

의사결정나무와 서포트 벡터 머신 모델을 활용한 토지이용 변화 시뮬레이션 : 통일 후 북한 도시를 대상으로 *

Simulating Land Use Change Using Decision Tree and SVM Model
: A Case Study of North Korea's City after the Unification

원석환 Won Seokhwan **, 황철수 Hwang Chul Sue ***

Abstract

This study aims to estimate land-use changes in North Korea's city after the unification through simulation. The simulation analysis consists of exploration of urban patterns, validation of simulated results, and analysis of the land-use changes in the future. The estimation accuracy of land use change was measured by polynomial regression analysis, decision tree analysis and support vector machine analysis. On this basis, suitable methodologies were selected for the modeling of the urban pattern search. The hybrid model uses the 'C5.0', a kind of decision tree algorithm and the 'RBF kernel', a kind of support vector machine algorithm. The land use change of North Korea's city (Nampo City) was estimated based on this model. The scenario of newly constructing roads and railroad stations, and developing new urban areas was used in the simulation. The simulated results showed that the urban sprawl impact through improvement of traffic accessibility and the development of new urban areas could be linked to the region and other regions. This study could provide a strategic guidance for policy when the unification of Korea will be on the process.

Keywords: Machine Learning, Land-use/Cover Change, North Korean Study, Urban Simulation Analysis, Inaccessible Area Analysis

I. 서론

1. 연구 배경 및 목적

한국전쟁 이후 남·북한은 시장경제체제와 사회주의 체제라는 상이한 사회·경제 구조를 유지하여 왔고,

이에 따라서 두 체제의 이질성은 심화되어 왔다. 분단으로 인해 만들어진 이질성은 남·북한 간의 대립과 맞물려 지속적으로 확대·재생산 되었으며, 이로 인해 통일이 이루어질 경우 확대된 이질성은 사회 전반에 걸쳐 다양한 문제점을 발생시킬 것으로 예상된다(변병설, 윤갑식 2001). 남·북한 사이의 이질성은 다양한

* 본 논문은 원석환(2018)의 박사학위 논문 '머신러닝 기법을 적용한 통일 후 북한 도시 토지이용 변화 시뮬레이션의 일부를 수정·보완한 것임.

** 경희대학교 지리학과 객원교수(제1저자) | Visiting Prof., Dept. of Geography, Kyunghee Univ. | Primary Author | wonsh@khu.ac.kr

*** 경희대학교 지리학과 교수(교신저자) | Prof., Dept. of Geography, Kyunghee Univ. | Corresponding Author | hcs@khu.ac.kr

분야를 포괄하고 있으며, 국토의 이용과 관리 역시 상당한 수준의 이질성을 보여주고 있다. 또한 분단으로 인해 효율적인 국토의 이용이 이루어지고 있지 못하다는 점 역시 주지의 사실이다. 과거 우리나라의 국토 개발과 토지이용의 경험에 비추어보면, 통일 이후에는 북한 지역의 국토개발이 가속화될 것으로 예상되며 동시에 이에 따른 많은 문제점도 수반될 것으로 예상되지만, 이에 따른 대안 마련을 위한 자료의 획득과 연구의 수행이 어려운 상황이다.

통일 이후의 문제점을 최소화하고 예상되는 문제점을 해결할 대안을 마련하기 위해서는 북한 지역에 대한 다각적인 이해와 연구가 필요하며, 공간의 변화를 예측할 수 있는 연구의 수행이 시급하다(이상준, 박세훈, 남기찬, 이건민 2014). 이 중 통일 이후 도시에 대한 연구는 매우 구체적인 수준의 연구가 필요하며, 이는 통일에 대한 방안과 정책에 대한 토론을 병행하여 준비되어야 하는 과정이다.

통일 이후 북한의 도시들은 큰 변화를 겪게 될 것이라는 점을 예상하기 어렵지 않다. 그러나 도시는 다양한 요소 간의 복잡한 상호작용 과정을 통해 변화하고 성장하기 때문에, 통일이라는 큰 변수가 발생할 경우 북한의 도시들은 어떠한 경로로 변화해 갈 것인지 예측하는 것은 매우 어려운 과정이 된다. 이러한 복잡한 문제들을 해결하기 위해서는 비슷한 선행사례를 분석하는 연구가 선행되어야 한다. 통일 이후 북한 도시 및 지역 계획, 다각도의 도시화 분석 등 통일을 대비한 대다수의 연구는 실증적인 연구가 진행되지 못하였으며, 과거 체제전환국의 도시변화 추세를 분석하여 향후 북한 발전 방향을 제시 혹은 추정하는 방법으로 주로 이루어졌다. 상대적으로 통일 이후 북한의 도시화 과정 및 현상을 구체적으로 추정하는 연구는 미비하였는데, 이는 북한이 현재 폐쇄적인 국가이며 접근불능지역이라는 지역적 특징 때문에 공간수치자

료를 비롯한 다양한 자료가 부재하다는 점에 기인한다(원석환, 황철수 2018).

