

토지특성 고저조사를 위한 공간정보 데이터 구축과 데이터 마이닝 분석

Spatial Information Data Construction and Data Mining Analysis for Topography Investigation of Land Characteristics

최진호¹⁾ · 김준현²⁾

Choi, Jin Ho · Kim, Jun Hyun

Abstract

The investigation of land characteristics is an important task for the calculation of officially land prices and standard comparison table of land price. Therefore, it should be done objectively and consistently. However, the current investigation system is mainly done by researcher's subjective judgment. Therefore, the objectivity and consistency of this investigation is not guaranteed and questionable. In this study, we first defined the problem by analyzing the current land topography investigation method. In addition, in order to investigate the land topography, the geometry of the parcel is quantified by spatial information and applied to the decision tree based method(C4.5) to produce the final result. This study intended to extract the parcel characteristics data of the topographic by the use of spatial information and to apply the information to the C4.5, there by suggesting a method for addressing the problems. The findings showed approximately 93.5% between the results of topography classification estimated with rules learned by C4.5

Keywords : Land Characteristics Investigation, Topography, Spatial Information, Data Mining

초 록

토지특성조사는 토지가격비준표 작성 및 표준지와 개별지의 특성차이 비교를 통한 지가 산정 과정에서 매우 중요한 과정이다. 따라서 토지특성조사는 최대한 객관적이고 합리적이며 일관성 있게 이루어져야 한다. 그러나 현재 토지특성조사는 지자체 공무원과 감정평가사의 경험이 상당수 반영되고 있기 때문에 객관성과 일관성을 보장하기 어렵다. 본 연구에서는 현행 토지특성조사 항목 중 고저의 조사방식을 분석하여 문제를 정의하고, 고저 분류를 위해 토지의 고저 정보를 공간정보 기술 기반으로 수치화하여 이를 데이터 마이닝 기법 중 하나인 C4.5을 적용하여 고저를 분류하는 방법을 제시하였다. 서울시의 표준지 고저 조사 결과와 필지의 공간정보를 C4.5 모델에 적용한 결과에서는 기존 감정평가사의 조사 결과 약 93.5% 일치 하는 것으로 나타났다.

핵심어 : 토지특성조사, 고저, 공간정보, 데이터 마이닝

Received 2019. 11. 20, Revised 2019. 12. 04, Accepted 2019. 12. 12

1) Public Announcement Research Department, Korea Appraisal Board (E-mail: k25830@kab.co.kr)

2) Corresponding Author, Member, Dept. of Geodetic and Cadastral Information, Taegu Science University (E-mail: kamcho78@tsu.ac.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

개별공시지가는 표준지 공시지가를 기준으로 개별지 토지특성과의 비교를 통해 산정된다. 그러므로 적정 개별공시지가 산정을 위해서는 유의미한 비교표준지 선정과 함께 정확하고 객관적인 토지특성조사가 필요하다. 토지특성조사는 지목, 면적, 용도지역·지구 등의 공적규제현황과 토지이용상황, 고저, 형상 등의 물리적 특성을 조사하는 것으로 1989년 공시지가 제도 도입 이후부터 현재까지 인력 기반의 현장 조사로 실시되고 있다.

그러나 토지특성조사는 현장조사에 의존하고 있는 한계로 인해 매년 많은 시간과 비용이 소요되고 있으며, 이마저도 조사 기준의 불명확성으로 인해 조사자에 따라 다른 결과가 도출되는 문제가 있다.

특히 특성항목 중 고저(지형지세)는 높낮이를 의미하는 고도와 토지의 기울어짐을 의미하는 경사가 중복되어 있어 조사 과정에서 이를 정확하게 구분하기 어려울 뿐만 아니라 ‘최유효이용’이라는 개념도 함께 고려되어 있어 조사방법에 대한 개선이 필요한 상태이다. 현행 고저의 조사는 해당 필지의 전반적인 경사도를 기준으로 간선도로 및 주위의 지형지세와 비교하여 조사하고 있다. 또한 간선도로가 주위의 지형지세보다 현저히 높거나 낮아 간선도로를 기준으로 하는 것이 부적절한 경우 또는 간선도로가 원거리에 있어 고저비교가 적절하지 않는 경우에는 주위의 지형지세를 기준으로 조사하도록 되어 있다 (Choi *et al.*, 2015).

이러한 문제에 기인하여 최근까지 고저 조사의 객관화 일환으로 GIS 공간분석 기능을 활용하는 연구가 진행되었으며 (Kim *et al.*, 2015), 수치지형도와 지적도를 활용하여 토지의 고저를 결정하는 방식까지 제안되었다(Lee and Koh, 2013). 또한 형상 항목에서 대해서는 데이터 마이닝을 활용하여 형상조사 객관화 방안을 제시한 연구까지 확장된 상태이다(Choi and Lee, 2017).

그러나 전술된 연구들은 GIS (Geographic Information System) 분석 등을 통해 객관적이고 일관성 있는 기준으로 고저 조사를 수행할 수 있는 장점은 있으나, 고저조사에 토지의 평균 경사만을 단일 변수로 활용하기 때문에 전체적인 지세로 필지의 고저를 판단하는 현재의 특성조사 과정과는 동떨어진 결과가 나타나는 한계가 있다. 특히 현재의 고저 분류기준이 ‘주변보다 현저히 낮다’ 등의 정성적 형태로 되어 있음에도 이를 보완하거나 정량적인 해석 부분이 부족하여 확장성을 가지지 못한 아쉬움이 있다.

현재 지가공시체계(표준-개별 방식)에서는 표준지와 개별지에 대해 동일한 기준으로 일관성이 있는 특성조사가 필수적이

다. 특히 지가 산정 과정이 표준지의 특성과 개별지의 특성 차이에 따라 토지 가격 비준율이 적용된다는 점에서 표준지의 조사 기준이 개별지에서 동일하게 적용되어야 균형적인 가격이 이루어질 수 있다.

