空間分析을 위한 퍼지분류의 理論的 背景과 適用에 관한 硏究

—慶尙南道 邑級以上 都市의 機能分類를 中心으로—

鄭 仁 喆*

본 연구는 퍼지이론을 공간분석에 적용하기 위한 이론적인 배경을 고찰하고, 퍼지분류법의 특성에 대해 살펴본 것이다. 이를 위해 필자는 공간정보의 모호성에 대해 살펴보고, 퍼지공간분석의 전제를 설정한 다음 퍼지분류법을 소개하였다. 그리고 퍼지분류법의 특성을 명확히하기 위해 경상남도 읍급이상 도시의 산업별 고용비율을 대상으로 퍼지분류를 행한 후, 퍼지분류와 전통적인 군집분석의 결과를 비교하였다. 그 결과, 공간정보의 모호성은 구체성의 부족, 인간행태, 인내치문제, 분류기준의 부족 등에 의해 발생하는데 기존의 공간분석기법으로는 공간의 모호성을 반영할 수 없으므로 퍼지기법을 도입한 퍼지공간분석의 필요성이 있음을 확인하였다. 퍼지분류법 중, 퍼지이산분류는 계산절차는 상대적으로 간단하나 분류결과가 집단간의 점이성을 고려하지 못하며, 퍼지중첩분류는 분류집단간의 점이성은 고려하나 분류결과가 지나치게 많아 적절한 분류수준을 선택하기 어렵고 결과해석이 상대적으로 난해하다는 문제점이 있음이 밝혀졌다. 또 경남의 도시기능분류는 분류기법에 따라 다르게 이루어졌지만 창원, 울산, 마산, 진해, 김해, 양산, 웅상, 장승포, 신현으로 구성된 제조업 군집과 단독군집 충무의 존재가 세 가지 분류 모두에서 공통적으로 확인되었다.

主要語：空間分析, 模糊性, 퍼지分類, 漸移性, 都市機能

1. 서 론

1) 연구목적

주어진 속성에 따라 공간단위(spatial unit)를 분류하는 것은 공간을 구성하는 요소들의 전체와 부분간의 관계를 알기 위한 것으로 공간분석의 한 형태로 자리잡고 있다(Racine and Reymond, 1973; Guttenberg, 1993). 이러한 연구의 대표적 형태로는 지역구분과 도시기능분류를 들 수 있다. 그러나 기존의 분류방식에 의하면 주어진 임계치를 기준으로 하나의 공간단위가 하나의 분류군에 속하느냐 속하지 않느냐는 이분법적인 결과만 알 수 있으며, 하나의 분류군에는 어느 정도 속하고 또 다른 분류군에는 어느 정도로 속하는 지에 대한 정보는 유실된다(Leung, 1984). 이러한 문제는 기존의 모든 분류법의 한계로 지적될 수 있다. 예를 들어 Harris의 도시기능분류에서 제조업인구가 제조업·소매업·도매업 총취업인구의 74%인 도시는 제조업도시 M′에 분류되나, 73%의 도시는 제조업도시 M으로 분류된다(홍경희, 1987). 즉 단 1%의 차이에 의해 다른 분류군에 속하게 되는 것이다. 그러므로 임계치를 기준으로 한 이분법적 분류로서는 정보의 손실을 피할 수가 없게 된다. 그리고 이분법적인 분류의 또 다른 단점은 임계치설정을 위한 객관적 기준의 설정이 매우 어렵다는 것이다. 하나의 지역이 다른 지역과 명확히 구분될 때도 있지만 이러한 경우는 현실적으로 매우 드문 실정인데, 이러한 상황에서 공간단위를 이분법적으로 구분하는 경계선을 설정하는 것은 임의적이라 할 수 있다. 그러므로 정보의 손실을

* 慶南開發研究院 交通·環境研究室長

최소화하면서도 경계선 설정의 임의성을 배재하기 위해서는 하나의 임계치를 기준으로 한 명확한 선적 경계 대신에 점이지대를 인정하고 고려하여야 한다. 이러한 관점에서 모호한 경계선 문제를 다루는 퍼지이론을 채택한 퍼지분류기법을 공간분석에 도입할 필요가 있다.

퍼지이론은 1962년[1] Zadeh에 의해 개발된 모호성(vagueness)을[2] 다루는 과학으로, 퍼지집합, 퍼지논리, 퍼지측정의 이론체계로 구성되어 있다. 이 퍼지이론을 공간단위의 분류에 적용한 최초의 도시는 프랑스 Dijon 공간경제학파의 Tranqui(1978)에 의해 이루어졌는데, 그는 퍼지분류기법을 개발하여 프랑스의 경제지역을 다양한 수준에서 구분하였다. 퍼지분류법은 명확히 수치화된 자료만 처리하는 것이 아니라 '좋다', '나쁘다'와 같은 언어변수, 또는 '약 10만원 정도'와 같은 모호한 자료도 처리가능하게 한다. 그리고 무엇보다도 공간구분의 결과가 임계치의 변화에 따라 변화하는 과정을 정확하게 추적하는 것을 가능하게 한다는데 그 이점이 있다.

그러나 퍼지이론의 이러한 장점에도 불구하고, 퍼지이론을 공간분석에 도입하기 위한 체계적인 연구는 아족 이루어지지 않고 있으며, 퍼지분류법을 이용한 공간단위의 분류가 기존의 분류방법과 어떠한 차이점을 갖고 있는지에 대한 이론적 연구 또한 Gale and Atkinson(1979)의 연구를 제외하고는 전혀 이루어지지 않는 실정이다.

이러한 관점에서 본 연구에서는 퍼지이론을 공간분석에 적용하기 위한 이론적 배경을 고찰하고, 퍼지분류법의 특성에 살펴보고자 한다. 그리고 경상남도 28개 읍급이상 도시의(그림 1)[3] 산업별 고용비율을 대상으로 퍼지분류를 행한 후, 퍼지분류법의 결과와 전통적 분류법의 결과를 비교하여 보고자 한다.

2) 연구방법 및 자료

퍼지이론을 이용한 공간분석의 이론적 배경을 고찰하기 위해서는 먼저 공간정보가 모호성을 갖게 되는 원인에 대해 살펴보아야 한다. 그리

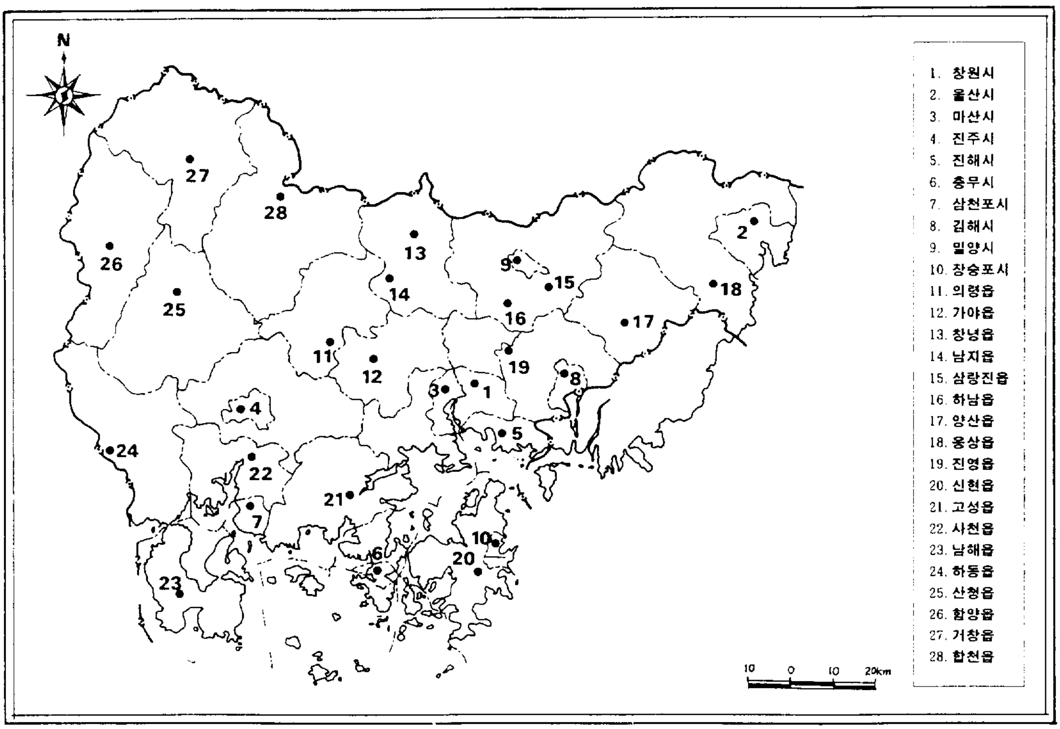

그림 1. 경남의 읍급 및 시급 도시(1994년)

