

도시인구분포모형 개발을 위한 GA모형과 회귀모형의 적합성 비교연구

A Comparative Study on the Genetic Algorithm and Regression Analysis in Urban Population Surface Modeling

최 내 영*

Nae-Young Choei

요약 본 연구는 최근 다수 도시개발사업들이 활발히 진행되고 있는 화성시 동부권을 사례대상지로 하여 행정구역 단위 인구데이터를 격자형 인구분포자료로 변환한 후 인구유인을 유발할 것으로 예상되는 주요 도시계획관련 공간변수들을 GIS로 측정 대입하여 제네틱 알고리즘기법과 회귀분석기법 두 가지 방법으로 일종의 도시인구분포모형을 구축하였다. 두 가지 모형의 분석결과를 통해 도시환경 해석에 있어서의 두 기법의 성능상 특징점을 비교해 보았으며, 분석결과 GA기법은 변수 설명력에 관한 변별력에 있어 일반회귀분석보다 우월한 특징이 있음을 알 수 있었고 따라서 회귀분석과 병행할 경우 매우 직관적이며 보완적인 도시분석기법이 될 수 있음을 확인할 수 있었다.

키워드 : 제네틱 알고리즘, 일반회귀분석, 단계적 회귀분석, 도시인구분포모형, 적합성 함수

Abstract Taking the East-Hwasung area as the case, this study first builds gridded population data based on the municipal population survey raw data, and then measures, by way of GIS tools, the major urban spatial variables that are thought to influence the composition of the regional population. For the purpose of comparison, the urban models based on the Genetic Algorithm technique and the regression technique are constructed using the same input variables. The findings indicate that the GA output performed better in differentiating the effective variables among the pilot model variables, and predicted as much consistent and meaningful coefficient estimates for the explanatory variables as the regression models. The study results indicate that GA technique could be a very useful and supplementary research tool in understanding the urban phenomena.

Keywords : Genetic Algorithm, Regression, Stepwise Regression, Population Surface Model, Fitness Function

1. 서론

제네틱 알고리즘(Genetic Algorithm)은 격자형 인구자료(gridded population)를 기반으로 제반 도시시설 적정입지 탐색 및 인구추정모형 구축에 적용되는 도시인구분포모형(population surface model) 중 하나로 Holland(1975)가 처음 원천 알고리즘을 개발한 이후 다양한 응용분야의 최적해 탐색기법으로 활용되어 왔으며 도시 및 환경관련 분야에서도 많은 실험적 연구가 시도된 바 있다[13][14]. 이때 공간을 정규 격자로 분해하여 비정형적 조사구역

공간의 인구자료를 격자에 수치적으로 공간 내삽(areal interpolation)하는 일련의 방법은 기존 인구자료의 공간적 제약을 극복하기 위한 매우 유용한 대안이라 할 수 있다[15][16]. 본 연구는 최근 동탄 신도시가 조성된 화성시 동부권을 대상으로 행정구역 단위 인구데이터를 격자형 인구분포자료로 변환한 후 인구유인을 유발할 것으로 예상되는 주요 도시계획관련 공간변수들을 GIS로 측정 대입하여 제네틱 알고리즘(이하 GA)기법과 회귀분석기법 두 가지 방법으로 일종의 도시인구분포모형을 구축하였다. 본 연구는 이렇게 구축한 모형들의 분석결과

* 홍익대학교 공과대학 건설도시공학부 도시공학전공 부교수, nychoei@hongik.ac.kr(교신저자)

통해 도시환경 해석에 있어서의 두 기법의 성능상 특장점을 비교해 보고자 하는데 그 연구 목적이 있다.

2. 연구의 공간적 범위 및 방법

2.1 사례대상지 개관 및 분석방법

본 연구는 2007년도에 조성 완료되어 수도권 내 성공적인 2기 신도시로 평가되는 동탄신도시 준공과 더불어 택지개발사업지구 조성 등이 활발히 진행되고 있는 화성시 동부권을 사례지로 선택하였다. 화성시는 2008년도 중반에 기반시설부담구역제도의 관내 적용을 위해 46개 읍·면·동·리 단위 행정구역별 정밀 자체 인구조사를 실시한 바 있다. 한편 인구주택총조사(센서스)의 경우 2010년 후반에 조사가 실시될 예정이며 그 이전 자료로는 2005년도 자료가 가장 근착자료이다. 따라서 이는 최근 조성된 수도권 신도시와 대규모 택지개발사업지구 입주개시 이후의 급격한 인구동태 변화를 반영하기에는 시차적 한계를 가진다고 하겠다. 이러한 맥락에서 본 연구에서 다루는 도시인구분포모형 분석결과의 시의성 제고를 위해 2008년도에 정밀 인구실사자료를 구축한 화성시를 사례대상지로 택함으로써 자료의 시차 한계를 최소화하고자 하였다. 그림 1은 화성시 동부권 행정구역과 함께 모형에 투입될 주요 면(面)적 공간요소(polygon feature)로서 개발진흥

용도지구 및 개발규제용도지역 지정현황을 보이고 있다. 그림 2 또한 모형에 투입될 점적 공간요소(point feature)로서 주요 기반시설(공원, 녹지, 학교, 도로진입점), 공공서비스시설(경찰서, 주민센터, 우체국, 소방서) 및 지역거점(고속도로 IC, 신도시 중심상업지역) 등의 입지현황을 화성시 동부권내 도시지역을 중심으로 확대하여 보이고 있다. 본 연구는 이러한 공간변수 값들을 GIS 9.3을 이용하여 측정하였고, GA모형은 GeneHunter 2.1을, 공간회귀분석(Spatial Regression)모형은 GeoDa 0.9.8을, 또한 단계적 회귀분석(Stepwise Regression)모형은 SPSS 15.0을 이용하여 분석하였다. 사용된 시스템은 Quad-Core 2.66GHz CPU, 4Gb RAM 하드웨어에 64-bit Windows 7 OS 환경에서 분석을 수행하였다.