따라서 본 연구에서는 조사자의 자의성을 최대한 배제하고 객관적 조사가 수행될 수 있는 고저 조사 방법을 제시하고자 하며, 이를 위해 고저 조사를 위한 경사도와 높이 정보를 공간정보 데이터 바탕으로 구축하고, 이를 저지, 평지, 완경사 등의 조사 분류 체계에 따라 분리 될 수 있도록 데이터 마이닝을 통해 규칙을 설정하도록 하였다.

2. 고저조사 방법의 현황 및 문제점

토지의 고저는 Table 1과 같이 표준지와 개별지 모두 간선도로를 기준으로 조사하고 있다. 여기서 간선도로는 국도, 지방도 및 대중교통수단이 하루 2회 이상 통과하는 현행도로를 의미한다.

조사과정에서 부득이하게 간선도로가 조사 대상 필지보다 현저하게 높거나 낮게 위치하고 있어 간선도로를 조사 기준으로 적용하기 어려운 경우 또는 조사 대상 필지와 비교 대상이 되는 간선도로가 비교적 멀리 떨어져 있어 비교 기준으로 적용하기에 적절하지 않은 경우에 주위의 지형지세를 기준으로 고저를 조사하고 있다. 또한 고저의 조사 및 판단이 애매한 경우에는 시·군·구와 미리 판단기준을 협의하여 결정하고 있다.

Table 1. Current topography investigation criteria

Type	Evaluation standard investigation
Low land	Very low elevation land than arterial road or surrounding ground
Flat land	Land of similar height to the adjacent arterial road or surrounding ground(or a land with little slope)
Mild slope	High elevation land than arterial road or surrounding ground(and slope less than 15 degree)
Steep slope	High elevation land than arterial road or surrounding ground(and slope 15 degree or higher)
High land	Very high elevation land than arterial road or surrounding ground

현행 고저조사는 간선도로와 주위의 지형지세를 기준으로 이루어지고 있다. 그러나 간선도로와 주위의 지형지세에 대한 경사의 정도와 높낮이의 변화는 일정하지 않다. 이로 인해 조사 기준점을 잡기 힘들고 특히 지형지세를 기준으로 하는 경우에

는 조사자에 따라 조사 기준을 다르게 설정할 수 있어 객관적인 조사가 어려운 것이 사실이다.

또한 현행 고저조사는 해당 토지를 간선도로 또는 주위의 지형지세와 비교하여 ‘현재지 낮은’, ‘비슷하거나’, ‘미미한’, ‘현재지 높은’ 등의 명확하지 않은 기준으로 비교하고 있고 완경사와 급경사를 구분하는 기준을 15°로 책정하고 있다고는 하지만 사람의 목측으로 경사도를 정확하게 측정하는 것은 사실상 불가능에 가깝다.

더욱이 표준지공시지가 조사·평가 업무지침에서는 토지의 고저는 최우효이용과 함께 해당 토지의 이용상황 뿐만 아니라 인근 토지들의 종합적이고 포괄적인 이용상황 등을 고려하여 전문가적 견해에서 판단하여야 한다고 명시하고 있어 경사 15°를 경사조사를 위한 적정 방법으로는 보지 않고 있다. 이는 조사자의 주관적 판단이 개입될 소지가 있음에도 측량 등을 통한 경사도 15°의 관측 방법은 비교 대상이 될 수 없음을 의미하는 것이다.

이러한 조사 여건과 한계로 인해 현행 토지특성조사는 조사자마다 다른 결과가 제시될 수 있는 가능성이 매우 높다. 결과적으로 이러한 조사결과의 차이는 민원인이 조사 결과에 이의를 제기할 경우 과학적인 근거나 산정근거를 제시하기 어려운 상황을 만들어 이해관계인을 설득하기 어렵게 할 수 있다.

특히 현재 공시지가 산정방식(표준-개별)에서는 토지특성조사의 오류는 공시지가 산정 오류로 직결되는 구조이기 때문에 특성조사의 정확성이 무엇보다 중요하다. 그러므로 어느 누가, 언제, 어디서든 분석과 조사를 실시하더라도 객관적이고 일관성 있는 결과가 나올 수 있는 체계가 필요하며, 객관적인 정보를 바탕으로 조사 결과가 도출될 수 있어야 한다.

3. 데이터 마이닝을 활용한 고저조사 설계

3.1 지대(지세)를 위한 범위 설정

토지의 고저는 해당 토지가 속한 지대의 전반적인 경사도를 기준으로 조사하고 있다. 고저의 조사를 해당 토지의 경사가 아닌 지대의 경사로 판단하는 것은 표준지와 그 주변 개별지가 유사한 형태로 조사되어야 지가산정이 균형 있게 이루어질 수 있는 현행 공시지가 산정 방식(표준-개별)의 특수성이 반영되었기 때문이다(Lee, 2018).

그러나 지대의 범위는 매우 주관적이기 때문에 조사자에 따라 매우 협소하거나 광범위한 범위로 설정될 수 있다. 또한 일반적인 방식과 같이 법정동과 같은 행정구역을 기준으로 정의하더라도 지역에 따라 매우 넓거나 좁은 범위로 설정될 수 있어 적합한 방법으로 보기 어렵다. 따라서 필지에 대한 주위의 지대를 정의하는 방법은 일정한 구획(또는 경계)으로 정의하기 보다는 당

해 토지를 기준으로 개별적으로 설정되는 것이 지역이나 유형에 따라 차이가 발생하지 않도록 하는 합리적 방법이 될 수 있다.