고 이러한 모호성을 취급하는 기법으로서의 퍼지이론에 대한 일반적인 논의가 필요하며, 이러한 기법의 적용이 어떠한 지리학적 의미를 갖는지에 대한 이론적인 연구가 필요하다. 이러한 관점에서 먼저 지리정보의 모호성에 대해 논하였다. 그리고 퍼지공간분석의 전제를 설정한 다음, 귀속함수를 이용하여 모호한 자료를 처리하는 방법을 소개하였다. 또 퍼지이론 중 퍼지공간분석 특히 퍼지분류에 응용되는 퍼지집합이론을 요약하고, 퍼지분류법의 절차를 소개하였다. 이 퍼지분류법을 이용하여 경상남도의 산업별 고용비율을 기준으로 한 도시유형화를 시도하였는데, 도시의 기능적 분류에 관한 문헌연구는 이미 국내에서 많이 이루어진 상태이므로[4] 생략하였다.

본 연구의 사례지역인 경상남도에는 1995년의 전면적인 행정구역개편으로 현재 10시 23읍이 존재하고 있다. 그러나 본 지역의 산업구조에 대한 통계자료취득이 현재로서는 불가능하므로 행정구역변경 이전인 1994년의 10시 18읍에 대한 통계자료를 이용하여 사례분석하기로 한다. 地方行政階層構造上 道-市로 이어지는 二層體制下에 있는 市와 郡의 단순한 하부행정구역으로 道-郡-邑의 三層體制下에 있는 邑을 함께 분류하는 것은 분류결과를 해석하는데 다소의 문제점을 야기하겠지만 경상남도 지역계획을 기초자료로 활용이 가능할 것이다.

산업별 고용비율은 한국도시연감(1994)의 '1992 한국 표준산업분류'에 의한 17개 산업 중 해당지역에 사업체가 하나도 없는 가사서비스업, 국제・기타 외국기관의 2개 부문을 제외한 ① 농업・수렵・관련 서비스업, ② 어업, ③ 광업, ④ 제조업, ⑤ 전기・가스・수도사업, ⑥ 건설업, ⑦ 도・소매 소비자 용품, ⑧ 숙박・음식점업, ⑨ 운수・창고・통신업, ⑩ 금융 및 보험업, ⑪ 부동산 임대・사업서비스업, ⑫ 공공 행정・국방・사회보장업, ⑬ 교육 서비스업, ⑭ 보건・사회복지사업 ⑮ 기타 공공사회・개인업의 15개 산업별 종사자수를 총 산업 종사자수에 대한 비율로 나누어 구하였다.

이상의 자료를 이용하여 퍼지분류한 결과가 전통적 분류법에 의한 결과와 어떠한 차이점이 있는지를 알아보기 위해 동일한 자료로 전통적 분류법 중 현재 가장 많이 사용되고 있는 군집분석을 이용하여 군집화하였다. 퍼지분류 프로그램은 필자가 작성하였으며 군집분석을 위해서는 SAS와 SPSS의 군집분석(Ward)법 프로그램을 사용하였다. 분류결과에 대해서는 상세한 원인분석은 피하고 간략히 해석하는 수준에 머물렀는데, 이는 시군통합으로 인한 행정구역변경으로 정밀한 분석의 필요성이 적기 때문이다.

2. 지리정보의 模糊性과 퍼지공간분석

1) 지리정보의 모호성

Dubois and Prade(1985)는 정보의 모호성을 논하면서, '모호한 정보(fuzzy information)'를 '부정확하고 불확실한 정보'라고 정의하였다. 이들이 말하는 부정확한 정보란 '일인당 주민소득이 약 100~200만원 정도'의 경우와 같이 정확하지 못한 값을 가진 정보이며, 불확실한 정보란 '가까운 거리'의 경우에서 보듯, 어느 정도의 거리가 가까운지에 대한 확신의 정도가 문제가 되는 정보이다. 모호성의 원인에 관한 철학자들의 설명에 의하면[5] 모호한 정보는 구체성이 부족하고 경계선 설정이 어렵고, 경계선으로 인해 정보의 손실이 야기되며, 정보관측자의 주관이 개재되는 속성을 가지는 정보이다(Rolf, 1980b). 이 가운데 구체성의 부족, 경계선 설정에 의한 정보손실은 정보의 부정확성과 관계가 있고, 정보관측자의 주관성은 정보의 불확실성을 의미한다고 말할 수 있다. 본고에서는 모호한 자리정보를 '부정확하고 불확실한 자리정보'로 정의하기로 한다.

이제 모호한 지리정보의 유형에 대해 살펴보기로 하자. 첫째로, 공간의 특성을 규정하는 敍述語가 부정확한 경우이다. 크다, 작다, 조금 작다, 약간 멀다 등의 인간이 日常的으로 사용하는 自然語에 의한 공간특성의 서술은 관측자가 판단하는 상황과 주관에 의존하게 되어 매우 모호하므로, 과학적 분석을 위해서는 부적절하다. 이러한 언어로 인한 공간표현의 주관성은

결국 판단기준의 모호성을 초래한다. 둘째로 자료의 부족 또는 관측자의 공간에 대한 지식의 부족으로 인해 공간의 특성을 정확하고 확실하게 기술하지 못하는 경우이다. 대표적 예로는 '석유매장 가능지역'을 들 수 있다. 그러나 이러한 모호성은 기술의 발달, 지식의 증가, 현장답사 등을 통해 감소된다. 셋째로, 認識의 主觀性으로 인해 모호성이 야기되는 경우가 있다. 認知空間(cognitive space)은 인간 인지의 필터가 인식을 제한하고 변형시킨 부정확하고 불확실한 공간이다. 넷째로, 기술하고자 하는 지리정보 자체가 모호하게 정의되는 경우이다. 대표적인 예로 機能地域의 경계를 들 수 있다. 기능중심부의 영향이 주변부에 균질하지 않기 때문에, 주변부는 항상 다소간 부정확하게 정의되며, 주변부의 실정에 대해 확신하기 어렵다.

부정확성과 불확실성의 문제 중 불확실성은 기존의 통계적 방법에 의해서도 어느 정도의 해소는 가능하다. 예를 들어 하나의 공간모형이 현실을 어느 정도 설명하느냐는 확실성의 문제는 상관계수 및 결정계수를 통하여 설명이 가능하다. 그리고 주어진 가설을 모수검정이나 비모수검정을 통하여 확인하는 통계추론에 의하여도 불확실성의 확인은 가능하다. 그러나 어차피 최초에 선택되는 표본들은 정확하고 또 확실한 것들이라고 전제되므로, 이러한 통계적 방법이 불확실성을 정교하게 취급하는 것은 아니다. 그리고 이러한 확률론적[6] 측면의 불확실성이란 어떤 사건이 일어날지 안 일어날지 모르는 불확실성이다. 이 불확실성 속에서 하나의 사건이 일어날 경우를 수량적으로 표현해 보는 것이 확률인데, 확률은 상대적으로 많은 관측에 의존한다. 즉 확률은 객관적인 사항의 관계이다. 그러나 퍼지이론의 불확실성은 주관성을 포함한다 (Rolland-May, 1987a).

이러한 관점에서 모호성을 다루는 퍼지이론을 공간분석에 도입할 필요가 제기된다. 퍼지공간분석은 부정확하고 불확실한 지리정보를 계측하여, 공간분석에 이용할 수 있게 한다. 그러면 이제 퍼지이론을 이용한 퍼지공간분석이 기존의 공간분석과 어떠한 차이점이 있는지 공간분석의 전제를 통하여 살펴보자.

2) 퍼지공간분석의 전제

퍼지공간분석법 역시 기존의 공간분석법처럼 공간의 규칙성과 구조를 파악하고 이로부터 공간조직 및 그 조직성에 영향을 미치는 변수들을 연구한다. 퍼지집합이 보통집합을 포함하므로 퍼지공간분석법 역시 기존의 공간분석법을 포함하고 있다고 말할 수 있다. 즉 전통적 공간분석은 퍼지공간분석의 특별한 경우로 볼 수 있는 것이다. 근본적인 차이는 공간분석의 전제가 다르다는 것인데, 퍼지공간분석의 전제조건은 다음과 같다(Fustier, 1979; Ponsard et al., 1988; Rolland-May, 1987b).

