한다. 수혜자는 선정 지역으로부터 일정 거리 안에 있는 기업 및 종업원으로 한정한다.

이 분석에 사용되는 데이터는 청주시내 기업으로 한정하고, 기업의 주소 정보를 이용한 지도상의 좌표정보와 종업원 수로 구성되었다. 여기에서 공간 데이터는 좌표상의 위치를 나타내는 x, y값이고, 비공간 속성으로는 종업원의 수이다. 또한 각 동별로 제조업체의 수 및 종업원 수의 합계와 평균을 구해본 결과, 상위 5개 동에 전체 기업의 62.8%가 밀집되어 있고, 전체 종업원 수의 91.9%가 포함되어 있다. 이는 공단 지역에 기업체가 밀집되어 있어서 어떤 알고리즘에 의해서 클러스터링 하더라도 밀집도가 상당히 높아 중심객체가

되기 위한  $MinPts$ 와  $MinWt$ 의 수준을 훨씬 초과하므로 차별성을 발견할 수 없는 상황이다. 따라서 이 문제에서는 공단 지역은 대부분의 조건에서 클러스터에 속하게 되므로 공단 이외의 지역에 대한 클러스터링에 관심을 가져야 하는 경우에 해당한다. 그러므로 밀집지역을 제외하고 기타로 분리된 나머지 37.2%에 해당하는 153개 기업에 관한 문제로 재정의 한다.

### 4.3 알고리즘 적용 및 결과

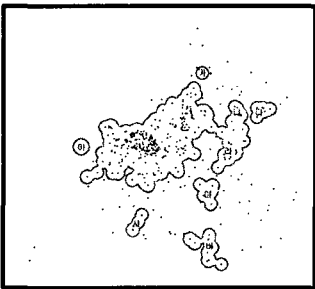
클러스터링 결과를 나타난 [그림 4]는 각 알고리즘 별 수행 결과로 그림 (a)와 그림 (b)는  $Eps=20$ ,  $MinPts=4$ 에서 수행되었다. 그리고 [그림 4]의 (c)는  $Eps=20$ ,  $MinPts=4$ ,  $MinWt=70$ 에서 수행되었다. 다소 크기의 차이가 있지만 세 알고리즘 모두 공단 지역은 커다란 클러스터를 구성하고 있음을 볼 수 있다. 그러면 그 나머지 지역에서 형성된 클러스터에 대해 특징을 정리해 본다. [그림 4]의 (a) DBSCAN은 7개 [그림 4]의 (b) DBSCAN-W는 6개의 클러스터를 발견하였다. 그리고 [그림 4]의 (c) DBSCAN-I는 공단 인접 지역이 공단 클러스터에 포함되었으나, 일정 거리가 있으므로 이 부분도 별도의 클러스터로 보아 7개의 클러스터로 처리한다.



(a) DBSCAN



(b) DBSCAN-W



(c) DBSCAN-I

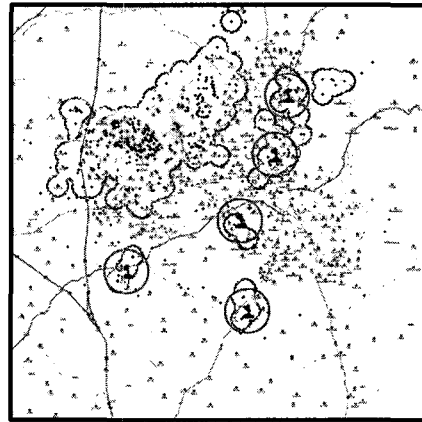
그림 4. 클러스터링 결과

이 문제는 보다 많은 기업체 종업원이 스포츠 시설을 이용할 수 있도록 하기 위한 것과 같은 입지 선정에 있다. 따라서 클러스터 영역 중 일정 범위 안에 최대한 많은 기업체 종업원이 있는 곳이 중요한 위치가 된다. [그림 4]에 나타난 클러스터 영역중 가, 나, 다, ..., 아로 표시한 곳에서 일정 반경 내 기업체의 종업원 수가 최대가 되었고, 이를 [표 1]에 요약하였다.