각 개별필지 마다의 지대를 구분하기 위해서는 거리(범위) 기준이 설정되어야 하는데 애초에 지대라는 개념은 대상과 그 인근지역을 설명하기 때문에 인근 지역이라는 개념에서 접근이 필요하다. 즉 정확한 고저조사를 위해서는 하나의 개별필지 중심이 아니라 인접지형과 연계되어 인접하고 있는 범위의 개념에서 접근되어야 한다.

토지에 있어서 고저나 경사와 같은 지형학적으로 연속적으로 구성되어 있는 토지특성을 정확하게 조사하기 위해서는 독립된 필지라는 개념에서 보다 확장하여 주변의 지리적, 물리적 현황인 지세 또는 지대와 같은 개념에서 조사되어야만 하나의 토지단위인 일필지에서 이웃하고 있는 인접필지와의 지형정보를 연속적으로 산정하거나 추정하여 정확한 특성값에 해당하는 지수를 반영할 수 있다.

현재 표준지 조사 및 평가 기준에서는 인근 지역을 당해 필지의 주변 500m 이내의 거리로 정의하고 있으며, 「역세권 장기전세주택 건립관련 지구단위계획 수립 및 운영 기준」에서는 승강장 중심에서 반경 250m를 1차 역세권으로 설정하고 있다. 또한 현재 비교표준지의 평균적인 범위가 약 230m내외인 점을 고려할 때 지대 설정을 위한 인근 거리는 대체적으로 약 250m 내외인 것으로 판단된다.

이에 본 연구에서는 여러 분야에서 정의하고 있는 인근 지역의 평균적인 범위(250m)를 기준으로 Fig. 1과 같이 지대의 기준을 설정하였다.

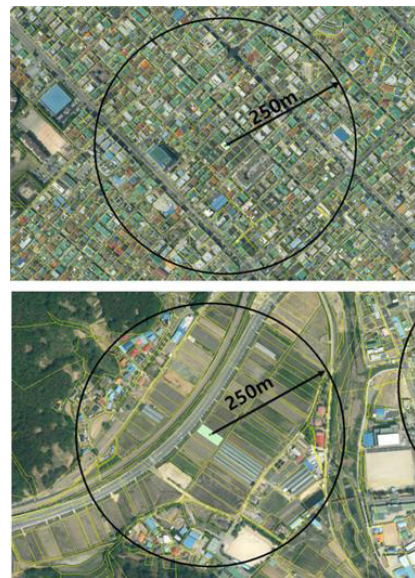


Fig. 1. Setting of zoning range

3.2 고저 구축을 위한 요인설정

고저 조사를 위한 가장 기초적인 정보는 경사와 고도(높이) 정보이다. 일반적으로 토지의 고도와 경사를 조사하는 방법은 실측(측량)과 인공위성을 활용하는 방법, GIS 이용 방법으로 구분될 수 있다.

실측에 의한 방법은 GPS (Global Positioning System) 측량, 토탈 스테이션 등의 기기를 이용하여 측량하는 방법으로 일반적으로 가장 많이 사용되는 방법이며 현지 지형을 가장 잘 반영할 수 있는 장점이 있다. 그러나 현장에서 기기를 활용하여 측량이 이루어지기 때문에 토지의 크기가 광대하거나 측량이 요구되는 토지가 많을 경우 많은 비용과 시간이 소요된다. 이러한 측면에서 전국 약 3,300만 필지에 대한 개별 측량은 사실상 불가능하다고 할 수 있다.

인공위성을 이용하는 방법은 인공위성에서 촬영된 영상정보로부터 높이 정보를 추출하는 방법으로 인공위성 촬영주기마다 정기적인 조사가 가능하고 상대적으로 정확한 데이터를 확보할 수 있다는 장점이 있다. 그러나 영상자료 취득에는 많이 비용이 소요되며, 영상정보를 처리할 수 있는 별도의 소프트웨어와 분석 인력이 필요하며 위성영상 자료의 공간 해상도에 따라 분석 결과가 달라지는 단점이 있다.

마지막으로 GIS를 이용하는 방법은 지리정보체계 프로그램을 이용하여 경사도를 분석하는 방법으로, 수치정보로 구축된 수치지도를 활용하여 고도와 경사도를 산출하는 방법이다. GIS를 이용하는 방법은 전국 단위로 구축된 수치지도가 구축되어 지속적인 갱신이 이루어지고 있고, 관련 소프트웨어의 대중화라는 장점이 있다(Choi and Park, 2008). 또한 현재 대부분의 자료가 무료로 공개되고 있어 경제적 효율성이 큰 이점이 있으나 정확도에서는 전술된 데이터들에 비해 상대적으로 떨어지는 단점이 있다.

토지의 고저 조사를 어떻게 규정하는가는 본 연구의 방향정립을 위한 과제라 할 수 있으나 효율적이고 경제적인 방법을 통해 객관적 기준을 제시하는 측면도 고려되어야 한다.

특히 경사도의 경우 하나의 대상 토지에 대해 분석에 활용하는 데이터의 종류와 경사를 계산하는 방법에 따라 서로 다른 결과가 도출될 수 있기 때문에 경사도 계산과정에서 일괄적인 계산방법과 조사 기준이 적용되지 않을 경우 비교 과정에서 오류가 발생할 가능성이 높다는 점을 감안해야 한다.

경사도를 계산하는 방법은 계산 가중치와 지형면의 요건에 따라 다양하게 제시될 수 있으며 대표적인 방법은 다음과 같다(Lee, 2016).