전제 1 : 퍼지공간은 주어진 장소들의 퍼지부분집합[7]이다.

장소의 집합 S는 적어도 2개 이상의 장소로 이루어진 집합이다. 지표 위에서 일어나는 모든 자연 또는 인문현상은 이 장소와 결합되어 일어난다. 이 장소는 바로 공간을 지리적으로 규정하는데 가장 중요한 성분이다. 장소의 집합이란 연구 대상지역을 의미하며, 대상지역은 장소 1, 2, 3, …, n의 집합 {1, 2, 3, …, n}으로 표시된다. 그리고 이 집합의 퍼지부분집합은 장소의 특성을 나타내는데 사용되고 퍼지공간을 형성한다.

전제 2 : 표준공간(referential space)은 보통집합으로 표시되나, 이 공간을 구성하는 요소들의 특성은 퍼지부분집합으로 표시된다.

표준공간이란 최초에 기준으로 정해둔 공간단위들의 집합을 의미한다. 예를 들어 1994년도 경상남도 28개 시급 및 읍급도시를 기능별로 분류한다고 하면 이 28개 도시의 전체집합이 표준공간에 해당한다. 이 경우 표준공간은 명확히 규정되나, 분류된 도시들이 각 기능집단에 속하는 소속도는 모호하다.

전제 3 : 퍼지공간에서 장소간의 거리는 모호하게 측정된다.

장소들 간의 관계는 거리에 의해 표시된다. 연구의 목적에 따라 거리의 측정 방법은 다양하게 결정된다. 직선거리, 도로거리, 시간거리, 비용거리, 사회적 거리, 심리적 거리 등이 단독

으로 또는 결합되어 공간분석을 위하여 많이 사용되어 왔다. 이 가운데 직선거리, 도로거리, 시간거리 등 객관적인 거리는 관측자에 상관없이 주어진 상황에서 일정하게 측정되는 거리이다. 즉 명확하고 확실하게 측정되는 거리이다. 그러나 심리적 거리는 사람에 따라 또 그 사람이 처한 상황에 따라 다르게 측정되는 부정확하고 불확실한 거리이다. 퍼지공간은 이러한 퍼지거리에 의해 측정될 수 있다. 그러므로 퍼지거리는 거리함수(metric)에 한정되지 않는다.

전제 4 : 퍼지공간의 형태는 모호하게 측정된다.

거리가 모호하기 때문에 공간의 면적 또한 모호한 것은 당연하다. 공간의 형태를 C라 하면 윤곽(configuration) 바깥의 부분은 윤곽의 보집합 $2C^C$에 해당한다. 표준공간을 U라 하면 C와 C^C의 교집합은 공집합이 아니며($C \cap C^C \neq \phi$), 이 둘의 합집합은 표준공간 U가 아니다($C \cup C^C \neq U$).

전제 5 : 퍼지공간의 속성은 모호하게 측정된다.

퍼지 공간의 속성들은 모호하게 측정되는 것이 용납된다. 예를 들어 공업화 비율을 측정하며 '약 5%', '30% 조금 넘게' 등으로 규정할 수 있다. 이렇게 모호한 측정치는 귀속함수를 사용하여 수치화한다.

전제 6 : 퍼지공간은 공간의 연속성과 등질성을 전제하지 않는다.

전통적인 공간분석법은 공간의 연속성과 등질성을 전제하여 왔으며, 최근의 공간분석기법 역시 이 전제를 채택하고 있는 경우가 많다. 예를 들어 지형분석을 위해 최근에 많이 사용되고 있는 불규칙 삼각망 역시 공간연속성을 바탕으로 한 것이다(Goodchild, 1992). 그러나 퍼지공간분석은 이러한 전제를 필요로 하지 않는다.

이상에서 퍼지이론을 이용하여 공간을 분석하기 필요한 이론적 배경을 살펴보았다. 이제 모호한 자료를 처리하여 계량화하는 방법에 대해 살펴보자.

3) 모호한 자료의 처리

모호한 자료의 처리는 귀속함수의 설정을 통해 이루어진다. 지리정보가 0.5나 0.7과 같은 구체적 수치가 아닌 '좋다', '나쁘다', '만족하다' 등과 같이 언어로 표현되거나 '50%에 가까운 값'과 같은 부정확한 자료는 귀속함수를 이용해 계량화한다. 그러나 이러한 언어나 부정확한 값의 처리는 연구자가 최초에 귀속함수값을 수치로 미리 정의해 두어야 처리 가능하다. 예를 들어 '만족도'의 관점에서 '매우 만족'은 귀속도 1, '조금 만족'은 귀속도 0.3, '매우 불만족'은 귀속도 0 등이 사전에 정의되어야 된다. 그러나 현실적으로 귀속도를 정할 자료의 수가 상당히 많을 경우에는 일일이 상대적인 귀속도를 사전에 할당하는 것이 거의 불가능하다. 그래서 연구자가 귀속도를 할당하는 대신, 귀속함수를 정하게 된다. 예를 들어 중공업부분에 종사하는 생산인구비율이 50%면 중공업지역에 속하는 귀속도가 얼마, 80%면 귀속도가 얼마라고 지적하는 대신, 10%이하면 0, 90% 이상이면 1, 그리고 10~90% 사이에서의 귀속함수를 $(x-10)/80$이라 정할 수 있다. 그러나 귀속함수를 만들기 위해서는 귀속함수치 0, 0.1, 0.2, …, 1 등에 대응하는 관측 값을 할당하여 이로부터 함수를 추출해야 함으로, 관측 값이 많을 시 상당히 복잡한 작업이 된다. 그리고 귀속함수는 개인의 주관적 판단에 따라 다르게 설정될 수 있다. 그러므로 귀속함수를 만드는 과정은 매우 주관적이면서도 복잡한 작업이라 할 수 있다. 귀속함수 설정을 위한 명확한 이론체계는 아직까지 정립되지 않고 있으며, 이에 대한 경험적인 연구도 부족한 실정이다(Zimmermann, 1991). 이러한 문제점으로 인해 현재까지는 퍼지이론이 모호한 자료를 완벽히 처리한다고 볼 수 없다. 그러나 귀속함수를 간단히 작성할 수 있는 방법에 대한 연구가 계속되고 있으므로(이상준, 1992), 이러한 문제점은 곧 보완될 수 있을 것이다. 이제 퍼지공간분석에 사용되는 퍼지집합의 속성에 대해 살펴 보기로 하자.

3. 퍼지공간분석을 위한 퍼지집합의 속성

퍼지집합을 이해하기 위해서는 먼저 보통집합(crisp set)에 대한 이해가 선결되어야 한다. 그러나 보통집합에 대한 소개는 생략하기로 하고, 퍼지공간분석 특히 퍼지분류를 위해 필요한 최소한의 퍼지집합의 속성에 대해서만 소개하기로 한다.

먼저 몇 가지의 기본용어와 기호를 정의하면, 대상물 전체로 이루어진 전체집합은 영문대문자(U)로 표시하고, 이 집합의 구성요소는 대문자(x)로 나타낸다. 그리고 전체집합 U의 부분집합을 A라 할 때 A가 U에 속한다는 것은 A⊂U로 나타내며, x가 A의 구성요소라는 것은 x∈A로 나타낸다. 이렇게 전체집합의 구성요소 x가 A에 속하는 정도를 정해주는 함수를 特性函數(Characteristic function)라 하며 다음과 같이 표시한다.

x∈A면 $\mu_A(\mathrm{x})=1$

x∉A면 $\mu_A(\mathrm{x})=0$

보통집할에서는 한 집합의 구성요소가 다른 집합에 소속되는 것이 불가능하다. 즉 특성함수는 x가 A의 성질을 만족시키면 1, 그렇지 않을 경우에는 0이다. 그러나 퍼지집합의 구성요소는 여러 집합에 소속되는 것이 가능하므로 특성함수는 0과 1사이의 모든 값을 가질 수 있다. 한가지 기억할 것은 퍼지집합은 보통집합(crisp set)과 반대되는 개념이 아니라, 보통집합을 포함한다는 사실이다.