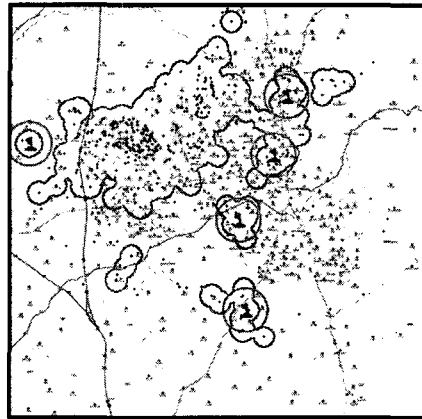
표 1. 영향권 내 기업 및 종업원 수 현황

위치 구분	가	나	다	라	마	바	사	아
기업체수	6	6	6	11	6	8	7	1
종업원수	18	42	10	20	77	52	19	54

따라서 영역 내 종업원 수를 기준으로 최대 수계가 가능한 지역 5개를 선택하자면 DBSCAN과 DBSCAN-W의 결과로는 마(77명), 바(52명), 나(42명), 라(20명), 사(19명) 위치를 선정하게 된다. 그러나 DBSCAN-I의 결과로는 마(77명), 아(54명), 바(52명), 나(42명), 라(20명) 위치를 선정하게 될 것으로 보다 많은 사람들에게 영향을 줄 수 있는 곳을 클러스터로 도출하였다. 이렇게 결정된 위치를 지도상에 나타내면 [그림 5]와 같다.



(a) DBSCAN, DBSCAN-W결과



(b) DBSCAN-I결과

그림 5. 입지 선정 클러스터링 결과

DBSCAN-I는 주변 객체의 수인 MinPts 뿐만 아니라 주변 영향력의 합인 MinWt를 함께 고려하고, 두 입력 파라미터를 조절함으로써 여러 형태의 다양한 분석이 가능해 진다.

## 5. 결론

공간클러스터링(spatial clustering)은 공간 데이터베이스 내에서 객체간의 거리, 연결성, 밀도를 기반으로 유사한 객체 들을 그룹화하는 것이다. 밀도기반 공간 클러스터링 알고리즘에 가중치를 고려하기 위해서 기존의 DBSCAN을 확장한 DBSCAN-W도 제안되었다. 본 논문에서는 영향력을 고려한 클러스터링을 위한 알고리즘인 DBSCAN-I를 제안하였다. DBSCAN-I는 기존의 DBSCAN과 DBSCAN-W를 확장한 것으로 공간클러

스터링 시 여러 비공간 속성을 영향력으로 변환하여 고려되는 속성에 의한 영향력이 클수록 클러스터에 포함될 확률을 높여주기 위한 알고리즘이다.

그리고 영향력을 고려한 클러스터링기법을 이용하여 적정입지 선정모델에 적용하였다. DBSCAN-I는 주변 객체의 수인 MinPts 뿐만 아니라 주변 영향력의 합인 MinWt를 함께 고려하고, 두 입력 파라미터를 조절함으로써 여러 형태의 다양한 분석이 가능해 진다. 분석 결과 기존의 알고리즘인 DBSCAN과 DBSCAN-W는 기업의 밀도에 크게 영향을 받은 반면, DBSCAN-I는 기업의 밀도 뿐만이 아니라 주요 변수인 종업원 수도 고려하여 클러스터를 발견함으로써 많은 종업원들이 고려되는 적정입지를 선정할 수 있었다.

DBSCAN-I 알고리즘은 주변에 있는 객체들이 특정 속성(feature) 중심으로 보았을 때, 영향력(influence)이 큰 객체임에도 불구하고 주변에 객체가 드물게 있다는 이유로, 또는 속성(feature)이 갖는 특성 값을 제대로 반영하지 못하여 클러스터에서 배제되게 되는 기존 알고리즘의 단점을 보완해 줄 수 있다.

