

필지의 기하학적 특성 정보와 의사결정트리 기법을 활용한 토지 형상 인식 방법론 연구

A Study on the Method of Land Shape Recognition Using the Geometric Feature Information of Parcel and the Technique of Decision Tree

최진호* · 이석희**

Choi, Jin Ho · Lee, Seok Hee

■ Abstract ■

The land feature survey needs to be conducted in an objective and consistent manner. Among the categories of characteristics survey, however, the land shape has had vague reference points, undermining the objectivity and consistency of the survey. As such, this study intended to extract the geometric characteristics information of the land shape by the use of GIS and to apply the information to the data mining algorithm C4.5, thereby suggesting a new method for addressing the problems of subjectivity and uncertainty. The findings showed approximately 81.6% agreement (21,570 parcels among the total 24,433 parcels) between the results of shape classification estimated with rules learned by C4.5 and the results of the existing surveys conducted by experts of Seoul City. Further, the kappa coefficient, which explains accidental or probabilistic agreement rate, was found to be 0.72, showing a considerable level of accuracy. The results of the experiment are significant in that the new method derived standardized rules that resolve the ambiguity of the existing survey. In particular, the method shows the effects of improvement in the consistency of characteristics information, which used to be classified variously according to surveyors. Given the inefficiency of the current land characteristics survey, which relies on manpower and field survey, it is deemed that the application of the computer simulation technique to the land characteristics survey will contribute to greater objectivity and efficiency in the land characteristics survey.

Key Words : Land Characteristics Survey, Land Shape, Data Mining

* 한국감정원 KAB부동산연구원 연구개발실 공시연구부 책임연구원 (주저자, k25830@kab.co.kr)

** 한국감정원 KAB부동산연구원 연구개발실 공시연구부 연구원 (교신저자, sirking_lee@uos.ac.kr)

1. 서론

1. 연구의 배경 및 목적

토지특성 조사는 토지의 지목, 이용 상황, 지형지세 등을 조사하여 기재하는 것으로 토지가 격비준표 및 개별공시지가 산정에 활용되는 중요한 자료이다. 그러므로 적정 공시지가 산정과 공신력 확보를 위해서는 무엇보다 정확하고 일관성 있는 토지특성조사가 필요하다.

그러나 현행 토지특성조사는 매년 전국 약 3,200만 필지를 조사하는 방대한 업무임에도 불구하고 매우 한정된 조사 인력과 짧은 조사 기간에 조사하고 있기 때문에 매년 조사의 불확실성 문제가 지적되고 있다.

특히, 토지 형상의 경우 조사 기준의 모호성으로 인해 조사자¹⁾에 따라 조사 결과가 달라지는 문제가 발생하고 있어 조사과정에 대한 방법론적 개선이 요구되어 왔다.

토지의 형상은 토지의 경계를 바탕으로 파악되는 개별 토지에 대한 고유 특성이기 때문에 토지의 분할 또는 합병이 발생하지 않는다면, 지적도에 나타난 본연의 형태가 토지의 형상으로 구분될 수 있다. 이는 필지 경계를 GIS(Geographic Information System)상의 폴리곤(polygon) 형태로 인식할 경우 토지의 기하학적 형태 정보로 인식할 수 있음을 의미한다.

즉, 토지 형상을 토지의 기하학적 구조 및 유효 면적 등의 개념이 결합된 하나의 패턴으로

분류할 수 있다고 가정한다면, 형상은 토지의 지오메트리(Geometry) 정보를 일련의 구조로 조합하여 분류별 특성의 패턴 분석 과정으로 해석할 수 있다는 논리가 성립된다.

이에 본 연구는 토지 형상 조사의 객관성과 일관성 제고를 위해 GIS를 활용하여 토지의 기하학적 특성 정보를 추출하고, 이러한 정보를 하나의 조합으로 나타낼 수 있는 데이터 마이닝 기법에 적용하여 해석하는 방안을 제시하고자 한다.

단, 본 연구는 다양한 토지 형상 유형의 객관적 분류와 분류기준 제시에 목적을 두고 있기 때문에 인근 토지와 가격 균형을 위한 형상 변경 등과 같은 특수한 상황 변수는 반영하지 않았다.

2. 연구내용과 방법

본 연구에서는 먼저 현행 토지형상 조사의 절차와 기준을 분석하여 문제점을 제시하고, 이를 개선할 수 있는 형상 조사 데이터 마이닝 기법을 탐색하였다.

적정 학습 알고리즘 탐색을 통해 형상 구분 알고리즘을 선정한 후 형상 조사 기준을 위한 데이터 웨어하우스 구축을 통해 선정된 데이터 마이닝 기법을 적용하여 최종적인 형상 추출 방식을 제안하였다.

연구지역은 서울시를 대상으로 하였으며, 대상 표본은 2016년 기준 서울시 표준지로 선정하였다.

형상 인식 결과에 대한 적합성 분석은 현행

1) 현재 토지특성조사의 평가주체는 감정평가사(표준지), 지자체 공무원(개별지)로 이원화되어 있음.

조사 체계에서 토지특성조사 정확도에 대한 절대적 기준이나 수치가 존재하지 않으므로, 현행 표준지의 토지 형상 조사 결과의 비교 분석을 통해 정확도와 타당성을 제시하였다.

연구에 활용된 데이터는 2016년 기준 연속지적도와 토지특성정보를 활용하였으며²⁾, 데이터의 입·출력 및 각 토지의 기하학적 특성은 QGIS 2.18 Las Palmas의 concave hull 플러그인을 활용하여 추출 하였다. 데이터 마이닝 분석은 뉴질랜드 와이카토(Waikato) 대학에서 개발하여 데이터 마이닝 분야에서 활발하게 사용되고 있는 오픈소스 프로그램 Weka 3.8.1을 활용하였다.