그러나 Zadeh는 퍼지집합에서 특성함수라는 용어대신 歸屬函數(membership function)라는 용어를 사용하였다. 그러므로 본고에서도 퍼지집합에 한해서는 특성함수란 용어대신 귀속함수를 사용하기로 한다. 이제 퍼지분류에 사용되는 퍼지집합의 특성과 연산에 관한 기본적인 용어들을 정의해 보자.

정의 1 : 전체 집합 U의 퍼지집합 A는 다음과 같이 순서쌍으로 표현된다(Kaufmann, 1973).

$\mathrm{A}=\{(\mathrm{x}, \mu_A(\mathrm{x})) : \mathrm{x}\in \mathrm{U}\}, \quad \mathrm{M}=[0,1]$

$\mu_{\mathrm{A}}(\mathrm{x})$는 귀속함수로 x가 A에 속하는 소속정도를 나타내며, 그 소속정도는 M영역에 포함된다. 영역 [0,1]은 0과 1사이의 모든 값을 가질 수 있음을 의미한다.

정의 2 : 교집합 A∩B에 대한 귀속함수 $\mu_{\mathrm{A}\cap\mathrm{B}}(\mathrm{x})$는 모든 x∈U에 대해

$\mu_{\mathrm{A}\cap\mathrm{B}}(\mathrm{x})=\min(\mu_{\mathrm{A}}(\mathrm{x}), \ \mu_{\mathrm{B}}(\mathrm{x}))$라 정의한다.

정의 3 : 합집합 A∪B에 대한 귀속함수 $\mu_{\mathrm{A}\cup\mathrm{B}}(\mathrm{x})$는 모든 x∈U에 대해

$\mu_{\mathrm{A}\cup\mathrm{B}}(\mathrm{x})=\max\{\mu_{\mathrm{A}}(\mathrm{x}), \ \mu_{\mathrm{B}}(\mathrm{x})\}$라 정의한다.

정의 4 : 퍼지집합간의 거리는 일반적으로 해밍거리(Hamming distance)[8] 또는 유클리드거리로 측정된다.

정의 5 : α절단(α-cut)과 분해 원리를 이용하여 퍼지집합을 보통집합으로 변화시킬 수 있다. α절단은 귀속함수의 값을 임의의 값 α와 비교하여 α보다 크면 1, 작으면 0으로 하여 귀속함수를 특성함수로 변화시키는 것이다. 그리고 분해원리란 퍼지집합을 α와 보통집합의 곱으로 분해하는 원리이다.

정의 6 : A, B⊂U인 부분집합들 간의 관계 $R=\{((\mathrm{x},\mathrm{y}), \ \mu_R(\mathrm{x},\mathrm{y})) \mid (\mathrm{x},\mathrm{y})\in \mathrm{A}*\mathrm{B}\}$를 연산자 *를 만족시키는 A, B 사이의 퍼지관계라 하는데, 일상의 언어로 표현하기 어려운 모호한 관계를 표현하는 수단으로 사용된다.

정의 7 : 두 개의 퍼지관계를 합성하여 새로운 관계를 추론하기 위해서는 합성법을 사용한다. R_1 이 A×B상의 퍼지관계이고, R_2가 B×C상의 퍼지관계라면 R_1과 R_2의 합성은 A에서 C로의 관계를 나타낸다. 최대-최소 합성(max-min composition)은 가장 널리 사펼되고 있는 합성법으로 귀속함수는 다음과 같이 정의된다.

$$\mu_{R_1 \cdot R_2}(x, z)=\max_y \min[\mu_{R_1}(x, y), \ \mu_{R_2}(y, z)]$$

정의 8 : 최소-최대 합성(min-max composition)의 귀속함수는 다음과 같이 정의된다.

$$\mu_{R_1 \cdot R_2}(x, z)=\min_y \max[\mu_{R_1}(x, y), \ \mu_{R_2}(y, z)]$$

정의 9 : A×B에서의 퍼지관계 R이 $\mu_R(x, x)=1$이면 반사관계(reflexive relation)이며,

$\mu_R(x, y)=\mu_R(y, x)$이면 대칭관계(symmetric relation)이다.

정의 10 : 집합 A에 대한 퍼지관계 R은 모든 $x, y, z \in A$에 대해 $\mu_R(x, y) \in R$, $\mu_R(y, z) \in R$이면 $\mu_R(x, z) \in R$일 때, 이행관계(transitive relation)를 갖는다.

정의 11 : 퍼지관계가 반사적이고 대칭적일 때 닮음관계(resemblance relation)라 하고, 반반사적(anti-reflexive)이고 대칭적인 경우는 비닮음관계(dissemblance relation)가 된다.

정의 12 : 퍼지관계가 반사성, 대칭성 그리고 최대-최소 합성의 관계를 만족할 때 유사관계(similitude relation)라 하며, 반반사성, 대칭성 그리고 최소-최대 합성의 관계를 만족할 때는 비유사관계(dissimilitude relation)가 된다.

정의 13 : 집합 A가 어느 특정 성질을 만족하고, 이 성질을 만족하는 A의 부분집합 사이의 교집합이 관계 R을 만족한다면, 이 관계는 폐쇄성질(closure property)을 가진다고 한다(이광형, 1991). 이행폐쇄란 연결관계를 추론하기 위한 하나의 수단으로 최대-최소 합성법을 이용한 최대-최소 이행폐쇄(max-min transitive closure)가 널리 사용되고 있는데 $\check{R}=R \cup R^2 \cup R^3 \cup \cdots$로 정의된다.

정의 14 : 한 요소가 집합에 속하는 정도가 아주 높거나 낮으면 모호성이 낮으며, 반대의 경우에는 모호성이 높다. 모호성의 측정을 위해서는 퍼지집합 A에 가장 가까운 크리스프 집합을 구한다. 귀속함수가 0에서 0.5사이이면 특성 함수는 0이며 0.5에서 1사이이면 1이다. 단 0.5의 경우에는 0 또는 1을 임의로 취할 수 있다. 이렇게 변형된 크리스프 집합을 A라 한다면 퍼지지수는 아래의 두 가지 방법으로 측정된다(Kaufmann, 1973).[9]

$$v(A)=\frac{2}{n}d(A,\ A)$$

$$\eta(A)=\frac{2}{\sqrt{n}}e(A,\ A)$$

이상에서 퍼지분류를 위해 꼭 필요한 퍼지집합을 기본적 용어만 소개하였다. 이제 퍼지분류법에 대해 구체적으로 살펴보기로 하자.

4. 퍼지분류법

퍼지분류는 표준공간을 주어진 장소들의 퍼지부분집합으로 표현하는 것이다. 퍼지분류법은 크게 퍼지이산분류(Fuzzy discrete classification)와 퍼지중첩분류(Fuzzy conjoint classification)로 구분된다.

1) 퍼지이산분류

퍼지이산분류는 거리에 의한 분류와 유사성에 기초한 분류로 구분된다. 그러나 이 둘은 분류 절차만 다르지 결과에서는 차이는 없다(Leung, 1984). 그러므로 본고에서는 퍼지유사성에 의한 분류절차만 소개하기로 한다.

먼저 장소간의 거리를 측정한다. 이를 위해서는 각 변수 값을 변수 최대 값으로 나누어 각 변수의 최대 값이 1이 되도록 한다. 그리고 각 장소간의 거리를 퍼지거리를 사용하여 측정하고 이를 유사성(1-거리)으로 변환하여 유사성 행렬로 만든다. 이 행렬은 반사적이며, 대칭적이다. 그러므로 행렬이 이루는 관계는 닮음관계이다. 그러나 유사성 분류가 가능해지려면 최대-최소 이행관계가 성립되어야 한다. 즉 유사성행렬이 유사성관계를 가져야만 분류가 가능한 것이다. 최대-최소 이행의 의미는 장소 a, b간의 유사성은 a, c간의 유사성과 c, b간의 유사성의 최소 값보다는 최소한 크다는 것을 의미한다. 즉 두 장소간의 유사성은 다른 장소를 경유한 유사성보다 항상 큰 것이다. 그러나 모든 유사성행렬이 최대-최소 이행이 가능한 것이 아니므로 비이행적인 유사성행렬을 이행적인 행렬로 변환할 필요가 있다. 이를 위해서는 이행폐쇄가 일어날 때까지 최대-최소합성을 계속하여, 최대-최소합성 $R^k=R^{k+1}$일 때의 R^k 행렬을 구한다. 이 행렬을 이용하여 α절단을 하고, 분해하여 분류하게 된다. α수준이 높을수록 군집수는 적어진다.