2.2 분석을 위한 원천자료

본 연구에서는 공간정보 구축을 위해 한국토지정보체계(KLIS : Korea Land Information System)를 활용하였다. KLIS는 필지, 지형, 용도지역지구 정보 등 주요 도시계획 공간정보가 구축된 국토해양부 종합 토지정보체계로서 본 분석에서 사례지에 대한 자연환경 및 용도지역지구 지정현황 등의 설명변수 속성값의 원천자료로 사용하였다. 건축물정보는 건축행정시스템(eAIS : electronic Architectural Information System)을 사용하였다. eAIS는 기존의 개별 건축물대장에 나타나 있는 건축연면적, 건폐율,

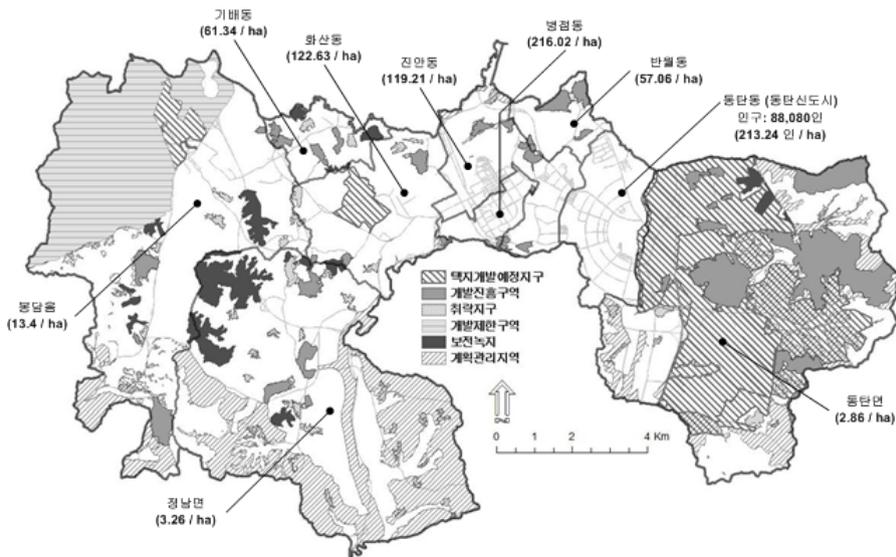


그림 1. 화성시 동부권 행정구역, 인구밀도 및 용도지역지구 지정현황

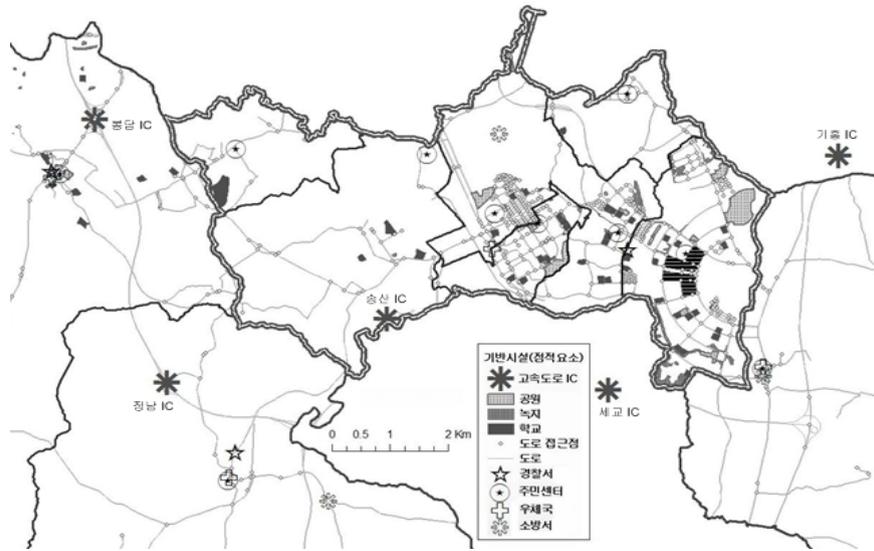


그림 2. 화성시 동부권 기반시설 및 공공서비스시설 분포현황

용적률, 층수, 건축물용도 등 공부자료를 전산화하여 구축한 비공간 속성DB이다. KLIS의 필지 지번과 eAIS의 대지 지번은 동일한 PNU(Parcel Numbering Unit) 코드체계로 작성되어 있어 두 데이터베이스를 PNU 코드를 매개로 결합(join)시킴으로써 KLIS의 토지대장 정보와 eAIS의 건축대장 정보를 개별 필지단위로 통합 연동하여 사용하였다. 또한 사례지자체인 화성시로부터 민원협조를 얻어 앞서 언급한 2008년도 자체 인구조사자료를 확보하여 사용하였다.

3. 격자인구자료 구축

2008년 현재 eAIS 상에서 확인된 화성시 동부권 소재 총 건축물 수는 일반건축물이 63,519건, 집합주택건물이 35,806건이다. 격자인구자료(grided population) 구축을 위해 KLIS와 eAIS를 결합할 때 사용된 PNU코드는 도와 시·군·구 및 읍·면·동·리 등의 행정구역 분류코드와 함께 개별 필지의 본면·부면을 포함한 총 19자리 정수코드로 일관되게 구성되어 있어 GIS Join과정을 통해 총 99,325 필지가 무난히 결합되었다. 그러나 결합결과의 검토과정에서 연립주택과 아파트 등 집합주택의 경우는 공부상 오류가 경미한 반면, 일반건축물의 경우는 제2종 근린생활시설에 실제로는 단독주택이 있는 경우가 61건, 건물 층구분 상 옥탑에 주거시설이 포함된 경

우가 39건, 그 외 우사, 축사, 창고시설에 주거시설이 있는 경우 다수 등 확인불가한 불특정 사례가 발견되어 위성사진자료 등 보조자료를 대조하여 상식적 판단을 통해 연구자 주관에 따라 재정리하였다는 점을 밝혀두고자 한다.