II. 이론 및 선행연구 검토

1. 토지특성조사(형상) 지침 및 문제점

1) 토지특성조사 지침

토지특성조사는 약 3,200만 필지(국세 또는 지방세의 부과대상토지)³⁾에 대한 토지특성을 조사하는 것으로, 토지가격형성에 중요한 요인으로 작용하고 있는 토지 관련 자료의 정보요인으로 가치가 있는 것을 말한다.

토지특성 조사는 비교표준지의 토지특성과의 비교를 통해 가격배율을 도출하고 개별공시지가

를 산정하는 기초자료이며, 이는 매년 1월 1일(공시기준일) 현재를 기준으로 한다. 조사는 토지(임야)대장, 건축물 대장 등 각종 공부조사 및 지가현황도면과 현장 확인을 통해 조사하고 있다.

조사 항목은 크게 공적규제 사항과 지형지세, 도로조건, 유해시설접근성 등 23개 항목으로 구성되어 있다.

형상 조사는 ‘공시지가 조사산정 지침’에 제시된 8개 형상 유형⁴⁾에서 가장 비슷한 것을 택하여 조사하며, 일단지내 토지는 일단지 전체의 형상을 기준으로 조사하고 있다. 이 중 다각형의 경우 토지의 효용도를 감안하여 형상을 파악하며, 각지에 위치하여 각각 정리된 토지의 경우가각이 없는 것으로 판단하여 조사한다.

현재 형상조사는 토지의 형상이 모호하여 형상을 구분하기 곤란한 경우에 관련 전문가와 협의하여 결정하도록 되어 있다. 또한, 구획정리가 진행된 지역 외에는 사각형 형태와 유사한 경우에는 정방형 또는 장방형으로 판단하며, 부정형 토지의 경우 형상으로 인해 최유효이용에 상당한 제약을 받는 경우로 한정하며, 택지의 경우 최유효이용에 장애가 될 정도로 심한 경우에 한해 부정형이나 자루형으로 구분하고 있다⁵⁾.

2) 형상조사의 문제점

현행 형상조사 상 가장 큰 문제점은 형상 구분

2) 각각의 데이터는 국토교통부 국가공간정보포털 오픈마켓(<http://market.ndsi.go.kr>)에서 다운로드 가능함.

3) 국·공유지의 경우 도로 등 공공용지는 제외할 수 있음.

4) 정방형, 가장형, 세장형, 사다리, 삼각형, 역삼각형, 부정형, 자루형.

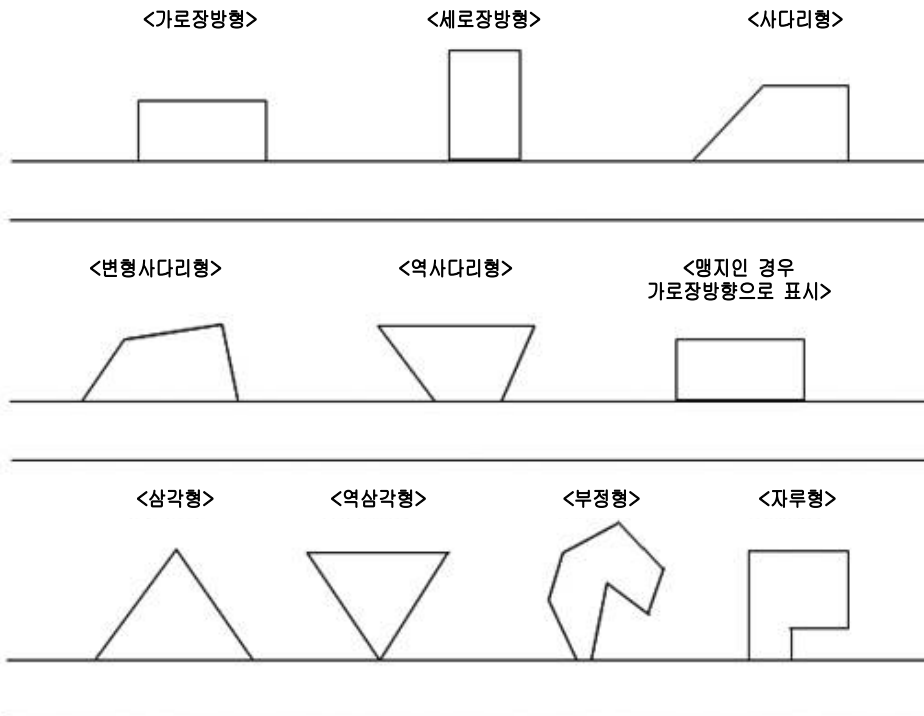
5) 택지에 있어 최유효이용에 상당한 제약을 받는 경우의 예로 형태가 띠모양 토지이거나 지적선의 굴곡이 심하여 통상의 건축에 제약이 있는 경우로 설정함.

기준의 모호성으로 볼 수 있다. 현행 형상 구분은 <표 1>과 같이 정방형, 가장형, 세장형, 사다리형, 삼각형, 역삼각형, 부정형, 자루형으로 구분하고 있다. 그러나 각각의 구분 기준으로는 해

당 필지가 사다리형인지 부정형인지 판단하기 어려운 경우가 많다. 또한, 토지특성조사 지침에서 제시되어 있는 <그림 1>의 예시와 같이 정형화된 토지의 형상은 소수에 불과하여 조사자에

<표 1> 형상 구분 기준

구분	구분 기준
정방형	정사각형 모양으로서 긴변과 짧은변의 길이가 비슷한 모양의 토지(양변의 길이 비율이 1:1.1 내외인 토지)
가장형	장방형의 토지로 넓은 면이 도로에 접하거나 도로를 향하고 있는 토지(가로장방형)
세장형	장방형의 토지로 좁은 면이 도로에 접하거나 도로를 향하고 있는 토지(세로장방형)
사다리형	사다리꼴 모양의 토지(변형 사다리형, 다각형의 불규칙한 형상이나 그로 인하여 최유효 이용에 상당한 제약을 받지 않는 토지 포함)
삼각형	삼각형의 토지로 그 한면이 도로에 접하거나 도로를 향하고 있는 토지
역삼각형	삼각형(역사다리형을 포함)의 토지로 꼭지점 부분이 도로에 접하거나 도로를 향하고 있는 토지
부정형	불규칙한 형상으로 인하여 최유효이용에 상당한 제약을 받는 다각형 또는 부정형의 토지
자루형	출입구가 자루처럼 좁게 생긴 토지



<그림 1> 형상 구분 예시

따라 서로 다른 판단을 할 수 있는 문제가 있다.