2) 퍼지중첩분류

거리행렬을 구한 다음 이를 바탕으로 프랑스의 의학자인 Pichat(1970)가 개발한 알고리듬을

사용하여 분류하는데, 그 절차는 다음과 같다.

① 거리행렬을 가지고 임의의 임계치(α 수준)를 정한 다음 이 값보다 작거나 같은 수를 1, 그렇지 않은 경우를 0이라 한다. 이 행렬은 반반사적이고 대칭적이다.

② 대각선과 대각선 이하는 행렬에서 삭제된다.

③ i행을 $\{i\}$와 $\{0$값을 가진 열들$\}$ 집합으로 기술한다.

④ i행의 집합들과 $i+1$행의 집합들의 합집합 연산을 행한다. 즉 모든 행을 대상으로 순차적으로 $\{i\}\cap\{i+1\}$, $\{i\}\cap\{i+1$에서 0값을 가진 열들$\}$, $\{i$에서 0값을 가진 열들$\}\cap\{i+1\}$, $\{i$에서 0값을 가진 열들$\}\cap\{i+1$에서 0값을 가진 열들$\}$을 구하는 것이다.

⑤ $X \cdot X=X$, $X+XY=X$, $X+X=X$의 규칙을 적용한다.

⑥ 연산결과에 대한 여집합을 구하여 열거하면 분류결과를 얻는다.

5. 사례연구 : 경상남도의 도시기능분류

1) 퍼지분류 결과

(1) 퍼지이산분류 결과

경남의 읍급이상 도시의 산업별 고용비율을 대상으로 상대적 해밍거리를 사용하여 유사성행렬을 만든 다음, 최대-최소 이행을 계속한 결과 R^8에서 이행폐쇄가 이루어 졌다. R^8행렬을 이용하여 퍼지이산분류를 행하여 유사성 1에서 0.7909까지의 27개 α 수준에서의 분류결과를 얻을 수 있었다(그림 2). 분류결과를 소개하면 먼저 창원과 장승포는 2단계(0.9725)에서 군집화되기 시작하므로 고용구조면에서 가장 유사한 도시로 볼 수 있다. 이 두 도시는 제조업 종사자 비율이 각각 86.36%와 85.70%로 매우 높은 반면, 농업·수렵·관련 서비스업, 광업, 교육 서비스업, 보건·사회복지사업은 종사자 비율이 0%이다. 창원과 장승포는 3단계(0.9695)에서 양산과 군집화된다. 그리고 제 4 단계(0.9429)에서는 전

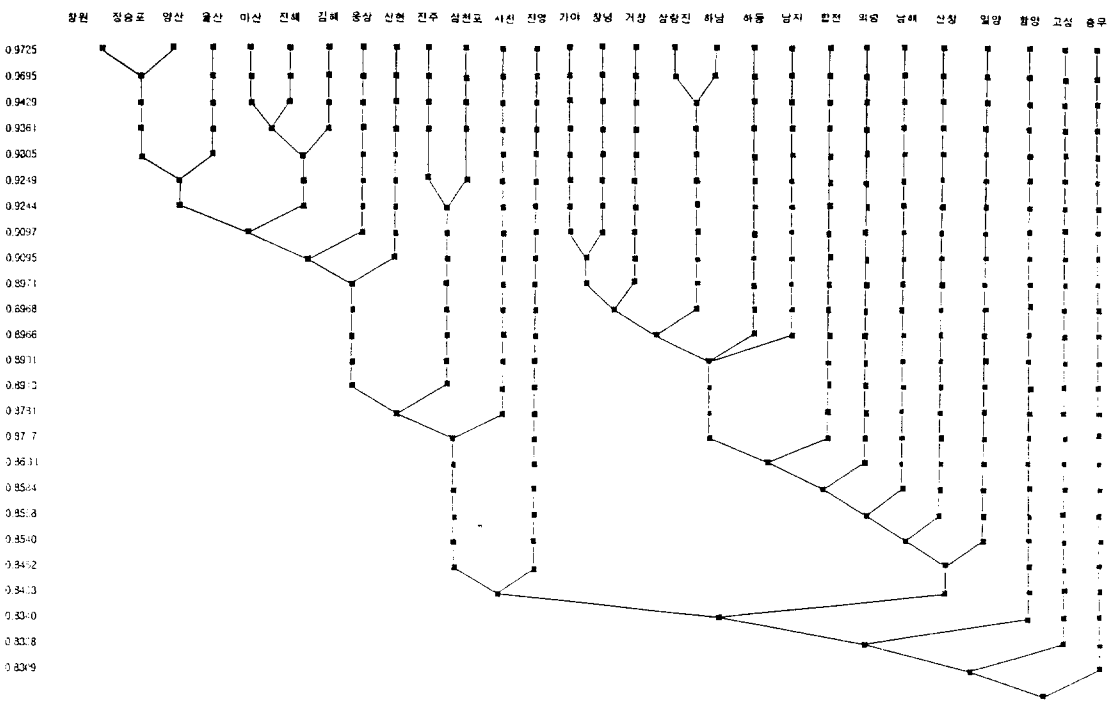

그림 2. 퍼지이산분류의 군집화 dendrogram

분야에서 아주 유사한 고용구조를 가진 삼랑진과 하남이 하나의 군집을 형성한다. 이와 같이 군집화과정은 27단계인 0.7909에서 완전히 하나의 군집화가 될 때까지 계속된다.

군집화가 마치면 군집의 수를 결정해야 하는데, 퍼지분류도 기존의 군집분석과 마찬가지로 군집수의 결정에 대한 명확한 정보를 제공하지 않는다. 그러므로 연구자가 각 군집의 특성을 파악하고 그 의미를 유추하여 군집수를 결정하게 된다. 이러한 측면에서 본 연구에서는 6개의 군집이 적당하다고 판단하여 0.8452수준의 결과를 이용하기로 하였다. 분류결과는 다음과 같다.

군집 1 : 창원, 울산, 마산, 진주, 진해, 삼천포, 김해, 장승포, 양산, 웅상, 신현, 사천

군집 2 : 밀양, 의령, 가야, 창녕, 남지, 삼랑진, 하남, 남해, 하동, 산청, 거창, 합천

군집 3 : 충무

군집 4 : 고성

군집 5 : 함양

군집 6 : 진영

〈표 1〉 도시별 퍼지집합 포함 빈도수 및 $v(A)$ 퍼지지수

도 시	빈도수	퍼지지수
창원시	6	0.102
울산시	3	0.104
마산시	10	0.241
진주시	12	0.408
진해시	9	0.230
충무시	1	0.181
삼천포시	2	0.369
김해시	13	0.295
밀양시	4	0.338
장승포시	3	0.083
의령읍	5	0.361
가야읍	6	0.369
창녕읍	9	0.483
남지읍	4	0.333
삼랑진읍	17	0.333
하남읍	15	0.322
양산읍	10	0.082
웅상읍	21	0.171
진영읍	3	0.104
신현읍	9	0.270
고성읍	4	0.390
사천읍	22	0.425
남해읍	2	0.237
하동읍	1	0.383
산청읍	3	0.162
함양읍	1	0.265
거창읍	12	0.473
합천읍	2	0.219

본 결과에 의하면 경남의 도시들이 제조업 종사자 비율에 의해 크게 구분됨을 알 수 있다. 군집 1의 도시들을 삼천포를 제외하고는 제조업으로 특화된 도시들로 8개 시급도시와 4개 읍급도시로 구성되어 있다.[10] 군집 2는 제조업 종사자 비율은 낮으나 교육종사자 비율이 평균 이상인 도시들로(남해제외) 밀양을 제외하고는 전부 읍급도시로 구성되어 있다. 충무, 함양, 고성 그리고 진영은 독립성을 유지하는데, 이는 이들 도시의 고용구조가 특이하기 때문이다. 충무는 건설업, 함양은 광업과 금융보험업, 고성은 전기·가스·수도업, 그리고 진영은 기타 공공사회·개인업에 종사하는 비율이 경남에서 가장 높다.