격자인구자료의 격자 크기는 연구목적에 맞게 임의로 설정하게 된다. 기존 격자분석사례를 보면 중국의 성(省) 단위 인구추정의 경우 1×1km[16]로부터 우리나라 시·군 단위 연구의 경우 200×200m [3][4]를 적용한 경우가 있다. 본 연구에서는 준도시급 공간분석의 경우로 편의상 200×200m 규모의 격자를 선택하였다. 이 경우 화성시 동부권 전체 격자시스템은 총 4,207개 셀로 분할되게 되며 공간적인 셀 분할 결과는 그림 4의 도면에서 보이는 바와 같다.

설정된 격자 시스템에 대해 각 셀 당 인구를 추계하는 과정은 우선, 1) 개별 동 및 리 각각의 1인당 평균 주거점유면적을 계산한다. 이는 eAIS로부터 각 행정구역별 주거건축물 연면적을 집계하여 인구조사자료의 각 해당 행정구역 인구수로 나누어 줌으로써 구한다. 2) 대상지 전체 KLIS 지적도 레이어를 베이스맵으로 하여 200×200m 격자 셀로 분할한 다음 KLIS와 eAIS Join DB로부터 각 셀에 속한 주거건축물 연면적을 합산한다. 그림 3은 화산동 소재 1개 셀과 진안동, 병점동에 걸쳐있는 1개 셀 등 2개소를 확대하여 귀속 필지 및 건물이 선택

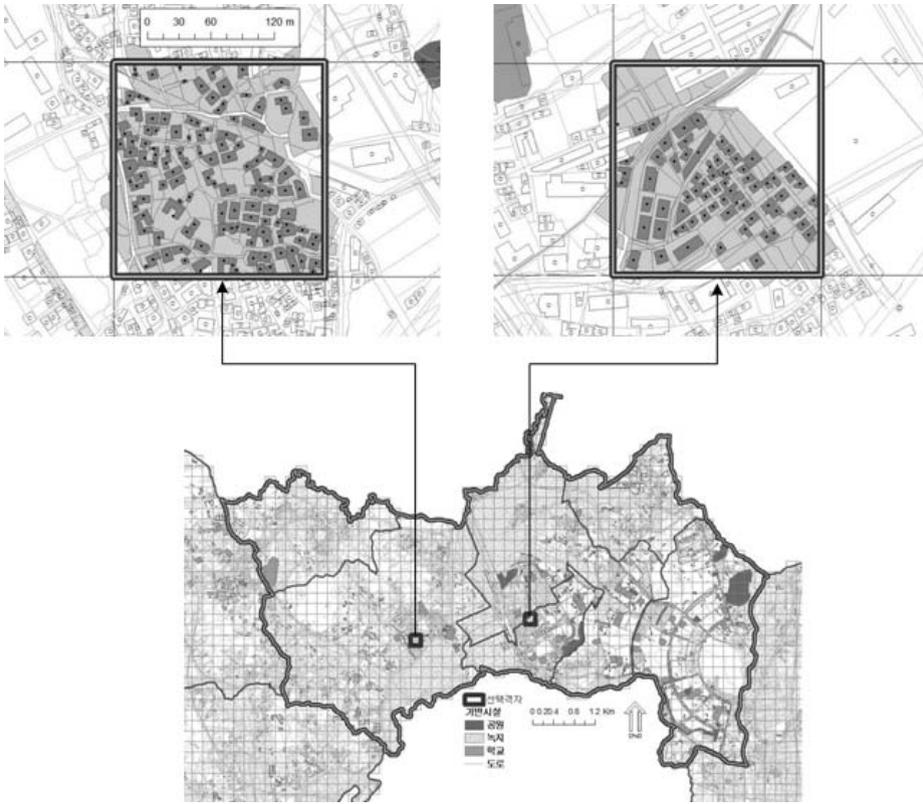
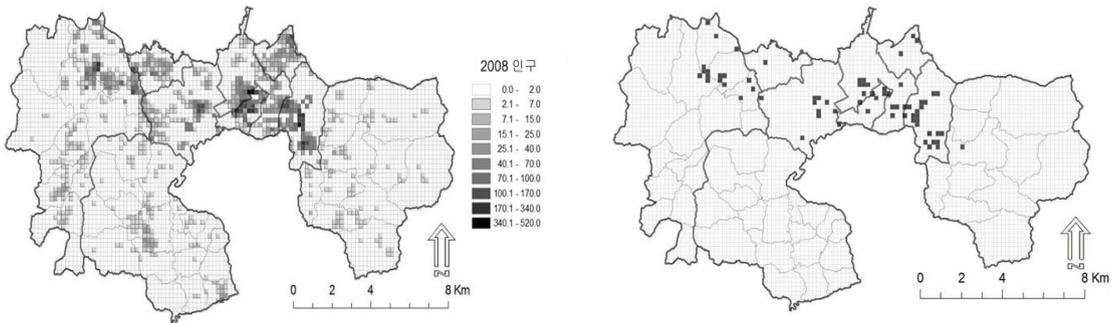


그림 3. 특정 Cell의 귀속 필지 선택결과와 예시도



a) 화성시 동부권 셀 분할 결과 및 인구분포현황

b) 사전분석을 통한 유효분석 셀 선정결과

그림 4. 화성시 동부권 셀별 인구분포 현황 및 70개 유효분석 셀 선정결과

(selection)된 결과를 예시하고 있다. 이때 격자의 경계로 분할된 연면적이 넓은 건축물의 경우 건축물 기준이 아닌 필지기준으로 필지면적의 절반(50%) 이상이 포함되는 셀에 건축물을 귀속시키는 규칙을 적용하였다. 3) 이렇게 구한 셀별 주거건축물 연면적을 1)에서 구한 1인당 주거면적으로 나누어 해당 셀의 인구를 추계하였다.