또한, 형상 조사의 문제는 ‘~ 내외인 토지(정방형)’, ‘~ 상당한 제약을 받지 않는(사다리형)’, ‘~ 불규칙한 형상(부정형)’ 등 매우 정성적으로 정해져 있는 조사 기준의 영향도 무시할 수 없다. 이러한 조사기준의 모호성은 조사자에 따라 동일한 형상의 필지에 대해 다른 조사 결과로 도출되는 문제를 야기하고 있다.

2. 선행 연구 검토

토지의 형상 조사 관련 선행 연구는 크게 두 가지로 나눌 수 있다. 첫째, 토지특성조사의 자동화와 객관화와 관련하여 형상 정보를 추출하는 연구와 둘째, 토지적성평가 또는 도시계획 과정에서 개발가능지 정형화를 위한 연구로 구분된다. 이에 대한 구체적인 내용은 다음과 같다.

먼저 토지특성조사 자동화와 관련된 연구로는 GIS와 위성영상 등을 활용하여 토지특성정보를 추출하는 연구가 주를 이루고 있다.

박성규 등(1998)은 GIS를 활용하여 토지특성 조사를 자동 수행할 수 있는 알고리즘 및 자동 산정 프로그램을 개발하였다. 형상 조사의 경우 필지의 경계선 호를 생성하여 호의 길이 시작_끝점 상의 식별자를 추출하여 시작점에서 각 점으로 방향각을 계산하여 변수의 수, 변의 변장비, 변과 변사이의 내각을 추출하였다.

박수홍 등(2003)은 수치지적도를 이용한 개별 공시지가 토지특성 자동 추출 방법의 개발 및 적용 연구에서 각 필지의 일반화 과정 이후 각

점(point)의 수를 파악 하여 점간 길이 추출 및 좌표 취득하고 좌표값을 기반으로 수선의 길이 추출을 추출하는 과정을 통해 각 필지의 포인트와 내각 등의 정보를 획득하고, 이를 기반으로 토지 형상 정보를 추출하였다.

성춘자·박재국(2007)은 항공사진과 수치지적도를 중첩하여 육안으로 토지의 형상을 구분하는 방법을 제시하였으며, 정우수 등(2015)은 무인항공기 정사영상과 지적도를 중첩하여 육안으로 형상을 확인하는 방법을 제시하였다.

토지적성평가 및 개발가능지의 정형화와 관련하여 토지의 형상에 대한 연구를 진행한 사례는 김정훈 등(2006)이 GIS를 활용하여 토지형상 지표(Land For Index) 및 조정된 선형성지표(Adjusted Related Circumscribing Circle Index)를 필지별로 산출하고 이를 기준으로 각 토지형상지표 표준영역을 벗어난 토지를 부정형 토지로 판단하는 연구를 진행한 사례가 있다. 또한, 이인성 등(2009)은 필지형상이 개발밀도에 미치는 영향 분석 연구에서 형태지수(Shape Index)를 기준으로 정형성과 부정형성을 추출하여 정형형태의 토지와 부정형의 토지를 추출하는 방법을 제시하기도 하였다.

해외의 경우 Demetriou, D et al.(2013)이 필지의 둘레, 넓이, 내각, 외각 점 수 등을 활용하여 필지 형태 지수(PSI, Parcel Shape Index)를 개발하여 각 필지의 형태에 따라 지수 점수를 부여하여 구분하는 방법을 제안하였다.

이러한 선행연구들은 경우에 따라 항공사진 등의 영상 정보를 활용하여 육안으로 형상을 구

분하는 방법도 제시하고 있지만, 대부분 GIS를 활용하여 필지의 형태(형상)를 필지의 기하학적 특성을 이용하여 구분한다는 공통점이 있다. 그러나 각 방법론은 대상토지의 형상을 단순한 기하학적 정보에 의거하여 판단하고 있거나, 연구자가 정한 주관적 기준에 따라 분류체계를 적용하고 있어 현행 조사기준과는 부합되지 않는 한계가 있었다. 또한, 매우 복잡한 연산과정을 통해 분류체계를 적용하고 있어 효율성 측면에서 한계가 나타났다.

박수홍 등(2003)의 연구가 본 연구에서 추구하는 목적과 가장 유사한 연구로 볼 수 있으나, 형상 추출 전 필지의 일반화 과정을 통해 임의의 버텍스(vertex)를 삭제하는 과정을 거치기 때문에 필지의 고유 형상 정보를 왜곡하는 문제와 함께 부정형 토지가 과다하게 추출되는 한계가 있었다.

이에 본 연구에서는 필지의 기하학적 특성을 기반으로 형상정보를 추출하되 토지 본연의 형상 정보 왜곡을 최소화하는 방법론을 적용하는 동시에 현행 조사 기준을 최대한 만족시킬 수 있는 형상 조사 방법을 제시하고자 한다.

III. 토지 형상 인식 방법 설정

1. 형상 인식 알고리즘 설정

데이터 마이닝은 1R(1-Rule), 베이저안망(Bayesian Network), 인공신경망(Artificial Neural

Network), 의사결정트리(Decision Tree) 등 다양한 방법론이 제시되어 있다.