먼저 등고선에 직각 방향인 방향선을 이용하여 경사를 계산하는 방법으로 경사도 측정을 위한 단면을 생성하고 이 단면에

나타나는 등고선들에 직각이 되도록 수선을 설정하여 계산하는 방법이다.

$$\tan(d) = H/D \tag{1}$$

where d denotes slope, H denotes elevation difference between contours, D denotes horizontal distance between contours

이 방식은 대부분의 도시계획에서 Fig. 2와 같이 지형이 평평한 면으로 변하는 경우에 2개 이상의 단면의 경사를 이용하여 2개 이상의 단면 중 최댓값을 해당 토지의 경사도로 산정하거나, 가중치를 기준으로 평균경사도를 산정하는데 활용되는 방법이다.

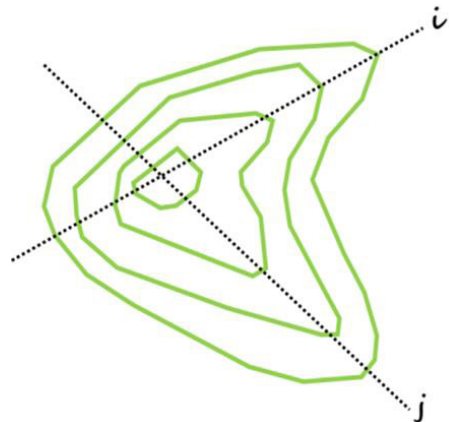


Fig. 2. Nonidentical topography

$$d = \frac{\sum d_{ai}l_{ai} + \sum d_{bj}l_{bj}}{\sum l_{ai} + \sum l_{bj}} \tag{2}$$

where, d denotes mean slope, d_{ai} denotes local slope of i^{th} segment in cross section a, l_{ai} denotes distance of i^{th} segment in cross section a, d_{bj} denotes local slope of j^{th} segment in cross section b, l_{bj} denotes distance of j^{th} segment in cross section b

경사면적을 활용하여 경사도를 계산하는 방법은 일정한 수준의 경사(S_i)를 가진 평면 면적(A_s)과 수평 투영 면적(A_p)의 수학적 관계를 Eq. (3)와 같이 이용한 방법이다.

$$\sec S_i = \frac{1}{\cos S_i} = \frac{A_{si}}{A_p} \tag{3}$$

where S_i denotes slope, A_{s_i} denotes surface area, A_π denotes orthogonal projected area of A_{s_i}

격자높이를 활용하는 방법은 해당 토지를 일정한 격자 간격으로 나누어 각 격자별 경사를 평균하여 토지의 평균 경사를 구하는 방법이다. 이는 Fig. 3과 같이 주위의 8개 격자 높이를 활용하여 중심에 있는 격자를 경사를 계산하는 것으로 주변의 8방위에 분포한 격자의 방향별 경사를 계산하여 중심격자의 경사 S 를 구하는 방법으로 유한차분법이라고도 명명되고 있다.

$Z_{i-1,j+1}$	$Z_{i,j+1}$	$Z_{i+1,j+1}$
$Z_{i-1,j}$	$Z_{i,j}$	$Z_{i+1,j}$
$Z_{i-1,j-1}$	$Z_{i,j-1}$	$Z_{i+1,j-1}$

Fig. 3. Grid separation

이상에서 과정을 종합할 때 토지의 높이와 경사도 조사의 방법은 GIS를 중심으로 활용하는 것이 경제적인 측면에서 가장 유리하며, 조사 기준(계산 방법)은 등고선을 직접 활용 방법의 경우(식 2와 식 3의 방법)는 등고선 정보를 포함하고 있는 수치지도 또는 측량정보가 가진 오차의 요인과 함께 활용자간 수직면 설정에 따른 편차 발생할 수 있기 때문에 유한차분법(격자 높이를 활용하는 방법)이 범용성 측면에서 유리하다고 볼 수 있다(Florinsky, 2012).

특히 최근 개정된 국토의 계획 및 이용에 관한 법률(이하 국토계획법)과 각 지자체별 도시계획 조례에서 토지개발행위허가 과정에 필요한 토지의 경사도를 산정하는 방법은 국토계획법 시행규칙 제10조의 2에서 격자의 높이를 활용하도록 규정하고 있어 범용성과 경제성, 제도적 체계의 일원화라는 측면에서 볼 때 개별요인을 비교하는 과정에서 경사도 산정을 격자 높이를 이용하여 토지의 경사를 산정하는 방법이 가장 효율적이라고 볼 수 있다. 또한 GIS를 활용한 유한차분법은 이미 법제도적으로 규정되어 있으면서 연속적인 수치로 토지의 경사를 나타낼 수 있어 비교 대상 토지간 요인 비교에 유리하다고 볼 수 있다.

이에 따라 고도 및 경사도 분석은 개발 행위와 제도적 인허

가 기준에 적합성을 함께 적용하기 위해 국토계획법 시행규칙 제10조의 2와 산지관리법 시행규칙 별표 1의 3 비고 2호에 따른 방법을 차용할 수 있다. 해당 방법은 기본적으로 GIS 프로그램을 기반으로 분석하며 격자 높이를 계산하여 토지의 높이와 경사도를 산출하는 유한차분법과 동일하다.

국토계획법 및 산지관리법에서 제시하고 있는 평균고도와 경사도의 측정방법은 수치지형도(1:5,000)를 기본으로 10m × 10m의 격자로 필지를 각각 분할하여 불규칙삼각망을 생성한 후 격자 내 삼각면의 고도와 경사도에 면적비율을 적용하여 측정 대상지의 평균고도와 경사도를 산출한다(Woo *et al.*, 2001). 높이 정보 추출 역시 국토지리정보원에서 제작한 1/5,000 수치지도를 기준으로 추출하였다. 수치지도는 2018년 12월말 기준 지도이다. 수치지도는 국토지리정보원에서 운영하고 있는 국토정보플랫폼(<http://map.ngii.go.kr>)을 통해 무상으로 구득가능하다.