4. 분석모형 설정

격자분석 결과 그림 1 좌측의 봉담읍, 정남면과 우측 동탄면 등 읍·면 단위 비도시지역의 경우 거주밀도가 일부 낮게 나타남을 알 수 있다. 반면 그림 중앙부 6개동(2006년 행정구역 분할 전 태안읍에 해당하며 그림 2의 붉은 경계선 내부 지역) 등

주요 도시지역의 경우 최근 개발사업지구를 중심으로 인구가 고도 밀집한 셀들이 다수 분포하고 있는 특징도 보이고 있다. 후자와 같은 기성시가화지역의 경우 지나친 돌출값(outliers)들을 제거하여 모형 신뢰도를 제고하기 위해 본 연구에서는 GeoDa 0.9.8의 백분위 맵(Percentile Map)과 상대리스크 맵(Relative Risk Map) 등 Choropleth Map 분석기법을 통해 셀들을 선별하였다. 우선 백분위 분석에서 고밀 셀과 저밀 셀들을 제거하여 유효 설명력을 가질 수 있는 셀들을 1차 선별한 후 상대리스크 분석을 통해 리스크 5.0 이상의 고 리스크 셀들을 추가 제외시켜 보수적인 수준에서 70개 분석 셀을 최종 선택하였다. 여기서 상대리스크란 모든 셀에 대한 총 사상(event) 발생 빈도로부터 구한 평균 리스크를 기초로 특정 셀의 사상발생 기대치를 계산하고 이를 실제 발생 관측치와 비교하여 구한 상대 비율을 의미하는 것으로 이때 그 수치가 높을수록 돌출값 내재 개연성이 높은 것이 된다[9]. 이러한 판별 기준에 의하여 선택된 70개 셀 조합은 사전분석(pilot test)과정을 통해 여타 대안기준을 적용한 다양한 셀 조합 중 가장 우수한 GA 추정결과를 보였다.

본 연구의 GA모형에 투입할 자연환경 요소인 표고 및 경사도 변수값은 GIS 상에서 200×200m셀 각각의 평균 해발고도(m) 및 평균 경사도(%)를 계산하여 구축하였다. 또한 점적 공간요소의 point feature와 용도지역지구 등 polygon feature의 경우 모두 GIS의 NEAR Tool 기능을 이용하여 각 feature boundary에서 각 셀에 대해 수선의 발을 찾아 최단거리를 m단위로 측정함으로써 변수 공간 정보를 구축하였다(설명변수들의 범주, 명칭 및 코

드는 6장 표 2의 분석결과 일람표 참조).

5. 제네틱 알고리즘의 작동기제

본 연구에서 GA모형과 GIS 툴 간의 연동기제의 도식은 그림 5과 같다. GIS에서 구축한 변수가 GA 적합성 함수 평가과정에 투입된 다음 교차(Crossover)와 변이(Mutation) 기제를 거쳐 차세대 모수(parameter) 개체군을 형성하고 그 최적해(optimal solution) 달성여부에 따라 진화가 반복되거나 멈추게 된다. 이때 GA는 생물 종의 진화과정을 설명하는 다윈의 적자생존(Darwinian survival of the fittest) 원리에 그 기초를 두게 된다. 즉, 추정코자하는 모수들의 십진법 수치는 전산처리 과정에서 0과 1의 2진법 숫자(binary string)로 변환되는데 이때 스트링을 유전자 염색체(chromosome) 구조와 같이 해석하여 그림 6의 도식과 같이 한 쌍의 선대(parents) 개체로부터 스트링 일부를 교차시켜 후대(offspring) 개체를 생성시킨다. 그림에서 a)는 4비트 스트링 네 개로 이뤄진 16비트 스트링 전체에 대해 1회 교차시킨 경우를, b)는 염색체 내 4비트 스트링 각각에 대해 모두 교차시키는 과정의 사례 도식을 보이고 있다. 새로운 자손 개체 값들이 모형의 적합성 함수(Fitness Function) 값을 향상시킬 경우 보다 적합(fitter)한 개체로 채택하여 차세대(next generation) 개체군에 진입시키고 그렇지 않을 경우 기각시킨다. GA는 또한 특정 염색체 스트링 일부의 위치를 바꾸는 돌연변이(Mutation)를 발생시킴으로써 일시적으로 달성한 모형의 국지적(local) 극대, 극소를 벗어나 전구간(global) 최대, 최소값 도달 가능성을 높이는 진화적 기제로 작용하

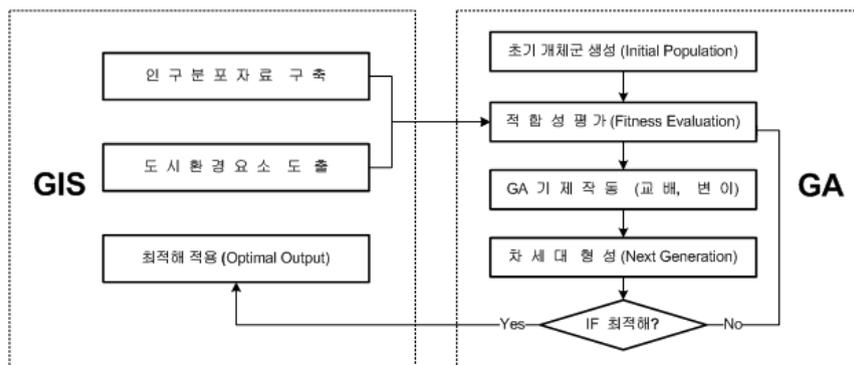
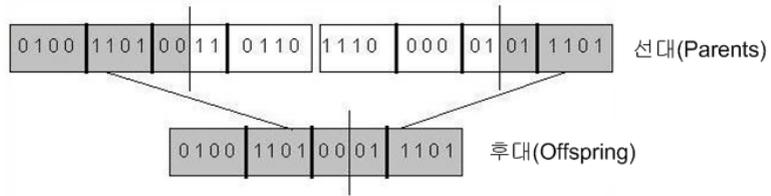
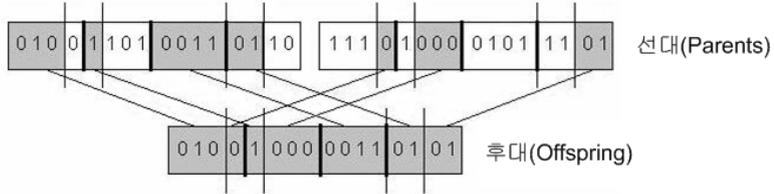


그림 5. GA와 GIS의 연동기제 도식



a) 16비트 염색체의 1회 교차(Crossover)



b) 16비트 내 4비트 염색체 각각의 교차(Crossover)