데이터 마이닝 분야에서 형상 조사(토지특성 조사)는 여러 가지 조건을 고려하여, 그 결과를 도출하는 개념학습(concept learning)으로 분류할 수 있다. 즉, ‘정방향’, ‘장방향’ 등의 클래스를 가지는 분류(classification)의 분야로 볼 수 있다.

이러한 측면에서 1R과 베이저안망, 인공신경망 알고리즘은 각각 하나의 속성 사용, 사전 지식의 부재 및 변수 간 조건부 독립성 가정, 블랙박스 방법론이라는 문제가 존재하고 있어 토지 형상 조사를 인식시키기 위한 데이터 마이닝으로는 적절하지 않다. 반면, 의사결정트리는 범주화를 통해 형상을 분류하는 하나의 규칙을 표현할 수 있어 형상과 같이 패턴 분류의 문제를 가지는 데이터를 처리하는데 효과적이다(김종윤 · 박수홍, 2004).

따라서 본 연구에서는 토지 형상을 추출하는 기법으로 의사결정트리를 설정하여 다양한 변수들이 형상을 결정하게 되는 과정과 규칙을 도출하고자 한다.

의사결정트리 알고리즘의 대표적 방법은 〈표 2〉와 같이 CART(Classification And Regression Tree), CHAID(Chi-squared Automatic Interaction Detection), C4.5가 있다. 그러나 본 연구에서는 입력변수가 3개 이상일 때 각 범주 만큼 가지 분리를 수행할 수 있어 다양한 가지 분리를 할 수 있는 장점과 함께 토지 형상 분류와 같은 객체 분류 문제를 상대적으로 처리하는데 용이한 C4.5 알고리즘을 주 모델로 설정하였다.

〈표 2〉 의사결정트리 알고리즘 비교

구분	CART	CHAID	C4.5
종속 변수	All type	All type	Nominal
독립 변수	All type	All type	All Type
분리 기준	Gini Index	Chi-square	Entropy Index
분리 개수	이지분리	다지분리	다지분리
가지 치기	포함	미포함	포함

2. 입력 변수 설정

현행 형상 조사 기준은 장방형에 대한 구체적인 구분 기준(세장비 1:1.1 내외)만 언급되고 있을 뿐 최유효이용과 도로 접면 비율, 불규칙한 형상 등을 판단할 수 있는 구체적인 구분 기준은 제시되지 않고 있다.

일반적으로 토지 형상은 토지의 효율성과 활용도 등을 나타내는 척도로 건부지 활용 시 건폐율 및 용적률의 적용조건으로 작용한다(김현영, 2013).

따라서 현행 토지특성조사에서의 토지 형상은 토지의 정형성과 복잡성을 포괄하고 있어 특정 조건으로 구분 기준을 규정하기는 어려운 것이 현실이다. 그러나 토지의 형상을 효율성과 활용도 측면에서 바라볼 경우 정형적인 형태로서 복잡하지 않고, 단순하며 길게 늘어져 있는 형태가 아닌 집중된 형태가 효율성과 활용도 측면에서 유리한 토지라고 볼 수 있다(오윤표·임재문, 2004).

기준 문헌을 살펴볼 때 이러한 토지 형상의 관련 지표는 전통적으로 조경생태학(Landscape

Ecology) 분야에서 CI(Compactness Index)와 SI(Shape Index) 등의 개념이 널리 적용되고 있다. 그러나 토지특성 조사의 복잡한 형상 개념을 인식하기 위해서는 필지 형상별로 표준화 가능 여부와 토지 가격 비준율에 적용할 수 있는 방법론적 설계가 필요하다⁶⁾.

이에 본 연구에서는 토지특성 조사에서 형상 조건의 복잡성을 인식할 수 있는 동시에 토지형상 분류의 직관성과 일관성을 기하고자 선행연구에서 제시된 형상 조사를 위한 변인들을 지적도상 필지의 기하적 정보로 추출하여 다음과 같이 입력변수로 설정하였다.

형상지수(SI, Shape Index)는 토지 형상 변형 비교값으로 해당 토지가 정형의 토지에서 얼마나 변형되어 있는지를 지수화 한 값이다. 형상지수는 토지가 복잡한 형태 일수록 높은 값을 나타내며, 정형의 토지에 가까울수록 낮은 값을 가진다(McGarigal and Marks, 1994).

$$SI = 0.25 \times (\text{필지둘레} / \sqrt{\text{필지면적}}) \quad (\text{식 } 1)$$

정형지수(STI, Standard Index)는 필지면적 대비 최소외접사각형의 면적 값으로 0~1의 값을 가진다. 1에 가까울수록 정방형의 형태를 가지고 0에 가까울수록 부정형에 가까워진다(최진호 등, 2015).

$$STI = \text{필지면적} / \text{최소외접사각형 면적} \quad (\text{식 } 2)$$

6) CI 지수와 LFI 지수 모두 정형성과 형태에 대한 상대적 값을 보여주지만 절대치로 해석하기는 어려운 문제가 있음(이인성 등, 2006).

세장비(WR, Widthdepth Ratio)는 필지의 폭(width) 대비 너비(length)의 값으로 형상의 세장비가 동일하면 1에 가까우며 폭과 너비의 차이가 클수록 0에 가까워진다(최진호 등, 2015).

$$WR = \frac{\text{최소외접사각형 폭}}{\text{최소외접사각형 너비}}$$

(식 3)

이들 지수는 <그림 2>와 같이 필지형태의 정형성, 비정형성 등을 나타내는 상대적인 값들이지만 하나의 값으로 형상 구분을 위한 절대적 기준으로 활용하기는 어렵다. 따라서 각각의 지수가 가지는 특징을 조합하여 필지 형태유형을 구분하는 규칙 생성의 변수로 투입하였다.