(2) 퍼지중첩분류 결과

퍼지중첩분류는 행렬차수(행렬차수−1)/2개의 α 수준에서 이루어질 수 있다. 그러므로 본 연구에서는 28(28−1)/2 즉 378개의 α 수준에서의 분류결과를 이용할 수 있다. 그러나 378개 수준에서의 분류결과를 모두 이용하는 것은 지나친 정보량으로 인하여 불가능하다. 그러므로 본 연구에서는 0.2108 수준의 분류결과만 이용하기로 한다. 0.2108을 선정한 이유는 이 수준이 일종의 임계적 성격을 갖고 있기 때문이다(Tranqui, 1978). 즉 바로 앞의 α 수준 0.2091과 뒤의 α 수준 0.2145와 상대적으로 많은 차이를 갖고 있다.

그러나 퍼지중첩분류에서의 적절한 수준을 결정하는 문제는 앞으로도 계속 연구되어야 할 과제이다. 분류결과는 38개의 퍼지부분집합으로 구성되는데 충무시와 함양읍을 제외하고는 전부 2개 이상의 집합에 포함된다. 도시들이 각각의 집합에 속하는 빈도수는 $v(A)$ 퍼지지수와 함께 〈표 1〉에 나타나 있는데, 이들을 비교한 결과

빈도수와 퍼지지수간의 기능적 상관관계는 찾을 수가 없었다. 즉 여러 퍼지집합에 동시에 속하는 도시의 퍼지지수가 반드시 높은 것은 아니었다. 그러나 단 하나의 α 수준에서의 비교를 통해 퍼지지수가 중첩분류의 빈도수와 무관하다고 결론 내리는 것은 불가능하다. 그러므로 이에 대해서는 보다 많은 연구가 필요할 것으로 사료된다. 이제 중첩분류 결과를 간략히 살펴보자.

—충무는 다른 어떠한 도시와도 결합되지 않는 독특한 고용구조를 갖고 있다.

—사천, 웅상, 삼랑진, 하남은 각각 22개, 21개, 17개 및 15개 집합의 구성요소이다. 이는 이들 도시가 매우 모호한 고용구조를 갖고 있음을 의미한다. 이외에도 마산, 진해, 김해, 거창 등의 도시가 고용구조면에서 모호성을 가진다. 그러나 충무, 함양, 합천, 삼천포, 남해, 하동은 집합에 속하는 반도수가 1∼2회에 지나지 않는 명확한 고용구조를 가진 도시들이다.

—함양은 진주, 삼천포, 사천, 거창과 하나의 집합은 형성한다. 이 집합의 특성은 어업, 도·소매 소비자 용품, 운수·창고·통신업, 기타 공공사회·개인업 분야의 비율이 상대적으로 낮고 건설업 및 금융·보험업의 비율은 높다.

—창원, 울산, 마산, 진해, 김해, 양산, 웅상은 제조업 종사자 비율이 높은 도시들이다. 장승포 또는 신현이 이 도시군과 결합하여 퍼지집합을 형성한다.

—울산은 항상 제조업 비율이 높은 도시들하고만 집합을 형성하나 창원은 제조업 비율이 낮은 삼랑진, 하남과도 집합을 형성한다. 창원이 속한 집합에는 항상 김해가 포함되어 있다.

—삼천포가 속한 2개의 집합에는 인근의 진주, 사천이 포함되어 있다.

—마산, 진해 그리고 김해는 제조업, 운수·창고·통신업 종사자 비율이 상대적으로 높으나 전기·가스·수도, 도·소매 소비자, 기타 공공사회·개인, 교육, 보건·사회복지분야가 열세인 지역이다. 이 세 도시가 중심이 되어 제조업 종사자 비율이 높은 창원, 울산, 장승포, 양산 및 웅상과 집합을 형성하기도 하고 제조업 종사자 비율이 평균인 진주 및 평균이하의 삼랑진과도 집합을 형성한다.

—의령, 삼랑진, 하남은 어업, 광업, 제조업, 전기·가스·수도업, 건설업, 운수·창고·통신업, 부동산 임대·사업서비스업의 비율은 낮으나 교육서비스업 및 기타 공공사회·개인업의 비율은 높다.

—가야와 창녕, 양산과 웅산, 남해와 하남, 그리고 남지와 삼랑진은 각각 전반적으로 유사한 고용구조를 갖고 있다.

2) 군집분석 결과와의 비교

이상의 퍼지분류결과가 기존의 군집분석의 결과와 어떠한 차이점을 갖고 있는지를 살펴보기 위해 군집분석을 실행하였다. 분석결과의 RSQ, Pseudo F, Pseudo t^2 값을 고려할 때, 적정군집의 수는 5개 정도로 볼 수 있는데, 그 결과는 다음과 같다.

군집 1 : 장승포, 양산, 창원, 마산, 진해, 웅상, 신현, 울산

군집 2 : 진주, 삼천포, 밀양, 의령, 사천, 함양

군집 3 : 삼랑진, 하남, 창녕, 거창, 가야, 남지, 합천, 남해, 하동, 산청

군집 4 : 충무, 고성

군집 5 : 진영

군집 1의 도시들은 제조업 기능이 특화된 도시이며, 군집 2는 운수·창고·통신업과 보건·사회·복지사업의 종사자 비율이 평균 이상으로 높은 도시들이다. 군집 3은 제조업 기능이 낮은 반면 도·소매 소비자 용품, 숙박·음식점 또는 교육서비스업의 기능이 다소 특화된 도시들이며, 군집 4는 건설업, 운수·창고·통신업, 금융 및 보험업, 부동산 임대·사업서비스업이 평균 이상인 고용구조를 갖고 있다. 반면 군집 5의 진영은 기타 공공사회·개인업이 압도적인 도시이다.

군집화과정을 살펴보면 제1단계에서 장승포와 양산이 군집화되며, 2단계에서 삼랑진과 하남이 군집화되는 등, 순차적으로 군집화과정이

이루어진다. 이 군집화과정을 퍼지이산분류의 군집화과정과 비교해 본 결과 다소의 차이점은 있지만[11] 상당한 유사성이 있음을 발견하였다. 최종분류결과에서도 역시 다소간의 차이점을 발견할 수 있지만 퍼지이산분류와 군집분석은 명확한 결과를 제공한다는 점에서 근본적으로 유사하다고 할 수 있다. 즉 퍼지이산분류는 퍼지집합이론은 이용하여 분류를 행하지만 그 분류결과는 명확하다.

반면 퍼지중첩분류는 군집분석과는 완전히 다른 결과를 제공한다. 군집간의 경계가 모호할 뿐만 아니라, 각 군집에 명확히 속하는 구성요소와 그렇지 않은 구성요소가 존재한다. 그리고 군집화과정을 파악하는 것이 현실적으로 불가능할 정도로 분류가능수준이 지나치게 많다. 본 사례연구에서 군집분석의 분류결과는 5개의 명확한 집합이지만, 퍼지중첩분류의 결과는 구성요소가 중복되는 38개의 퍼지집합이다. 퍼지집합의 수는 α 수준의 선택에 따라 달라지므로 퍼지중첩분류가 반드시 군집분석보다 많은 수의 군집을 형성한다고 말할 수는 없지만, 일반적으로 퍼지중첩분류는 군집분석보다 많은 수의 군집을 형성한다. 그리고 이것은 퍼지중첩분류의 결과해석을 어렵게 하는 하나의 요인으로 작용하고 있다.

6. 요약 및 결론

본 논문에서는 퍼지이론을 공간분석에 적용하기 위한 이론적 배경을 고찰한 다음, 퍼지분류법에 대해 살펴보았다. 그리고 경상남도 읍급이상 도시의 산업별 고용비율을 가지고, 퍼지분류법을 적용하여, 퍼지이론이 군집분석에 비하여 어떠한 차이점이 있는지를 비교하여 보았다. 본 연구의 내용을 요약하면 다음과 같다.

첫째, 정보의 보호성은 자료의 구체성 부족, 인간행태, 인내치문제, 분류기준의 부족 등에 의해 발생한다. 기준의 공간분석기법으로는 모호한 정보를 처리할 수 없으므로 새로운 기법을 도입할 필요가 있는데, 퍼지이론은 모호한 정보를 처리가능하게 하며, 처리결과는 모호한 공간현실을 반영한다. 퍼지이론을 공간분석에 적용하기 위해서는 본문에서 소개한 여섯 가지의 전제조건이 필요한데, 이는 전통적 공간분석의 전제를 확장한 것으로 볼 수 있다.