그림 6. 교차(Crossover) 작동기제 사례 도식

게 된다. 최종적으로 적합성 함수가 200세대 이상 값이 향상하지 않을 때 적정해가 찾아진 것으로 간주한다[10][11]. 본 연구에서는 도시인구분포모형의 피설명변수인 격자셀 추정인구를 다음과 같이 설정하였다.

$$GA_{cell} = M \left[W_S \left(\sum_{s=1}^{n_s} w_s sc_s \right) + \sum_{R=1}^{n_R} W_R \left(\sum_{r=1}^{n_r} w_r rt_r \right) \right] \quad (\text{식 1})$$

식에서 W_S 는 점수화한 자연환경 범주의 가중치, w_s 는 n_s (=2)개의 자연환경 변수 가중치, sc_s 는 변수의 점수이다. 마찬가지로 W_R 은 여타 n_R 개 범주들 각각의 가중치, w_r 은 각 범주내 n_r 개 변수들의 가중치, rt_r 은 변수 최대값에 대한 비율로 산출한 변수 값이다. M 은 괄호 안의 수식을 관찰값에 근접시키는 임의의 승수(Multiplier)이며, 특히 각 범주 또는 계수들의 상대적인 기여도를 비교하기 위해, W_S , W_R , w_s , w_r 등 모든 가중치는 각각 그 합이 1이 되도록 제약조건을 설정한다.

모형성능을 판단하는 적합성함수(Fitness Function)로는 관측된 격자인구 P_{cell} 과 식 1에 의해 계산된 GA_{cell} 간의 상관계수 r 과 상대평균오차(mean absolute relative error) e_{ma} 로 정의하고 r 은 최대화, e_{ma} 는 최소화하는 것으로 모형의 성능을 판단하였다[14]. 이때 n_c 개의 관측셀에 대한 r 과 e_{ma} 는 다음 식과 같이 계산된다.

$$r = \frac{\left[\sum_{n=1}^{n_c} (P_{cell_n} - \bar{P}_{cell}) (GA_{cell_n} - \overline{GA}_{cell}) \right]}{\sqrt{\sum_{n=1}^{n_c} (P_{cell_n} - \bar{P}_{cell})^2} \sqrt{\sum_{n=1}^{n_c} (GA_{cell_n} - \overline{GA}_{cell})^2}} \quad (\text{식 2})$$

$$e_{ma} = \frac{1}{n_c} \sum_{n=1}^{n_c} |(P_{cell_n} - GA_{cell_n}) / P_{cell_n}| \quad (\text{식 3})$$

6. 모형 추정결과

6.1 GA와 회귀분석의 초기모형 비교

우선 표 1에서 GA모형 적합성 함수로 채택한 상관계수 r 의 최대값은 0.558, 상대평균오차 e_{ma} 의 최소값은 0.163으로 나타난 반면 회귀모형에서의 상관계수 r 은 0.632, 상대평균오차 e_{ma} 는 0으로 나타나 두 가지 적합성 함수 모두에서 회귀모형이 일견 다소 나은 결과를 얻은 것으로 보인다.

그러나 회귀분석의 통계량(표 2)을 보면 모든 투입변수에 있어 90% 이상의 신뢰도를 갖는 변수가 없는 것으로 나타나는데 이는 산지와 농지로 이루어

표 1. 초기모형의 적합성 함수 값

적합성 함수	GA모형	회귀모형
	$M = 49.80392$	$R^2 = 0.40042$
r	0.55838	0.63171
e_{ma}	0.16306	0.00000

표 2. 초기변수에 대한 회귀분석 계수 추정치

코드	범주	코드	요소	회귀모형			
				β (계수)	표준오차	t-통계량	확률
1	자연경	1.1	표고	0.03519	0.15499	0.22706	0.82130
		1.2	경사	-0.92849	0.56748	-1.63616	0.10809
2	기반시설	2.1	수계	0.00177	0.00562	0.31449	0.75445
		2.2	도로	-0.02328	0.01717	-1.35587	0.18123
		2.3	학교	-0.00088	0.00789	-0.11134	0.91179
		2.4	공원	-0.00326	0.00342	-0.95442	0.34446
		2.5	녹지	0.00157	0.00290	0.54048	0.59127
3	지역공공서비스	3.1	경찰서	0.00063	0.00225	0.28021	0.78047
		3.2	주민센터	0.00138	0.00344	0.40135	0.68987
		3.3	우체국	0.00234	0.00275	0.85130	0.39866
		3.4	소방서	0.00014	0.00326	0.04254	0.96624
4	주요지역거점	4.1	고속도로 IC	0.00097	0.00355	0.27453	0.78481
		4.2	동탄중심상업	0.00339	0.00387	0.87659	0.38490
5	개발진흥용도지구	5.1	택지개발지구	0.00096	0.00369	0.25992	0.79599
		5.2	개발진흥지구	0.00056	0.00424	0.13120	0.89614
		5.3	취락지구	0.00152	0.00361	0.42144	0.67524
6	개발규제용도지역	6.1	개발제한구역	0.00255	0.00382	0.66862	0.50682
		6.2	보전녹지	0.00168	0.00236	0.71195	0.47980
		6.3	계획관리지역	0.00253	0.00260	0.97337	0.33506

어진 배후 저밀지역에 국지적으로 고밀 택지가 조성되고 있는 화성시 지역 특수성에 기인하는 것으로 판단된다. 가장 설명력이 큰 변수 몇 가지를 살펴보면 경사도 변수가 신뢰구간 10%에 가장 근접하고 있는데, 경사도가 한 단계(3%) 증가할 경우 90%의 큰 인구감소를 유발하는 것으로 나타나고 있다. 다음으로 영향력이 큰 변수는 도로접근성으로서 접근도로에서 1m 멀어질수록 약 2.3% 인구밀도 하락이 발생하는 것으로 나타난다. 그 다음으로는 순차적으로 개발규제용도지역 중 계획관리지역 변수, 기반시설 중 공원까지의 거리, 주요지역거점 중 동탄신도시 중심상업지역 근접성 등의 설명력이 높은 것으로 나타난다.