본 논문에서 제안한 클러스터링 기법은 다차원 속성이 개입하는 복잡한 문제에서 각 속성들이 갖는 영향력을 고려하여 종합적인 분석이 가능하고, 또한 중요한 일부의 부분 속성들에 의해 선택적인 분석이 가능하며 공공시설의 입지 선정 문제, 상권분석 등에서 각 객체의 속성을 충분히 고려할 수 있고, 분석과정의 객관성과 결과에 대한 합리성을 제공할 수 있다.

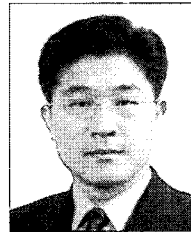
### 참고문헌

[1] A. Weber, "Über den Standort der Industrien", 1909.  
 [2] D. G. Bunn, Wright, "Interaction of judgemental and statistical forecasting methods: Issues and Analysis", Management Science, Vol. 37, pp. 510-518, 1991.  
 [3] D. J. Maguire, M. F. Goodchild, and D. W. Rhind, "Geographic Information systems", Principles and applications, Vol. 1, Longman Scientific & Technical, New York, 1992.  
 [4] R. H. Guting, "An introduction to spatial database systems", The VLDB Journal, 3, pp. 357-400, 1994.  
 [5] M. J. Egenhofer, "Spatial Query Languages", UMI Research Press, University of Maine, 1989.  
 [6] W. Lu, J. Han, and B. C. Ooi, "Knowledge discovery in large spatial databases", In Proc. Far East Workshop Geographic Information Systems, pp. 275-289, June 1993.

[7] M. Ester, H. P. Kriegel, and J. Sander, "Spatial data mining: A database approach", In Proc. Int. Symp. Large Spatial Databases(SSD'97), pp. 47-66, July 1997.  
 [8] H. J. Miller, and J. Han, "Discovering Geographic knowledge in Data Rich Environment: A Report on a Specialist Meeting", SIGKDD Explorations Vol. 1 No. 2, pp. 105-107, 2000.  
 [9] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Database with Noise", Proc. of Int. Conf. on Knowledge Discovery and Data Mining, 1996.  
 [10] 김호숙, 임현숙, 용환승, "공간 데이터 마이닝에서 가중치를 고려한 클러스터링 알고리즘의 설계와 구현", 한국지능정보시스템학회 논문지, pp. 101-103, 2002.

### 김 병 철(Byung-Cheol Kim)

[정회원]



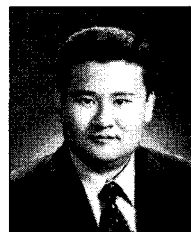
- 1992년 2월 : 청주대학교 응용통계학과 (경제학사)
- 1997년 8월 : 청주대학교 전자계산학과 (공학석사)
- 2005년 8월 : 충북대학교 전자계산학과 (이학박사)
- 2006년 8월 ~ 현재 : 남서울대학교 지리정보공학과 겸임교수

<관심분야>

Data Mining, GIS, 시공간데이터베이스

### 오 상 영(Sang-Young Oh)

[정회원]



- 1992년 2월 : 청주대학교 응용통계학과 (경제학사)
- 1996년 8월 : 청주대학교 경영학과 (경영학석사)
- 2001년 2월 : 충북대학교 경영학과 (경영학박사)
- 2002년 3월 ~ 현재 : 청주대학교 경영학부 교수

<관심분야>

KMS, 혁신이론, System Thinking

류 근 호(Kem-Ho Ryu)

[정회원]



- 1976년 송실대학교 전산학과(이  
학사)
- 1980년 연세대학교 산업대학원  
전산전공(공학석사)
- 1988년 연세대학교 대학원 전산  
전공(공학박사)
- 1986년 ~ 현재: 충북대학교 전  
기전자컴퓨터공학부 교수

<관심분야>

시공간 데이터베이스, GIS 및 지식기반 정보검색 시스템,  
데이터 마이닝, 데이터베이스 보안, 바이오 인포메틱스