정보 추출은 수치지도(Fig. 4(a))에서 등고선 레이어(Fig. 4(b))를 추출하고, 이를 불규칙삼각망(TIN, Triangular Irregular Network) 생성(Fig. 4(c)) 후 이를 다시 디지털수치모델(DEM, Digital Elevation Model)로 변환(Fig. 4(d))하였다. 변환 정보는 GIS 분석으로 경사(degree)로 표현(Fig. 4(e))하였다.

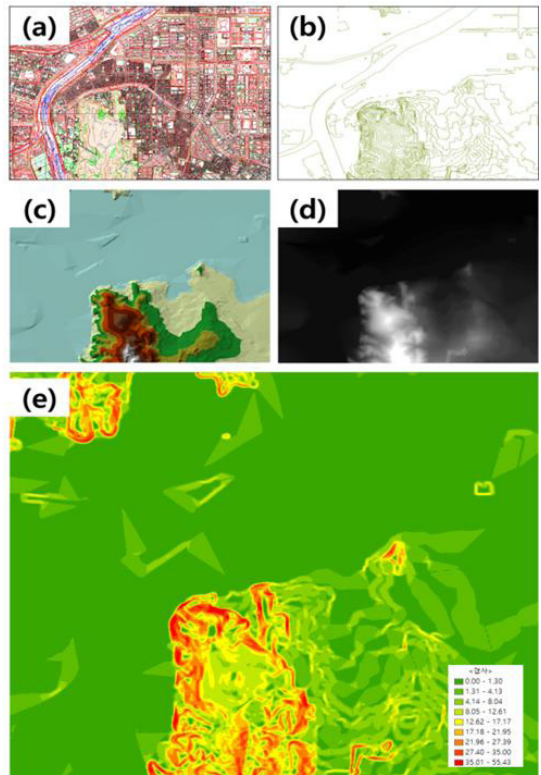


Fig. 4. Process of slope analysis

생성된 고도와 경사정보는 각 필지와 중첩하여 다음과 같이 필지 내 경사정보를 격자별(10m*10m) 단위로 생성하였다. 격자별 고도와 경사정보는 Eq. (4)와 같이 격자별 경사도와 격자면적(또는 고도와 격자면적)의 총 합을 필지의 면적으로 나눈 값으로 치환하여 Fig. 5와 같이 각 필지의 격자별 고도와 경사도를 바탕으로 해당 필지의 최종적인 고도와 경사도 값으로 추출하게 된다(Lee and Cho, 2002).

$$Mean Value = \frac{\sum_{i=1}^n (GI_i \times GA_i)}{\sum_{i=1}^n GA_i} \quad (4)$$

where GI_i denotes slope value of grid, GA_i denotes area of grid

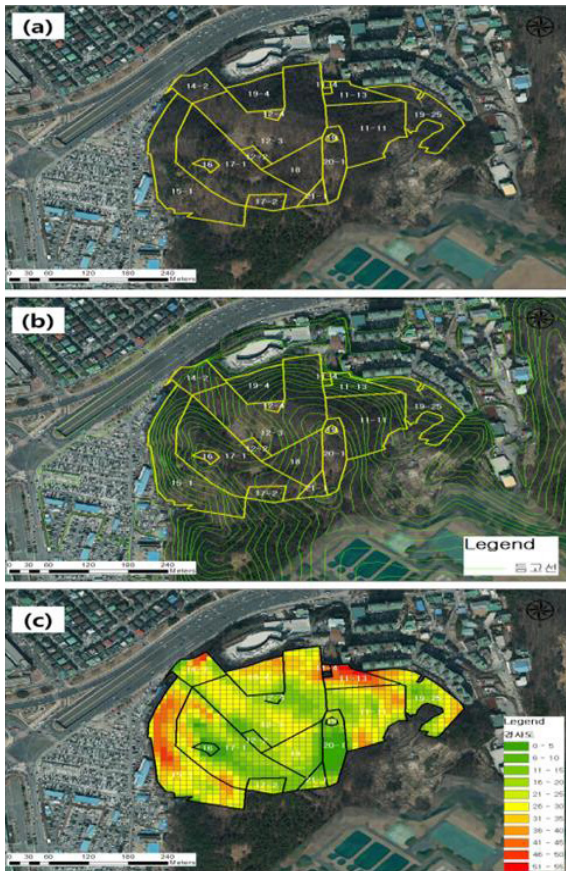


Fig. 5. Process of slope by cell

이러한 방법에 따라 고저 조사를 위한 기본 조사 정보로서 Table 2와 같이 필지별 평균 고도와 경사, 지세별 평균 고도와 경사를 구축하였다.

Table 2. Applicable topography classification criteria

Type		Content
Elevation	Parcel	Average elevation
	Zoning	Average elevation of a 250m radius zone with respect to the parcel center point
Slope	Parcel	Average slope
	Zoning	Average slope of a 250m radius zone with respect to the parcel center point

3.3 데이터 마이닝에 의한 고저의 분류

고저의 분류 단계는 구축된 고도와 경사 정보를 기준으로 조사기준에 의거하여 수행할 수 있다. 그러나 현재 고저 조사기준은 주위의 지형지세와 비교하여 ‘현저히 낮은’, ‘비슷하거나’, ‘미미한’, ‘현저히 높은’ 등의 정성적인 기술 형태로 제시되어 있어 정량적 분석 적용이 불가능하다. 따라서 분석정보를 바탕으로 객관적 분류 과정이 이루어지기 위해서는 새로운 형태의 분류 기준이 요구된다.