결과적으로 GA모형과 비교할 때 회귀분석 결과의 가장 큰 차이점은 첫째, 이와 같이 회귀분석 투입변수의 설명력이 전반적으로 낮게 나타날 경우 모형 향상을 위해 진입과 제거를 위한 유효 투입변수 변별력이 거의 파악되지 않는다는 점이다. 둘째, 회귀분석의 경우 개별 변수만의 설명력을 추정할 수 있을 뿐 유사성격으로 집단화할 수 있는 변수군

범주들에 있어 범주 간 유효 설명력의 상대적 기여도를 파악할 수 없다는 점을 들 수 있다.

이와 대비하여 전체 변수를 회귀모형과 동일하게 투입한 GA모형의 경우(표 3), 먼저 범주 단위의 설명력을 별도로 파악할 수 있음을 알 수 있다. 범주부터 살펴보면, 전체 설명력을 1(가중치의 합)로 볼 때 2) 기반시설 범주 해는 10% 이하(0.067), 그리고 3) 공공서비스와 4) 지역거점의 해는 0으로 나타나 나머지 3개 범주와 달리 그 영향력이 거의 없거나 매우 미미함을 명확히 알 수 있다. 이는 2)의 각종 기반시설(도로, 학교, 공원, 녹지)과 3)의 제반 공공서비스시설(경찰서, 주민센터, 우체국, 소방서)의 경우 앞서 그림 2에서도 볼 수 있었듯이 현재 개발이 활발한 화성시 동부권 도시지역에서는 어느 위치에서나 접근성이 거의 균질하게 이미 시설배치가 이루어져 인구유인에 작용하는 공간적 차별성이 없음을 의미하는 것으로 판단된다. 4)의 지역거점 변수에서도 도시지역 일대에 무려 5개의 IC가 거의 균등 분포하고 있고(그림 2 참조), 동탄신도시 중심상업지역의 경우는 아직 입주완료 이전이라 지역거점

표 3. 초기변수에 대한 GA모형 변수 가중치

코드	범주	코드	요소	GA모형	
				범주 해 (WS,R)	요소 해 (ws,r)
1	자연환경	1.1	표고	0.49804	0.00784
		1.2	경사		0.99216
2	기반시설	2.1	수계	0.06667	0.43529
		2.2	도로		0.06667
		2.3	학교		0.00000
		2.4	공원		0.46667
		2.5	녹지		0.03529
3	지역공공서비스	3.1	경찰서	0.00000	0.47843
		3.2	주민센터		0.19608
		3.3	우체국		0.17255
		3.4	소방서		0.17255
4	주요지역거점	4.1	고속도로 IC	0.00000	0.71765
		4.2	동탄중심상업		0.29412
5	개발진흥용도지구	5.1	택지개발지구	0.14902	0.87843
		5.2	개발진흥지구		0.00784
		5.3	취락지구		0.00000
6	개발규제용도지역	6.1	개발제한구역	0.34902	0.01569
		6.2	보전녹지		0.41569
		6.3	계획관리지역		0.56863

으로서의 영향력이 크지 않은 결과로 짐작된다. 이렇듯 일반적으로 매우 중요한 도시계획시설들임에도 불구하고 화성시 동부권의 경우 현재의 도시현황과 입지특성에 의해 이러한 변수들은 모형에서 제외시킬 필요가 있음을 GA모형을 통해 명확히 판별할 수 있다. 나머지 3개 범주 중 1) 자연환경의 경우 50% 가까이 압도적인 영향력을 가지며, 5) 개발진흥용도지구가 15%, 6) 개발규제용도지역이 35% 정도의 영향력을 가짐으로써 인구구성을 설명하는 유효한 변수 범주들로 나타났다. 이는 1)의 자연환경 변수들이 지역에 관계없이 도시개발과 인구집중에 중요한 공통요소임을 의미하며, 5) 개발진흥용도지구의 긍정적 인구유발 효과보다는 6) 개발규제용도지역의 인구증가 억제효과가 더 크다는 점을 시사한다.

6.2 GA와 회귀분석의 축소모형 비교

다음으로는 위 전체 19개 변수를 모두 투입하여 실험(pilot test)한 초기모형의 결과로부터 보다 유효한 설명력을 가지는 것으로 보이는 변수들만을 선별 압축한 축소모형을 구축하고 그 분석결과를

고찰해 보았다. 이를 위해 우선 압도적 영향력을 보이는 자연환경변수들은 그대로 남기고 5) 개발진흥용도 중 변수 해가 0인 취락지구와 6) 개발규제용도 중 범주 내 가중치가 1.5%로 가장 낮은 개발제한구역 변수를 제거하여 6개 변수만으로 모형을 실험해 보았다. 물론 변수 진입, 제거 기준을 더 엄격히 할 수도 있겠으나 여기서는 변수가 지나치게 적지 않도록 고려한 모의실험의 성격을 감안한 것이다.

우선 그림 7은 6개 변수를 투입한 GA모형 적합성 함수 r 의 상승궤적을 나타낸다. 그림에서 보듯 초기값은 매우 낮으나 세대를 거듭하면서 점차 빠르게 향상되어 약 200세대를 넘으면 최대값 0.5559에 근접함을 알 수 있다. 과정 중 국지적 기복은 r 의 최대화와 e_{ma} 의 최소화를 동시에 달성하기 위한 타협적 조절 선택의 결과로 보인다. 표 4는 6개 변수를 투입한 GA모형과 일반회귀분석, 그리고 차후 언급할 단계적 회귀분석(Stepwise Regression)의 적합성 함수 값을 요약한 것이다.