2016년 기준 서울시 표준지는 27,389필지이나 일단지 토지에 대한 구분 정보가 명확하지 않고 지적도와 비매칭 되는 필지가 존재하여 이를 제외한 나머지 26,433필지만을 실험 대상으로 선정하였다.

2016년 기준 서울시 표준지의 형상 분포는 정방형 4,507필지, 가장형 4,271필지, 세장형 7,820필지, 사다리형 7,955필지, 삼각형 124필지, 부정형 1,498필지, 자루형 258필지이다. 본 연구에서는 이중 필지의 형상 정보 외 도로 조건이 관여되는 요소를 1차적으로 제외하기 위해 장방형(가장형, 세장형), 삼각형(삼각형, 역삼각형), 부정형(부정형, 자루형)으로 통합하였다(<표 3>).

IV. 실험 및 분석








1. 실험

본 연구의 실험 적용은 서울시 표준지를 대상으로 하였으며, 이를 토지를 대상으로 GIS와 C4.5 알고리즘을 적용하여 토지형상을 분류하였다.

<표 3> 실험 대상 형상 분포

	빈도(필지)	비율(%)
정방형	4,507	17,1
장방형	12,091	45,7
사다리형	7,955	30,1
삼각형	124	0,5
부정형	1,756	6,6
계	26,433	100,0

<그림 2> 필지 형상 관련 지수 검토 예

도형							
SI	1,194	1,166	1,000	1,060	1,102	1,348	1,120
STI	0,500	1,000	1,000	0,792	0,625	0,577	0,800
WR	0,544	0,323	1,000	0,510	0,494	0,622	0,980

표준지의 형상 조사는 담당 감정평가사가 필지의 형태, 필지의 세장비, 필지의 최유효성 등을 고려하여 형상을 판단하게 된다. 이러한 패턴 또는 규칙성을 찾아내기 위해 STI, SI, WR 지수를 바탕으로 형상을 판별하는 규칙을 도출하였다. 도출된 규칙을 통해 추정된 형상 분류 결과는 표준지 형상 정보와의 교차 검증(10중 교차검증, 10 fold cross validation)을 진행하였다. 검증 결과는 정확도와 재현율로 분류의 정확도를 비교하였다.

2. 결과 분석 및 고찰

1) 결과 분석

C4.5 알고리즘을 통해 학습된 형상 분류결과와 혼합 매트릭스(confusion matrix)는 <표 4>와 같다.

서울시 표준지 26,433필지에 대한 형상 추정 결과 일치율(correctly classified instances)은 81.6%(21,570필지)로 높은 수준의 정확도를 보

였다.

각각의 형상 구분별 정확도(Precision Rate) 81.4%, 재현율(Recall Rate) 81.6%, TP Rate(True Positive Rate) 81.6%, FP Rate(False Positive Rate) 9.3%로 나타났다. 따라서 실험을 통해 학습된 규칙이 형상 정보를 인식하는데 상당 수준의 연관관계가 있음을 알 수 있다. 또한, 표준지 형상 정보와 학습을 통해 추정된 형상의 확률적 일치율을 나타내는 Kappa 계수가 0.72(0.6~0.8)로 나타나 상당 수준(substantial)의 일치도가 나타났다.

반면, 기존 조사자의 형상 정보와 일치하지 않는 비율이 약 18.40%(4,863필지)로 나타났다. 이는 학습된 규칙이 형상을 나타내는 일반적 규칙으로 설명되지 않거나, 모델을 구축하는데 활용된 독립 변수가 토지 형상을 완벽하게 설명할 수 없다는 의미로 해석할 수 있다.

그러나 실험 결과를 통해 생성된 tree가 총 139개의 node로 구성된 것으로 볼 때 현재 형상 조사 결과 속에는 현행 분류 체계와는 다소 이

<표 4> 실험 대상 형상

표준지→ ↓ 추정결과	정방형	장방형	사다리형	삼각형	부정형	계
정방형	3,544	711	247	0	5	4,507
장방형	479	10,628	948	0	36	12,091
사다리형	218	850	6,488	16	383	7,965
삼각형	0	0	47	58	19	124
부정형	47	110	737	10	852	1,756
계	4,288	12,299	8,467	84	1,295	26,433

질적인 유형이 혼재되어 있는 것으로 추측할 수 있다. 예를 들어 부정형으로 분류되어 있는 필지들 사이에는 폭이 좁거나 긴 사다리형 필지, 폭이 넓고 짧은 형태의 사다리형 필지, 굴곡이 심하게 나타나 부정형에 가까운 사다리형 필지 등이 혼재되어 있는 것으로 볼 수 있다.

2) 고찰

학습된 토지형상 구분 규칙을 활용하여 형상을 추출한 결과 기존 조사 결과와 약 81% 일치하는 것으로 나타났다. 이는 토지 형상을 구분하기 위한 학습 규칙이 토지형상을 추출하는 과정을 완벽하게 설명하고 있지 못하고 있음을 의미한다. 이는 학습된 규칙이 실험 대상 전체를 위한 일반적 규칙이기도 하지만, 한편으로는 현행 형상 조사가 매번 일정한 규칙에 의해 선정되고 있지 않음을 의미하기도 한다. 또한, 현재의 형상 조사가 조사자 각자의 주관성이 상당부분 반영되어 조사되고 있음을 의미하기도 한다.