둘째, 퍼지분류기법에는 퍼지이산분류와 퍼지중첩분류가 있는데, 퍼지이산분류는 분류기법상 퍼지이론을 적용함으로 모호한 자료의 처리는 가능하나 분류결과는 명확하다. 즉 두개 이상의 분류집단에 동시에 속하는 구성요소가 존재하지 않으므로, 배재된 중앙의 법칙이 성립한다. 그러므로 퍼지이산분류는 분류집단긴의 점이성을 고려하지 못한다. 분류집단의 수는 α 수준의 선택에 따라 달라지는데, 퍼지유사성에 의해 분류를 행할 경우 α 수준이 높을수록 군집수는 적어진다. α 수준의 선택은 연구자의 각 집합에 대한 특성파악과 의미부여에 의한 판단에 의해 이루어진다.

셋째, 퍼지중첩분류 결과는 분류집단간의 interface를 고려하므로, 모호한 공간현실을 잘 반영하는 분류법이라 할 수 있다. 그리고 소수의 집합에만 속하는 구성요소와 많은 수의 집합에 동시에 속하는 모호한 구성요소의 구별이 가능하다. 그러나 퍼지중첩분류는 분류결과가 지나치게 많아 적절한 α 수준을 선택하기 어렵고, 결과해석이 상대적으로 난해하다는 문제점을 갖고 있다. α 수준은 α 값의 변화정도와 연구자의 판단에 의해 결정된다.

넷째, 퍼지분류의 α 수준은 군집분류의 계수(fusion coefficient)와 비슷한 개념으로 볼 수 있으며, 단계(stage)로도 간주할 수 있다. Pseudo F 등의 지수를 참조하여 연구자가 임의적으로 군집수를 결정하는 군집분석과 마찬가지로[12] 퍼지분류의 군집수 결정 역시 연구자의 임의적 판단에 의존한다. 퍼지이산분류는 명확한 결과를 갖는다는 데서 군집분석과 유사하다고 말할 수 있으나, 퍼지중첩분류는 군집분석과는 완전히 다른 결과를 제공한다.

다섯째, 경남의 읍급이상 도시의 기능을 퍼지이산분류, 퍼지중첩분류 및 군집분석에 의해 분류한 결과를 살펴보면, 먼저 퍼지이산분류는 크게 제조업 종사자 비율이 높은 도시와 교육종사

자 비율이 높은 도시로 구분한다. 제조업비율이 높은 도시군 등의 확인이 가능하다는 면에서 퍼지중첩분류의 결과도 퍼지이산분류의 결과와 약간의 유사성을 갖고 있다. 그러나 퍼지중첩분류는 하나의 도시가 여러 개의 집합에 동시에 속하는 점이성을 반영한다. 그리고 군집분석에서는 제조업 기능 등에 의해 5개의 명확한 도시군이 형성된다. 이상의 세 가지 분류에 걸쳐 공통적으로 발견되는 군집은 창원, 울산, 마산, 진해, 김해, 양산, 웅상, 장승포, 신현의 제조업 비율이 높은 군집이며, 충무는 독특한 고용구조로 인해 모든 분류에서 단독군집을 형성한다.

여섯째, 분류결과의 차이점과 공통점에도 불구하고 어느 한 가지 분류법에 의한 결과가 경상남도 읍급이상 도시의 기능적 성격을 다른 분류결과보다 정확히 규정한다고 말할 수 없다. 사용하는 분류기법 또는 알고리듬에 따라 분류결과가 달라지는 것은 당연한 현상으로 각각의 분류결과는 공간구조를 밝히기 위한 의미 있는 지표가 된다(Rolland-May, 1984).

현재 퍼지이론은 날로 발전하고 있으며 그 응용영역은 넓혀가고 있다. 본고에서 소개한 퍼지이산분류와 퍼지중첩분류 외에도 퍼지교사분석(Fuzzy Supervised Classification) 및 퍼지 c-means 기법이 개발되어 원격탐사에 이용되는 등(Wang, 1990; Foody, 1992), 퍼지분류법의 종류 또한 다양해지고 있으며, 퍼지회귀분석법, 퍼지요인분석법 등의 개발도 본격화되고 있다. 따라서 보다 효율적인 공간분석을 위해서는 퍼지이론에 대한 지속적인 연구와 공간분석에 적용하기 위한 시도들이 필요할 것으로 생각된다.

註

1) 국내외의 대부분의 서적과 논문에서는 퍼지 이론이 1965년 Zadeh에 의해 창설되었다고 기술되어 있다. 그러나 Gaines는 이를 잘못이라고 규정하고 그 기원을 1962년으로 명확히 밝힌 바 있다.

2) 엄밀히 말하면 모호성과 퍼지성이 동일한 것은 아니다. 원칙적으로 모호성이란 언어에 관한 특성이며, 퍼지성은 대상에 관한 특성이라는 데서 이 두 개념은 구분된다. 예를 들어 구름과 같이 공간적 경계(spatial borderline)가 불명확한 대상은 모호한 대상이 아니라 퍼지한 대상(fuzzy object)이라고 말할 수 있다. 그러나 이러한 엄격한 '퍼지성'과 '모호성'의 구분은 별다른 의미를 갖지 못한다. 왜냐하면 퍼지이론 연구가들이 언어와 대상의 모호성을 구별하지 않기 때문이다(Rolf, 1980a).

3) 1995년 6월 현재 충무시는 통영시, 장승포시는 거제시, 삼천포시는 시천시로 통합된 상태이다.

4) 예를 들면 홍경희(1987), 윤정섭·이기석·최병기(1979)의 연구를 들 수 있다.

5) 모호성의 원인에 관한 이론은 크게 다음의 네 가지이다. 첫째, 물리적 이론으로, 물리적 법칙이 표현 대상과 표현의 관계를 규정하고 이로 인하여 모호성이 생긴다고 주장하는 이론이다. 이 이론은 Russel(1923)에 의해 제안되었는데, 그는 모호성과 그 반대 개념인 정밀성(precision)은 언어적 표현의 특성이라고 주장한다. '표현'과 '표현하고자 하는 대상'의 관계가 '일 대 일'의 관계가 구체성의 부족과 경계선 설정의 어려움에 의해 발생한다는 것이다. 둘째로 행태주의 이론에 의한 모호성의 설명을 들 수 있다. 행태주의에서 말하는 모호성이란 관측자의 언어나 대상에 대한 응답의 실제 행동에서 야기되는 모호성이다(Black, 1937; Hempel, 1939). 예를 들어 상징 T가 대상 x를 의미하는 지에 대해 관측자들에게 물어, 긍정적 대답의 수를 전체 질문의 수와 비교하여 이를 모호성이라고 정의하는 것이다. 이 모호성 지수는 0과 1 사이에서 변화하며, 0.5에 가까울수록 모호성이 높다. 즉 상징과 그것이 의미하는 것에 대한 독자의 행태가 곧 모호성과 직결된다. 행태주의적 모호성의 또 다른 특징은 언어를 사용하는 사람에 따라 모호성의 정도가 변한다는 것이다. 셋째로, Dummett(1975)와 Wright(1975)는 모호성이 인내치(tolerance)의 문제로 인하여 야기된다고 주장한다. 인내치 문제는 변화의 정도와 직결되는데, 그 정도가 너무 작아 보통 사람에게는 인지가 불가능하기 때문에 모호성이 야기된다는 이론이다. 넷째로, Zadeh를 중심으로 하는 퍼지이론가들은 대상을 분류할 경우 각 분류 집단에 속하는 원소를 구분하는 정확한 기준의 부족 때문에 모호성이 발생한다고 보고, 이 문제를 해결하기 위해 퍼지 집합이론을 제시하였다.

6) 확률론 역시 복잡한 이론적 체계를 가지므로 일괄적인 비교는 할 수 없다. 그러므로 여기에서는 Kolmogorov에 의해 공리화된 수학적 이론으로서의 확률적 의미에 국한하기로 한다.
7) 퍼지집합은 항상 전체집합의 부분집합으로 정의되므로 "부분"이라는 말을 일반적으로 생략한다.
8) 일반 해밍거리 : $d(A, B)=\sum_{i=1}^{n} |\mu_A(x_i)-\mu_B(x_i)|$

상대적 해밍거리 : $\delta(A, B)=\frac{1}{n}\sum_{i=1}^{n}|\mu_A(x_i)-\mu(x_i)|$
9) 이외에도 엔트로피를 사용하여 퍼지성을 측정하는 것도 가능하나, 효율적인 측정법은 아닌 것으로 알려져 있다.
10) 창원공단, 울산공단, 울산미포공단, 온산공단, 마산수출자유지역, 진주상평공단, 진해공단, 진해마천주물단지, 옥포공단, 양산공단, 양산웅상공단, 죽도공단과 사천진사공단이 이들 도시의 제조업 기능을 주도하고 있다.
11) 예를 들어 퍼지이산분류에서는 장승포와 창원이 가장 먼저 군집화되고 다음에 양산이 이 군집에 합류하지만, 군집분석에서는 장승포와 양산이 군집화된 다음에 창원이 합류한다.
12) 군집분석의 군집수 결정에 대한 명확한 기준은 아직 마련되지 않은 상태이다(Everitt, 1979).