표 4에서 6개 변수로 압축한 GA와 회귀분석 축소모형들 중 이번에는 r 값은 GA가, e_{ma} 값은 회귀분석이 다소 우수하게 나타나고 있다. 이는 양 모형

표 4. 축소모형의 적합성 함수 값

적합성 함수	GA모형	회귀모형	단계적 회귀모형
	$M = 64.31373$	$R^2 = 0.30571$	$R^2 = 0.31550$
r	0.55599	0.55209	0.56108
e_{ma}	0.16115	0.15949	0.15973

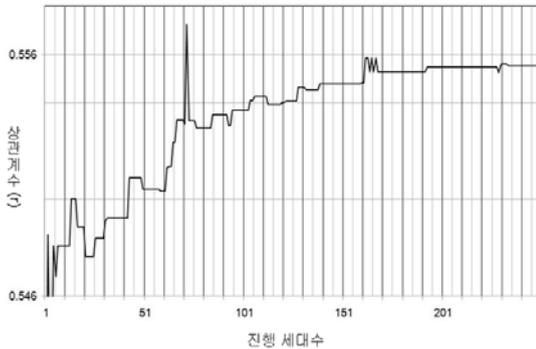


그림 7. GA 축소모형 적합성 함수 향상과정

간 우월을 단편적으로 비교하기 어려울 만큼 GA 최적치가 상대적으로 향상되었음을 의미한다.

두 가지 축소모형 중 먼저 표 5의 회귀분석 결과를 보면 우선 모형의 설명력(R^2)은 오히려 초기모형보다 다소 감소하였으나, 1.2) 경사, 6.2) 보전녹지, 6.3) 계획관리지역이 모두 90-95% 이상의 신뢰도를 보여 변수의 신뢰도가 크게 향상되었음을 알 수 있다. 그러나 나머지 3개 변수들의 신뢰도는 40-50% 수준으로 낮아 결과적으로 모형의 설명력(R^2)을 다소 낮춘 것으로 보인다.

표 6의 GA 축소모형에서도 결과는 이와 유사하다. 먼저 범주별로 보면, 5) 개발진흥용도의 영향력은 11%로 이전과 크게 변함이 없으나, 1) 자연환경이 전보다 다소 낮은 37%로, 6) 개발규제용도가 전보다 높은 50%로 나타나 초기 모형과 다소 차이가 있지만 여전히 가장 영향력이 큰 2개 범주임을 나타낸다. 동시에 변수별 영향력을 볼 때도 초기모형과 매우 유사한 결과가 도출되었음을 알 수 있다. 이러한 축소모형 분석결과를 앞 절의 초기모형과 비교할 때 결과적으로 초기 회귀모형으로부터는 변수 변별력이 없었으나 GA모형으로부터는 변별력이 유효하였다. 뿐만 아니라 GA모형에서는 특히 초기 모형과 축소모형 두 가지의 변수 영향력에 있어서도 변수 진입 제거에 따라 모형의 투입변수 구성이

표 5. 회귀분석 축소모형 계수 추정치

코드	범주	코드	요소	회귀모형			
				β (계수)	표준오차	t-통계량	확률
1	자연환경	1.1	표 고	-0.05306	0.12268	-0.43252	0.66684
		1.2	경 사	-0.88211	0.44712	-1.97287	0.05290
5	개발진흥용도지구	5.1	택지개발지구	-0.00091	0.00134	-0.67742	0.50062
		5.2	개발진흥지구	0.00134	0.00238	0.56295	0.57547
6	개발규제용도지역	6.2	보 전 녹 지	0.00163	0.00092	1.76221	0.08288
		6.3	계획관리지역	0.00301	0.00107	2.80506	0.00668

표 6. GA 축소모형 변수 가중치

코드	범주	코드	요소	GA모형	
				범주 해 (WS,R)	요소 해 (ws,r)
1	자연환경	1.1	표 고	0.37255	0.13333
		1.2	경 사		0.87451
5	개발진흥용도지구	5.1	택지개발지구	0.11373	0.88235
		5.2	개발진흥지구		0.12157
6	개발규제용도지역	6.2	보 전 녹 지	0.50196	0.37255
		6.3	계획관리지역		0.63922

크게 달라졌음에도 불구하고 일단 유효한 것으로 선택되어 재투입된 변수들의 영향력(가중치)은 크게 변화하지 않고 비교적 일관된 결과가 계속 유지된다는 특징을 알 수 있다.

6.3 단계별 회귀분석 추정결과 비교

마지막으로 SPSS 패키지를 이용해 회귀분석에 대해서만 초기모형의 전체 19개 변수에 대해 단계적 회귀분석(Stepwise Regression) 방법으로 유효 변수를 선택해 고찰해 보았다. 이는 앞서 살펴본 바와 같이 회귀분석만으로는 초기모형 추정결과로부터 유효변수들을 직관적으로 변별할 수 없었기 때문이다. 표 7은 그 결과를 보이는데 3단계 진입, 제거를 거쳐 오직 3개 변수만이 최종 채택되었다. 이 중 1.2)의 경사와 6.3)의 계획관리지역이 위 6개 변수 모형과 일치한 반면 그 외 5.3) 취락지구가 선택되어 GA모형의 선택결과와 큰 대조를 보인다. 또한 채택된 3개 변수 모두 97.5% 이상의 높은 신뢰도를 보이고 있긴 하나 앞의 표 4에서 보인 바와 같이 모형의 R²값은 GA분석을 통해 6개 변수를 선별 투입한 축소 회귀모형과 큰 차이가 없음을 알 수 있다.

특히 화성시 동부권 취락지구는 도시지역 외 농림지역에 대개 잔존하고 도시지역의 미개발 지역 일부에만 남아 있는 상황이다. 따라서 이러한 취락지구는 현재의 동부권 인구증가가 도시개발에 의한 현상임을 감안할 때 인구유인을 유도할 개연성이 거의 없다고 판단되는 변수이다. 그럼에도 불구하고 단계별 회귀분석의 주요변수로 채택되고 있는데, 이는 GA 초기모형에서 영향력이 전무한 것으로 나타난 변수가 채택된 것으로써 순수한 통계적 산출결과인 것으로 보이며 따라서 도시계획적 고려와 판단을 위해서는 한계가 있는 분석결과로 여겨진다.