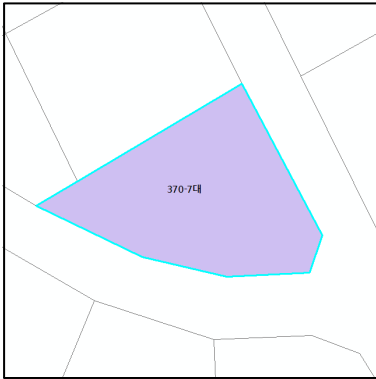
그러나 무엇보다 가장 큰 영향요인은 조사자가 판단하기에 어려운 부분이 존재하기 때문으로 볼 수 있다. 몇 가지 사례를 살펴보면 먼저 <그림 3>과 같이 현행 조사 결과에서는 장방형으로 조사 되었지만, 추출 결과는 정방형으로 추출된 필지를 볼 수 있다. 해당 필지는 가장 긴 변의 길이 17.98m, 짧은 변의 길이 16.45m 이기 때문에 정방형 기준(1:1.1)의 조건 비율을 만족함으로 정방형 토지로 구분 되는 것이 합당하다. 그러나 이러한 길이 비율을 사람의 목측으

로만 판단하기에는 한계가 분명히 존재할 것이다. 따라서 이러한 부분을 조사하는 과정에서 오류가 발생할 수 있다.

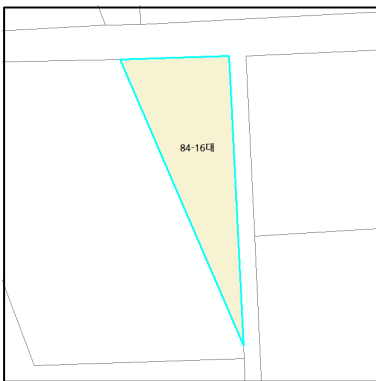


〈그림 3〉 형상 구분 오류 예시
(장방형 → 정방형)

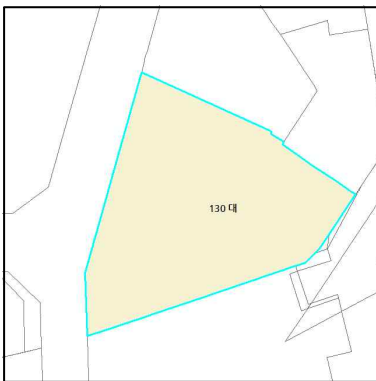
〈그림 4〉는 기존에 삼각형으로 조사되었으나, 추출 결과 사다리형으로 조사된 사례이다. 해당 필지는 꼭지점의 개수가 3개 이면서 내각의 합이 180° 라는 개념의 삼각형과는 거리가 멀다. 현행 조사기준에 따라 역사다리형 개념을 포함하여 삼각형 토지의 개념을 확장시키더라도 <그림 5>의 삼각형(조사: 삼각형, 추출: 삼각형)과 다르게 조사되어 있다. 오히려 <그림 6>과 같이 사다리형(사다리형→사다리형)의 토지에 더 유사한 형태로 나타나고 있다. 이러한 분류 오류는 수많은 토지를 조사하는 과정에서 일관성을 유지하는데 제약이 존재하는 것으로 보인다.



〈그림 4〉 형상 구분 오류 예시
(삼각형 → 사다리형)



〈그림 5〉 삼각형 토지 예시
(삼각형 → 삼각형)

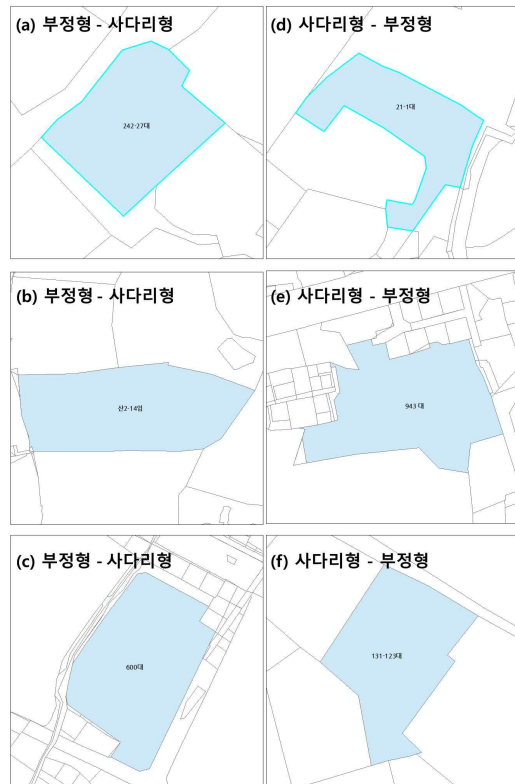


〈그림 6〉 사다리형 토지 예시
(사다리형 → 사다리형)

현행 조사 과정에서 가장 많은 문제가 발생하고 있는 사다리형과 부정형의 구분에서도 다양한 사례를 찾아볼 수 있다.

〈그림 7-a〉부터 〈그림 7-c〉는 기존 부정형으로 조사되었지만 사다리형으로 추출된 사례이며, 〈그림 7-d〉부터 〈그림 7-f〉까지는 사다리형에서 부정형으로 추출된 필지 사례이다.

해당 사례를 통해 볼 때 부정형과 사다리형의 경계가 서로 상반되거나 〈그림 7-c〉와 〈그림 7-f〉와 같이 부정형과 사다리형을 각각 구분하기 모호한 상황이 나타나고 있다.



〈그림 7〉 형상 구분 오류 예시
(부정형 → 사다리형 : a, b, c)
(사다리형 → 부정형 : d, e, f)

사다리형과 부정형의 문제는 불규칙한 형상과 최유효 이용 원칙이라는 정성적 판단기준으로 인해 조사 결과가 상반되게 나타날 수 있다. 즉, 조사 기준의 모호성으로 인해 조사자의 판단이 서로 상반되게 나타날 수 있음을 의미한다.

이와 같이 토지 형상 조사는 전술된 여러 사례와 같은 문제로 인해 일관적이고 객관적으로 이루어지기 어렵다.

연구의 서두에서 밝힌바와 같이 특성 조사의 비밀관성은 지가 불균형의 문제를 초래하는 동시에 토지가격 비준율의 오류를 발생시킨다. 또한, 토지특성정보 공개 과정에서 수많은 민원을 초래할 여지를 가지고 있다.