文 獻

내무부, 1994, 한국도시연감.

김동주, 1981, "군집분석을 이용한 권역설정," 국토계획, 16(2), 60-66.

문정희, 권인근, 1993, "우리나라 읍도시의 도시시설특성 연구," 한양대학교 산업공학연구소 논문집, 37, 29-44.

양순정, 1994, "CHAID 기법에 의한 도시기능의 시론적 연구," 대한지리학회지, 29(3), 360-368.

엄정국, 1991, 퍼지이론, 박영사, 서울.

윤시운, 고상선, 1988, "주성분분석법에 의한 우리나라 도시의 유형화에 관한 연구," 동아대학교 부설 해양자원연구소 연구보고 1, 69-79.

윤정섭, 이기석, 최병기, 1979, "한국도시(서울, 부산 및 5개 거점도시)의 기능, 형태 및 구조에 관한 연구," 국토계획, 14(2), 3-30.

이광형, 1991, 퍼지이론 및 응용, 홍릉과학사, 서울.

이상준, 1992, "다속성 의사결정 시스템의 퍼지모델링," 중앙대학교 전자계산학 박사논문.

이우종, 1994, "도시의 기능과 토지이용상황에 의한 도시분류체계에 관한 연구," 국토계획, 29(2), 83-96.

홍경희, 1987, 도시・촌락조사법, 법문사, 서울.

Beguin, H. and Thisse, J.F., 1979, An axiomatic approach to geographical space, *Geographical Analysis*, 11, 325-341.

Black, M., 1937, Vagueness: An exercise in logical analysis, *Philosophy of Science*, 4, 427-455.

Dubois, D. and Prade, H., 1985, *Théorie des possibilités*, Masson, Paris.

Dummett, M., 1975, Wang's paradox, *Synthese*, 30, 301-324.

Everitt, B., 1979, Unresolved problems in cluster analysis, *Bicmntrics*, 35, 169-181.

Foody, G.M., 1992, A fuzzy sets approach to the representation of vegetation continua from remotely sensed data: an example from Landsat heath, *Photogrametric Engineering & Remote Sensing*, 58(2), 221-225.

Fustier, B., 1979, *Les interactions spatiales en économie*, Sirey, Paris.

Gaines, B.R., 1983, Precise past-fuzzy future, *Int. J. Man-Machine Studies*, 19, 117-134.

Gale, S. and Atkinson, M., 1979, On the set theoretic foundations of the regionailzation problem, S. Gale and G. Olsson(eds.), *Philosophy in Geography*, Reidel Pub., Dordrecht, 65-107.

Goodchild, M.F., 1992, Geographical information science, *Int. J. Geographical Information Systems*, 6, 31-45.

Guttenberg, A., 1993, *The Language of Planning: Essays on the Origins and Ends of American Planning Thought*, University of Illinois Press, Urbana.

Hempel, C.G., 1939, Vagueness and logic, ***Philosophy of Science***, 6, 163-180.

Kaufmann, A., 1973, ***Introduction à la théorie des sous-ensembles flous***, Masson, Paris.

Körner, S., 1960, ***The philosophy of mathematics***, Huticinson, London.

Leung, Y., 1984, Towards a flexible framework for regionalization, ***Environment & Planning***, A, 16, 1613-1632.

Lowell, K., 1992, On the incorporation of uncertainity into spatial data systems, *GIS/LIS*, Conference and Exposition, 10-12.

Pichat, E.,1970 ***Contribution à l'Algorithimique non numérique dans les ensembles ordonnées***, PhD thesis Univesity of Grenoble.

Ponsard, Cl. et al., 1988, ***Analyse économique spatiale***, PUF, Paris.

Ponsard, Cl. and Tranqui, 1985, Fuzzy economic regions in Europe, ***Environment & Planning***, A, 17, 873-887.

Racine, J.B. and Reymond, H., 1973, ***L'analyse quantitative en Géographie***, PUF, Paris.

Rolf, B., 1980a, Black and Hempel on vagueness, ***Zeitschrift für allgemeine Wissenschaftstheorie***, 11, 332-346.

Rolf, B., 1980b, A theory of vagueness, ***Journal of Philosophical Logic***, 8, 315-325.

Rolf, B., 1981, ***Topics on vagueness***, PhD. Thesis, Lund.

Rolland-May, C., 1987a, Fuzzy data processing 1: concepts, formalization and statistical indices, ***Sistemi Urbani***, 193-211.

Rolland-May, C., 1987b, La théorie des sous-ensembles flous et la géographie, ***Espace Géographique***, 41-50.

Russel, B., 1923, Vagueness, ***Australasian Journal of Philosophy***, 1, 84-92.

Tranqui, P., 1978, ***Les régions économiques flous. Application au cas de la France***, Librairie de Univcsité de Dijon, Dijon.

Wang, F., 1990, Improving Remote Sensing Image Analysis through Fuzzy Information Representation, ***Photogrametric Engineering & Remote Sensing***, 56(8), 1163-1169.

Wright, C., 1975, On the Coherence of Vague Predicates, ***Synthese***, 30, 325-365.

Zadeh, L.A., 1962, From circuit theory to system theory, ***Proceedings of Institution of Radio Engineers***, 50, 856-865.

Zadeh, L.A., 1965, Fuzzy Sets, ***Information and Control***, 8, 338-353.

Zimmermann, H.J., 1991, ***Fuzzy set theory and its applications***, Kluwer Academic Pub., Norwell.

The Application of Fuzzy Classification Methods to Spatial Analysis

Jung, In-Chul*

Summary

Classification of spatial units into meaningful sets is an important procedure in spatial analysis. It is crucial in characterizing and identifying spatial structures. But traditional classification methods such as cluster analysis require an exact data base and impose a clear-cut boundary between classes. Scrutiny of realistic classification problems, however, reveals that available information may be vague and that the boundary may be ambiguous. The weakness of conventional methods is that they fail to capture the fuzzy data and the transition between classes.

Fuzzy subsets theory is useful for solving these problems. This paper aims to come to the understanding of theoretical foundations of fuzzy spatial analysis, and to find the characteristics of fuzzy classification methods. It attempts to do so through the literature review and the case study of urban classification of the Cities and Eups of Kyung-Nam Province. The main findings are summarized as follows:

1. Following Dubois and Prade, fuzzy information has an imprecise and/or uncertain evaluation. In geography, fuzzy informations about spatial organization, geographical space perception and human behavior are frequent. But the researcher limits his work to numerical data processing and he does not consider spatial fringe. Fuzzy spatial analysis makes it possible to include the interface of groups in classification.

2. Fuzzy numerical taxonomic method is settled by Deloche, Tranquis, Ponsard and Leung. Depending on the data and the method employed, groups derived may be mutually exclusive or they may overlap to a certain degree. Classification pattern can be derived for each degree of similarity/distance α. By taking the values of α in ascending or descending order, the hierarchical classification is obtained.

3. Kyung-Nam cities and Eups were classified by fuzzy discrete classification, fuzzy conjoint classification and cluster analysis according to the ratio of number of persons employed in industries. As a result, they were divided into several groups which had homogeneous characteristics. Fuzzy discrete classification and cluster analysis give clear-cut boundary, but fuzzy conjoint classification delimit the edges and cores of urban classification.

4. The results of different methods are varied. But each method contributes to the revealing the transparence of spatial structure. Through the result of three kinds of classification, Chungmu city which has special characteristics and the group of industrial cities composed by Changwon, Ulsan, Masan, Chinhai, Kimhai, Yangsan, Ungsang, Changsungpo and Shinhyun are evident in common. Even though the appraisal of the fuzzy classification methods, this framework appears to be more realistic and flexible in preserving information pertinent to urban classification.

Key Words: spatial analysis, vagueness, fuzzy classification method, interface, urban classification, Kyung-Nam province.

* Senior researcher, Kyung-Nam Development Institute