어떠한 분류 규칙의 생성은 결과에 영향을 미치는 정도를 분석을 통해 밝혀내고 이를 통해 기준이 세우는 것이 일반적이다. 그러나 토지특성조사와 같이 이미 수십년의 시간 동안 적용되어온 결과이면서 현재와 같이 정량적 분류가 어려운 기준이 설정되어 있는 경우 기존 분류된 결과로부터 행태적 정보를 추출 또는 학습을 통해 새로운 규칙을 생성하는 것도 하나의 방법이 될 수 있다.

최근 들어 더욱 주목받고 있는 데이터마이닝은 주어진 데이터로부터 다양한 형태의 유용한 정보를 추출하기 위해 모델링 방법을 적용하거나, 관측된 패턴의 유용성을 판단하고 데이터 속에 포함되어 있는 데이터간의 상관관계나 패턴, 규칙 등을 찾아낼 수 있다. 따라서 고저 분류에서도 현재까지 조사된 고저 조사 결과를 바탕으로 분석된 요인 관측치(고도, 경사)를 바탕으로 마이닝 과정을 통해 새로운 분류 기준 제시에 활용할 수 있다.

데이터 마이닝에는 Decision Tree, 1-Rule, Bayesian Network, Artificial Neural Network 등 다양한 방법이 있다.

고저 조사는 여러 조건과 자료를 활용하여 최적의 결과를 만들어내는 개념학습(concept learning) 분야라고 할 수 있다. 즉, ‘저지’, ‘평지’, ‘완경사’, ‘급경사’, ‘고지’ 등의 고저 조사 분류 체계를 자동으로 분류(classification)하는 데이터 마이닝의 한 분야라고 할 수 있다.

그러므로 1-Rule과 Bayesian network, Artificial neural network은 각각 하나의 속성 사용, 블랙박스적 방법이라는 한

계가 있어 토지 고저 조사를 인식시키기 위한 데이터 마이닝 방법으로 적절하지 않다. 반면에 Decision Tree는 범주화를 통해 고저를 분류하는 하나의 규칙을 표현할 수 있어 패턴 분류의 문제를 가지는 데이터를 처리하는데 효과적이라고 할 수 있다 (Choi and Lee, 2017). 이에 본 연구에서는 고저를 분류하는 데이터 마이닝 방법으로 의사결정트리로 설정하였다.

의사결정트리에는 Table 3과 같이 CART (Classification And Regression Tree), CHAID (CHi-squared Automatic Interaction Detector), C4.5 등의 알고리즘이 존재한다. 각 알고리즘은 활용할 수 있는 종속 및 독립변수의 데이터 유형, 분리기준, 분리 개수 등에서 차이가 있다. 이번 실험은 고저를 분류하는 과정에서 해당 필지의 고도와 경사뿐 만아니라 주변지역의 고도와 경사 정보와 같이 3가지 이상의 독립 변수(입력 변수)가 활용되고, 다양한 조건(고도와 경사 조합)을 검토하여 분리 객체를 분리되어야 한다. 또한 객체 분류 문제를 처리하기 용이한 알고리즘이 필요하다. 이에 따라 본 연구에서는 의사결정트리 중 다지분리가 가능하고 객체분류 문제 처리에 상대적으로 용이한 C4.5 알고리즘을 해석 모형으로 설정하였다.

Table 3. Comparison of algorithms in decision tree

Type	C4.5	CHIAD	CART
Input value	Nominal	All type	All type
Output value	All type	All type	All type
Classification tree	Entropy index	P_value of Chi square statistic	Gini Index
Split criterion	Multiple split	Multiple split	Binary split

4. 데이터 마이닝 분석결과

4.1 데이터 마이닝 분석

고저 조사를 위한 실험을 위해 서울시를 대상으로 고도와 경사 정보를 구축하여 고저 분류를 위한 요인 정보로 활용하였다.

다음으로 분류 기준 생성을 위해 2019년 기준 서울특별시 표준지를 대상으로 C4.5 데이터 마이닝 알고리즘을 적용하였다. 여기서 학습 기준 생성을 위해 표준지를 사용한 이유는 개별지에 비해 전문적 조사가 이루어져 기존 조사 오류가 적고 현재 표준-개별방식에 따라 표준지에 분류 기준이 나머지 개별지에 적용이 되는 것이 합리적이라 판단했기 때문이다.

2019년 기준 서울특별시 표준지는 총 29,152필지이나 일단지(용도상 불가분의 관계에 있는 2필지 이상의 일단의 토지)에 대한 구분이 명확하지 않거나 현 시점 연속지적도와와의 연결이 되

지는 않는 178필지는 제외하여 총 28,974필지를 학습 대상으로 선정하였다. 2019년 기준 서울특별시의 표준지 고저 분포는 전체 28,974필지 중 저지 48필지, 평지 23,530필지, 완경사지 4,690필지, 급경사지 517필지, 고지 189필지가 분포하고 있다.

기존 표준지 고저 조사 결과를 바탕으로 공간정보기술로 구축된 고도와 경사 정보를 C4.5 데이터 마이닝을 통해 분류 기준을 학습하였고, 학습 결과를 표준지를 제외한 나머지 개별지에 대해 적용하여 기존 개별지 조사 결과와 교차 검증을 통해 실험 결과의 정확도 및 재현을 검증하였다.

검증 대상인 개별지는 학습된 표준지와 동일하게 일단지 토지와 지적도 매칭이 이루어지지 않는 토지와 조사 대상으로 보기 어려운 하천, 도로 등의 공공 및 특수 용지는 검증 대상에서 제외하여 Table 4와 같이 총 583,090필지에 대한 분류 실험을 진행하였다.