본 연구를 통해 추출된 형상 구분 규칙을 하나의 통일된 식으로 규정하여 자치단체별로 일치율을 살펴보면, <표 5>와 같이 지역 간 상당한 차이가 발생하고 있는 것을 볼 수 있다⁷⁾.

<표 5> 자치단체별 형상 추출 일치율 비교

	표준지수	일치율(%)
A구	1,090	88.7
B구	910	88.5
C구	992	87.3
(중략)
L구	1,164	83.0
M구	614	82.9
N구	1,141	82.7
(중략)
W구	1,242	76.0
X구	974	74.6
Y구	1,674	74.4

이러한 차이는 토지 형상 조사 과정에서 조사자마다 기준이 다르게 적용되어 일관성이 유지되지 않고 있는 것으로 볼 수 있다. 이러한 문제는 비전문가인 지자체 공무원이 조사하는 개별지로 확대할 경우 더욱 심각한 문제로 나타날 수 있다. 즉, 지자체별 표준지 조사의 비밀관성과 비통일성의 문제 뿐만 아니라 지자체 내 표준지와 개별지(지자체 공무원 조사)간의 문제도 함께 발생할 수 있음을 의미한다.

현행 토지특성조사는 조사 결과에 대한 명확한 해답을 구하기 어렵다. 이는 현재까지 연구된 대부분의 연구에서 지적하고 있는 문제이기도 하다. 즉, 토지특성 조사의 명확한 규정이 없기 때문에 어떠한 특성으로 분류되어 조사되어도 이를 옳고 그름의 잣대로 판단하기 어렵다는 것이다.

본 연구에서 기존의 표준지 형상조사 데이터로부터 학습된 형상조사 규칙은 애매모호한 형상 조사 요령을 정형화된 규칙의 형태(그림 8)로 전환함으로써, 형상 조사를 보다 객관적이며 효율적으로 확보할 수 있다. 특히, 2만여 필지를 분류하는데 5분 내외의 시간이 소요됨을 고려할 때 현행 조사 체계의 문제점을 개선하는데 효과적으로 적용될 수 있을 것으로 판단된다.

VI. 결론 및 연구의 한계

본 연구에서는 GIS와 데이터 마이닝 분석 기

7) 자치단체별 표준지 조사자와 연결될 수 있어 지자체명은 표기하지 않음.

```

Source Code
// Generated with Weka 3.8.1
//
// This code is public domain and comes with no warranty.
//
// Timestamp: Sun Mar 05 18:19:33 KST 2017

package weka.classifiers;

import weka.core.Attribute;
import weka.core.Capabilities;
import weka.core.Capabilities.Capability;
import weka.core.Instance;
import weka.core.Instances;
import weka.core.RevisionUtils;
import weka.classifiers.Classifier;
import weka.classifiers.AbstractClassifier;

public class WekaWrapper
    extends AbstractClassifier {

    /**
     * Returns only the toString() method.
     *
     * @return a string describing the classifier
     */
    public String globalInfo() {
        return toString();
    }

    /**
     * Returns the capabilities of this classifier.
     *
     * @return the capabilities
     */
    public Capabilities getCapabilities() {
        weka.core.Capabilities result = new weka.core.Capabilities(this);

        result.enable(weka.core.Capabilities.Capability.NOMINAL_ATTRIBUTES);
        result.enable(weka.core.Capabilities.Capability.NUMERIC_ATTRIBUTES);
        result.enable(weka.core.Capabilities.Capability.DATE_ATTRIBUTES);
        result.enable(weka.core.Capabilities.Capability.MISSING_VALUES);
        result.enable(weka.core.Capabilities.Capability.NOMINAL_CLASS);
        result.enable(weka.core.Capabilities.Capability.MISSING_CLASS_VALUES);

    }

    /**
     * only checks the data against its capabilities.
     *
     * @param i the training data
     */
    public void buildClassifier(Instances i) throws Exception {
        // can classifier handle the data?
        getCapabilities().testWithFail(i);
    }

    /**
     * Classifies the given instance.
     *
     * @param i the instance to classify
     * @return the classification result
     */
    public double classifyInstance(Instance i) throws Exception {
        Object[] s = new Object[i.numAttributes()];

        for (int j = 0; j < s.length; j++) {
            if (!i.isMissing(j)) {
                if (i.attribute(j).isNominal())
                    s[j] = new String(i.stringValue(j));
                else if (i.attribute(j).isNumeric())
                    s[j] = new Double(i.value(j));
            }
        }

        // set class value to missing
        s[i.classIndex()] = null;

        return landshp.classify(s);
    }

    /**
     * Returns the revision string.
     *
     * @return the revision
     */
    public String getRevision() {
        return RevisionUtils.extract("1.0");
    }
}

```

〈그림 8〉 형상 추출 규칙 쿼리

법 중 하나인 C4.5을 활용하여 토지 형상을 추출하는 연구를 진행하였다. 이는 현행 표준지 형상 정보를 바탕으로 지적도 상 필지의 형상

정보를 수치화하여 일반적인 규칙으로 학습한 후, 이를 통해 토지 형상을 추출하는 방법을 활용한 것이다.

연구 대상인 서울시의 표준지 형상을 학습된 규칙으로 추출하여 교차 검증한 결과 전체 일치율 81.6%, Kappa 지수 0.72가 나타나 실험 대상 26,433 필지 중 21,570 필지가 기존 형상 조사 결과와 일치하게 나타나는 것을 확인할 수 있었다.