7. 결론

본 연구에서는 대상지 행정구역별 인구조사자료를 격자형 인구분포자료로 변환한 후 GA와 회귀분석 두 가지 방법으로 도시인구분포모형을 구축하고 두 가지 기법의 설명력과 적합성을 비교분석해 보았다. 인구 유입에 영향을 줄 수 있는 다양한 입지 특성 변수들을 투입한 초기모형 분석결과, 변수 영향력에 관한 변별력에 있어 GA모형이 회귀분석보다 다소 우월함을 알 수 있었다. 특히 GA모형에 있어서는 회귀분석과 달리 개별변수 설명력뿐만 아니라 유사한 변수군 범주들 간의 상대적 기여도를 파악할 수 있다는 특징을 알 수 있었다. 또한 초기모형으로부터 보다 유효한 변수로 압축한 축소모형에 있어서도 GA모형은 비교적 일관된 변수 영향력을 유지하며 단계적 회귀분석(Stepwise Regression)모형보다 유효 설명변수 선택에 있어 보다 합리적인 선택결과를 보이는 것으로 판단된다.

그러나 회귀분석기법과 비교할 때 GA모형에 있어서는 채택한 개별변수들의 설명력을 통계적 의미에서 정량적으로 검증할 수 없다는 점과 도출된 적합성 함수값들도 항상 회귀분석을 능가하는 분석결과가 도출되는 것은 아니라는 점에서 한계가 있다. 따라서 GA모형을 사용할 경우 이러한 한계를 감안하여 운용할 필요가 있을 것으로 사료된다. 결과적으로 도시모형에서 보편적으로 사용되는 회귀분석 기법과 비교할 때 GA기법은 하나의 유용한 선택적 분석대안으로서 변수선택 등에 있어서 회귀분석과 병행하여 사용할 경우 두 기법의 장단점을 보완하여 보다 효율적인 모형구축에 활용될 수 있을 것으로 판단된다.

본고의 분석과정에서 근본적으로 아쉬운 점은 사전분석을 통해 가장 유효설명력을 보인 인구 셀들만을 선별했음에도 불구하고 GA와 회귀분석 모두 e_{ma}오차가 더 낮은 수준으로 수렴치 못한 점이다. 물론 이는 앞서도 언급한 바와 같이 화성시 동부권

표 7. 단계적 회귀분석 단독 추정결과

코드	범주	코드	요소	단계적 회귀분석			
				β (계수)	표준오차	t-통계량	확률
1	자연환경	1.2	경사	-0.91543	0.27085	-3.37983	0.00122
5	개발진흥용도지구	5.3	취락지구	0.00381	0.00166	2.29369	0.02500
6	개발규제용도지역	6.3	계획관리지역	0.00233	0.00078	3.00262	0.00378

이 고도로 도시화된 지역 주변에 저밀인구의 산악 지역, 농림지역이 함께 병존하고 있는 이질적 공간 특성에 기인하는 것으로 판단된다. 이 점에서 혹시 최근 인구실사자료를 확보할 수 있는 여타 유사 지자체가 있다면 대상을 변경하여 지역 간 횡단비교연구를 수행해 보는 것도 매우 의미있는 확장연구가 될 수 있을 것으로 사료된다.

참 고 문 헌

[1] 건설교통부, 2006, 기반시설부담금에 관한 법률 산정기준 마련을 위한 연구.

[2] 건설교통부, 2007, 기반시설부담금제도의 실태분석 및 발전방안에 관한 연구.

[3] 국토해양부, 2008a, 기반시설부담구역제도 시행 및 운영활성화를 위한 연구.

[4] 국토해양부, 2008b, 기반시설부담구역제도 운영 편람.

[5] 김재철, 이규철, 2009, "u-GIS 국토정보 제공 시스템", 한국공간정보시스템학회논문지, 제11권, 제1호, pp. 1-8.

[6] 김형복, 전병은, 최내영, 2006, "기반시설부담금에 관한 법률 및 그 운영에 대한 특강자료", 대한국토·도시계획학회 국토도시아카데미.

[7] 최내영, 2009, "기반시설설치구역 지정을 위한 공간정보 적용방안 연구", 한국공간정보시스템학회 논문지, 제11권, 제3호, pp. 40-45.

[8] 화성시, 2004, 화성시 기반시설부담구역 지정 및 부담계획 기준수립 연구.

[9] L. Anselin, N. Lozano, and J. Koschinsky, 2006, "Rate Transformations and Smoothing," A Discussion Paper of the Spatial Analysis Laboratory, University of Illinois at Urbana-Champaign.

[10] P. Balakrishnan, and V. Jacob, 1996, "Genetic Algorithms for Product Design," Management Science, vol. 42, pp. 1105-1117.

[11] S. Chatterjee, M. Laudato, and L. Lynch, 1996, "Genetic Algorithms and Their Statistical Applications: An Introduction," Computational Statistics & Data Analysis, vol. 22, pp. 633-651.

[12] H. Holland, 1975, Adaptation in Natural and Artificial Systems, The University of Michigan Press, Ann Arbor.

[13] X. Li, and A. Yeh, 2005, "Integration of Genetic Algorithms and GIS for Optimal Location Search," International Journal of Geographical Information Science, vol. 19, pp. 581-601.

[14] N. Muttill, and J. Lee, 2005, "Genetic Programming for Analysis and Real-time Prediction of Coastal Algal Blooms," Ecological Modelling, vol. 189, pp. 363-376.

[15] T. Yue, Z. Du, and Y. Song, 2008, "Spatial Models and Geographic Information Systems," Encyclopedia of Ecology, pp. 3315-3325.

[16] T. Yue, Y. Wang, S. Chen, J. Liu, D. Qiu, X. Deng, M. Liu, and Y. Tian, 2003, "Numerical Simulation of Population Distribution in China," Population and Environment, vol. 25, pp. 141-163.

논문접수 : 2010.8.10
 수정일 : 1차 2010.10.22 / 2차 2010.10.28
 심사완료 : 2010.10.30



최 내 영
 1981년 서울대학교 건축학과(공학사)
 1986년 일리노이주립대 건축학과
 (건축석사)
 1996년 펜실베이니아대학교 도시 및 지역계획학과(도시 및 지역계획학 박사)
 1997년~현재 홍익대학교 건설도시공학부 도시공학전공 부교수
 관심분야는 기반시설부담구역제, 단지계획, 주택정책