Table 4. Result of the topography investigation (existing result)

Type	N	Ratio(%)
Low land	868	0.1
Flat land	437,532	75.0
Mild slope	121,272	20.8
Steep slope	17,027	2.9
High land	6,391	1.1
Total	583,090	100.0

4.2 분석결과 및 고찰

표준지 필지와 주변 지세의 평균고도와 경사도를 바탕으로 학습된 고저 분류 결과를 바탕으로 개별지에 적용한 결과는 다음과 같다.

서울특별시 개별지 583,090 필지에 대한 분류 결과는 기존 조사 결과와 비교 시 약 93.5%(548,536/586,595)로 높은 수준의 정확도가 나타났다.

각각 구분 항목별 정확도는 저지 99.9%(867/868), 평지 약 99.5%(435,435/437,532), 완경사 약 82.6%(100,132/121,272), 급경사 약 43.2%(7,358/17,029), 고지 약 74.2%(4,744/6,391)로 나타났다. 또한 표준지 고저 정보와 데이터 마이닝을 통해 조사된 고저 정보의 확률적 일치율을 나타내는 Cohen's kappa 계수가 0.84로 나타나 거의 완벽한 일치율(almost perfect)이 나타났다.

이에 따라 이번 실험을 통해 구축된 정보와 데이터 마이닝을 통해 학습된 규칙을 바탕으로 고저를 분류하고 조사하는데 상당한 수준의 연관관계와 성과가 있음을 확인할 수 있으며, 이러한 결과는 이번 실험을 통해 학습된 규칙이 고저를 조사할

수 있는 일반적인 규칙으로 적용될 수 있으며, 분석 대상의 약 93.5%에 대해서 사람에 의한 조사 결과와 동일한 결과물을 제시할 수 있음을 의미한다.

Table 5. Result of the topography investigation test

existing \ ↓ testing	Low land	Flat land	Mild land	Steep land	High land	Total
Low land	867	0	1	0	0	868
Flat land	0	435,435	1,875	222	0	437,532
Mild land	0	19,681	100,132	1,459	0	121,272
Steep land	0	523	9,148	7,358	0	17,029
High land	0	76	1,360	211	4,744	6,391
Total	867	455,715	112,516	9,250	4,744	583,092

반면에 실험결과에서 총 34,556필지(약 6.5%)가 기존 고저 조사 결과와 일치하지 않는 것으로 나타났다.

이는 기존 개별지 고저 조사가 잘못되었기 때문에 실험결과와 비교과정하는 과정에서 비일치로 도출되었을 수도 있고, 다른 측면에서는 데이터 마이닝을 통해 학습된 고저 분류 규칙으로는 설명할 수 없는 부분이 있을 수 있음을 의미하기도 한다.

실험과정에서 적용한 C4.5에서는 tree가 약 182개의 node로 구성되어 나타났다. 이는 현재의 조사결과에는 다소 이질적인 결과도 상당수 포함되어 있음을 의미한다. 즉 고저를 저지, 평지, 완경사 등의 5가지 항목으로 구분하는 과정이 약 180개로 분류하여 판별하는 과정으로도 설명되지 않는 부분이 상당수 포함되어 있다고 볼 수 있다.

이와 같은 사항은 다음의 예시에서 찾아 볼 수 있다.

Fig. 6은 서울 강동구 둔촌동 산150사례로 기존 지자체 조사 결과로는 ‘평지’로 분류된 토지이나 데이터 마이닝 결과로는 ‘급경사로’ 분류된 토지이다.



Fig. 6. Example of result difference (1)
Existing(flat land) → Testing(steep land)

해당 토지는 임야용 토지로서 필지 내 등고선 분포로만 봐도 상당한 경사와 높이차가 발생하는 토지로 볼 수 있어 평지로 판단하기는 무리가 있는 상태이다. 따라서 해당 토지는 급경사로 분류되는 것이 타당하다.

Fig. 7은 서울 은평구 불광동의 사례로 기존 지자체 조사 결과에서는 ‘급경사로’ 조사된 반면 데이터 마이닝 결과에서 ‘완경사로’ 분류된 토지이다. 해당 토지는 차량의 통행이나 건축이 불가능한 수준이 아니기 때문에 급경사로 분류하기는 어려운 상태로 볼 수 있다.



Fig. 7. Example of result difference (2)
Existing(steep land) → Testing(mild land)

Fig. 8의 사례는 강남구 자곡동 산2-0의 사례로 지자체 조사 결과 ‘급경사로’ 조사된 토지이나, 데이터 마이닝 결과에서는 ‘평지’로 분류된 토지이다. 해당 토지는 평균 약 2°의 경사도 수준을 가지고 있어 매우 평평한 수준이지만 주변경사가 다소 완만한 형태로 형성되어 경우에 따라 완경사로 판단할 수 있으나 급경사의 토지로 보기에는 무리가 있다. 따라서 해당 토지는 데이터 마이닝의 분류 결과와 같이 ‘평지’로 조사되는 것이 타당할 것으로 판단된다.



Fig. 8. Example of result difference (3)
Existing(steep land) → Testing(mild land)

Fig. 9의 사례는 강남구 개포동의 사례로 현황 고지로 조사되어 있는 토지이다. 해당 토지는 종교시설이 들어서 있는 토지로 약간의 경사지임에는 분명하나 해당 토지가 주변 지대에 비해 절대적으로 높은 '고지'를 형성하고 있지는 않았다. 따라서 데이터 마이닝 분류 결과와 같이 '완경사'의 토지로 조사되는 것이 타당한 것으로 판단된다.



Fig. 9. Example of result difference (4)
Existing(steep land) → Testing(flat land)

이와 같이 현재 토지의 고지 조사는 조사자에 따라 일부 조사 오류를 내포하고 있는 것을 확인할 수 있다. 따라서 이번 실험을 통해 나타난 결과 중 기존 조사자의 결과와 불일치한 약 6.5%의 결과물은 입력오류나 조사자의 착오로 인해 결과로 생각해 볼 수 있다.