학습된 형상 추출 규칙은 조사자의 현장조사 결과가 아니라 컴퓨터에 의해 학습된 정형화된 결과라는 것을 고려하면 상당 수준의 연구 목적을 달성했다고 볼 수 있다. 특히, 연구 대상 전체를 학습하여 결과를 도출하는 과정이 5분 내외로 소요된 점과 현행 표준지 조사의 애매모호함을 하나의 정형화된 규칙으로 도출했다는 점에서 연구의 의의를 찾아 볼 수 있다. 또한, 이러한 추출 과정에서 모두 오픈 소스와 오픈 데이터를 활용했다는 점에서 예산의 절감 효과도 기대할 수 있다.

그러나 본 연구가 실제적 효용을 가지기에는 몇 가지의 문제점과 한계를 가지고 있었다. 먼저 학습을 위한 대상 정보가 불규칙적인 결과를 내포하고 있었기 때문에 학습 과정에서도 이러한 불규칙성이 함께 학습되어 반영되어 있다는 것이다. 또한, 현행 토지특성조사가 하나의 제도라는 점을 감안할 때 각각의 결과 도출 과정을 하나의 규칙으로 정의할 수 없는 문제점이 있었다. 이러한 문제는 현행 조사 규정이 보다 객관화되고 수치화할 경우 학습을 위한 규칙을 보다 명확히 할 수 있어 개선할 수 있을 것이다.

본 연구에서 제시한 GIS와 데이터 마이닝 기반 토지 형상 인식은 조사의 효율성과 객관성을 개선시킬 수 있을 것이며, 고저, 향 등 다른 특성 항목에서도 적용할 수 있는 기초자료가 될 것으로 판단된다. 그러나 데이터 마이닝에는 본 연구에서 활용된 C4.5 외에도 다양한 방법이 존재하기 때문에 향후 다양한 알고리즘을 적용하는 후속 연구가 필요할 것으로 보인다.

참고문헌

1. 김정훈 · 조춘만 · 채명기, 2006, 「GIS 기반 개발가능지 정형화를 위한 토지형상기준 적용에 관한 연구」, 『한국측량학회지』, 24(1): 99-110.
2. 김종윤 · 박수홍, 2004, 「의사결정트리를 이용한 개별 공시지가 비교표준지의 자동 선정」, 『한국지리정보학회지』, 7(1): 9-19.
3. 김현영, 2013, 「필지형상 변경에 따른 토지특성 및 가격변화 분석」, 서울시립대학교 대학원 석사학위논문.
4. 박성규 · 임승현 · 박정남 · 조기성, 1998, 「GIS를 이용한 토지 특성의 자동추출에 관한 연구」, 『한국지형공간정보학회지』, 6(2): 59-67.
5. 박수홍 · 홍성언 · 김현석, 2003, 「수치지적도를 이용한 개별공시지가 토지특성 자동 추출 방법의 개발 및 적용」, 『국토계획』, 38(5): 61-77.
6. 성춘자 · 박재국, 2007, 「개별공시지가산정을 위한 토지특성조사에 GIS 공간분석기법의 적용」, 『한국지형공간정보학회지』, 15(1): 31-38.
7. 오윤표 · 임재문, 2004, 「도시토지의 획자규모 및 최적형상결정 모형 구축에 관한 연구」, 『국토계획』, 39(6): 57-70.
8. 이인성 · 임상준 · 김충식, 2006, 「필지형상이 개발 밀도에 미치는 영향 분석 - 서울시 강동구 천호 · 암사 지구단위계획구역을 대상으로」, 『한국도시설계학회지』, 10(4): 151-162.
9. 정우수 · 정성혁 · 임노열 · 김광호 · 이승기 · 최석근, 2015, 「고품질 지형공간정보를 이용한 토지특성조사」, 『한국지형공간정보학회지』, 23(3): 57-67.
10. 최진호 · 심종원 · 김봉준, 2015, 「토지특성조사 개선방안 연구」, 『부동산 공시제도 선진화 방안 연구』, 한국감정원.
11. Demetriou Demetris · Linda M See · John Stillwell, 2013, "A Parcel Shape Index for Use in Land Consolidation Planning", *Transactions in GIS*, 17(6): 861-882.
12. Kevin Mcgarigal · Mark J Barbara, 2013, "Spatial Pattern Analysis Program for Quantifying Landscape Structure", FRAGSTATS V2.0, Oregon State Univ, Corvallis.

논문접수일 : 2017년 3월 13일

심사(수정)일 : 1차 2017년 5월 12일

게재확정일 : 2017년 5월 26일

국문초록

본 연구는 GIS와 C4.5 모델을 활용하여 토지특성(형상) 조사 과정에서 나타날 수 있는 주관 개입 문제와 불확실성을 개선할 수 있는 새로운 방법론을 제시 하였다. 서울시의 표준지 형상 조사 결과와 필지의 기하학적 특성 정보를 C4.5 모델에 적용한 결과, 학습된 규칙에 의해 추정된 형상 분류 결과는 기존 전문가의 조사 결과와 약 81.6%(전체 24,433필지 중 21,570필지) 일치 하는 것으로 나타났다. 또한, 우연적 또는 확률적 일치율을 설명하는 kappa 지수에서도 0.72로 나타나 상당한 수준의 정확도를 가지고 있음이 확인되었다. 이러한 실험 결과는 현행 조사의 모호성을 하나의 정형화된 규칙으로 도출했다는 점에서 의의가 있으며, 특히 조사자에 따라 달리 분류되어 나타나던 특성 정보를 일관성을 유지하며 조사할 수 있다는 점에서 개선 효과를 찾아 볼 수 있다. 현재 토지특성조사가 인력과 현장조사에 의존하여 비효율적 집행이 진행되고 있음을 감안한다면, 향후 토지특성조사에 이러한 컴퓨터 시뮬레이션 기법을 적용하여 발전시킨다면 보다 객관적이며 효율적인 토지특성조사가 될 수 있을 것으로 판단된다.

주제어 : 토지특성조사, 토지 형상, 데이터 마이닝