본 연구는 기존의 표준지 고지 조사 데이터로부터 학습된 고지 조사 규칙을 바탕으로 비교적 정성적인 고지 조사 요령을 보다 정형화된 규칙으로 대체하고자 하였다. 따라서 기존에 일반적으로 규칙으로 분류되는 약 93.5%의 토지를 제외한 나머지 6.5%의 토지에 대해서 어느 누가, 언제라도 조사하여도 애매 모호한 결과가 아닌 직관적으로 분류할 수 있는 장점이 있다고 판단된다.

특히 현재와 같이 표준지와 개별지의 조사가 이원화되어 있는 상태에서 보다 감정평가사 조사한 고지 분류 정보를 학습하여 나머지 개별지에 대한 고지를 분류하는 과정을 통해 조사의 일관성을 유지하는 방안이 될 수 있을 것으로 판단된다.

5. 결론

본 연구에서는 GIS 공간분석과 데이터 마이닝 분석 기법 중 하나의 C4.5을 활용하여 토지의 고지를 분류하기 위한 데이터 구축과 분류 실험을 진행하였다. 이를 위해 공간정보기술을 바탕으로 필지의 경사도와 고도 정보를 추출하여 데이터 마이닝

을 통해 학습한 후 토지 고지를 분류하였다.

2019년 서울시 개별지 총 583,092필지에 대한 실험을 진행한 결과 약 93.5% 일치율과 0.84의 Kappa 지수가 나타났다. 이는 기존 고지 조사결과와 이번 실험을 통해 분류된 고지 조사 결과가 실험 대상 583,092 필지 중 548,536 필지가 동일하게 조사됨을 나타내는 것으로 컴퓨터에 의한 학습 결과물이라는 것을 고려한다면 상당한 수준의 결과가 도출되었음을 알 수 있다.

또한 불일치한 나머지 6.5%의 결과물을 살펴본 결과 대부분의 경우가 사람에 따라 애매모호한 판단이 이루어질 수 있는 대상이거나 조사 착오 사항이라는 점을 감안한다면 실제적인 불확실성은 더 줄어들 수 있을 것으로 판단된다.

이러한 결과로 볼 때 공간정보와 데이터 마이닝 기법을 기반으로 제시된 토지 고지 분류 방법은 조사의 효율성과 객관성을 개선시킬 수 있을 것으로 판단된다. 또한 제시된 방법을 발전시킬 경우 도로접면, 방위 등 공시지가 산정을 위한 다른 물리적 특성항목에 대해서도 적용이 가능할 것이다.

그러나 데이터 마이닝에는 다양한 방법이 존재하고 이에 따른 분류 정확도도 다르게 나타날 수 있다. 따라서 향후 다양한 알고리즘을 적용하여 최적의 분류 기법을 찾는 후속 연구와 함께 지역성을 반영하기 위해 지역별 연구가 필요할 것으로 판단된다.

또한 분류과정을 요약하거나 보여줄 수 없는 한계점을 보완하기 위해 행태적 분석 과정의 도입도 고려되어야 할 것으로 판단된다.

References

- Choi, J.H. and Lee, S.H. (2017), A study on the method of land shape recognition using the geometric feature information of parcel and the technique of decision tree, *Journal of Real Estate Analysis*, Vol. 3, No. 1, pp.19-34. (in Korean with English abstract)
- Choi, J.H., Shim, J.W., and Kim B.J.(2015), *A Study on Improvement of Land Characteristic Survey; A Study on Advancing Real Estate Disclosure System (IV)*, Research Institute of Korea Appraisal Board, Daegu, pp. 30-31. (in Korean with English abstract)
- Choi, S.Y. and Park, J.I. (2008), A study on GIS data sorts and format for real estate appraisal, *Journal of the Korean Cadastre Information Association*, Vol. 10, No. 1, pp.77-94. (in Korean with English abstract)
- Florinsky, I.V. (2012), *Digital Terrain Analysis in Soil Science and*

Geology, Elsevier Inc., San Diego, USA.

- Kim, T.W., Kang, I.J., Park, D.H., and Hwang, D.Y. (2015), Method to objectify individual factors of GIS-based real estate appraisal, *Journal of the Korean Society for Geospatial Information Science*, Vol. 23, No. 4, pp.35-41. (in Korean with English abstract)
- Lee, B.K. (2016), Improving the slope calculation method for evaluating the feasibility of the land development, *Journal of the Korean Society for Geospatial Information Science*, Vol. 24, No. 3, pp.85-92. (in Korean with English abstract)
- Lee, G.S. and Cho, G.S. (2002), The analysis of DEM accuracy and evaluation of resolution impact GIS environment –valley, peck, ridge-, *Journal of the Korean Society of Civil Engineers D*, Vol. 22, No. 3, pp.583-590. (in Korean with English abstract)
- Lee, H. (2018), A study on the effect of slope on housing prices: focusing on apartment complexes in seoul area, *Journal of the Korea Planning Association*, Vol. 53, No. 1, pp.153-177. (in Korean with English abstract)
- Lee, H.S. and Koh, J.H. (2013), A study on the improvement measuring and actual situation of determination of parcel's height land for calculation of publicly notified individual land price, *Journal of the Korean Society of Cadastre*, Vol. 29, No. 2, pp.53-62. (in Korean with English abstract)
- Woo, J.H., Koo, J.H., Hong, C.H., and Kim, T.H. (2001), A study on interpolation methods and size of grid to the various topographical characteristics for the construction of DEM, *Journal of the Korean Open GIS association*, Vol. 3, No. 2, pp.5-9. (in Korean with English abstract)