본 논문은 북한 도시 토지이용 변화의 양과 변화의 방향을 추정하기 위한 적절한 방법론을 모색하고 지리 정보시스템을 활용하여 접근불능지역의 데이터를 수집하며, 이를 바탕으로 북한 도시화의 미래 현상을 특정한 가정에 따라 추정하기 위해 연구를 수행하였다.

기존 연구의 데이터 한계점을 최대한 극복하기 위해 수치지형도, 위성영상, 통계연보 등을 통해 다양한 속성을 갖는 자료를 구축하였다. 이를 바탕으로 북한 도시를 대상으로 토지이용 변화 시뮬레이션 모델을 구축하기 위해 통계 기반의 다항로지스틱 회귀분석과 머신러닝 기반의 의사결정나무(Decision Tree: DT) 모델 및 서포트 벡터 머신(Support Vector Machine: SVM) 모델을 적용하였다. 다음으로 토지이용 현황별 추정 정확도가 높은 모델을 각각 선정하여 통합 모델을 구축하였다. 이를 통해 각종 시나리오별 공간 변화로 인해 발생할 수 있는 북한 도시의 토지이용 변화를 시·공간적으로 추정하고자 하였다.

II. 선행연구 검토 및 착안점 도출

공간 현상 및 토지이용 변화 시뮬레이션의 모델링 방법은 다수의 연구에서 다양한 방법으로 진행되고 있다. 방법론적으로 가장 우수한 방법을 탐색하기 위해 다양한 방법을 시도하고 있으며, 획득 데이터의 유형 및 상태에 따라서 상이한 방법론을 적용하여 분석이 진행되고 있다.

통계기법 중 선형회귀분석이 종속변수와 독립변수 간의 선형관계를 전제로 하는 데 반해, 로지스틱 회귀 분석은 비선형의 로지스틱 형태를 취하며 종속변수와 독립변수 간의 상관관계를 밝히고 추정이 가능한 통계 기법이다. 특히 종속변수가 명목적으로 측정된 결과일

경우에는 일반적인 회귀분석을 시행하기 난해하다. 하지만 로지스틱 회귀분석은 일반적인 회귀분석 기법이 내재하는 단점인 독립변수들 간의 상호작용 효과와 독립변수의 개수에 대한 민감도의 한계를 극복하지 못하는 문제점을 지니고 있다(구유희 2007). 통계 기반 분석은 상대적으로 해석이 용이하다는 장점을 지니기 때문에 다양하게 활용되지만, 통계적으로 가정해야 할 요소가 다수 존재하며, 변수들 간의 상관관계가 높은 경우 설명력이 떨어질 수 있다는 단점을 지니고 있다. 반면에 복잡계 기반의 분석은 변수들 간의 상호작용 특징을 포함한 분석이 진행된다는 특징을 지니고 있다.

복잡계 기반의 시뮬레이션 분석 방법은 주로 셀룰러 오토마타와 에이전트 시스템 분석 방법이 사용된다. 셀룰러 오토마타는 복잡한 자연계 현상의 모델링을 위한 수학적 방법론 중 하나로 이산적인 시·공간에 기초한 비선형 동적 모델로 정의되며(정재준 2001), 에이전트 시스템은 지능 독립체와 이와 관계되는 타 독립체 그룹과의 상호작용을 고려한 분석을 통해 복잡한 문제의 해결책 제시를 지원하는 컴퓨터 시스템이다(Jegou 2005; 원석환 2010에서 재인용).

복잡계 모형은 연구자가 정의한 변수와 변수 간의 규칙을 기초로 결과를 산출하는 반면에, 머신러닝 모형은 연구자가 관여할 수 없는 학습과정을 바탕으로 결과를 산출한다는 측면에서 본 논문에서 채택한 머신러닝 기법은 기존 시뮬레이션의 차별적 접근방법으로서의 의의를 지닌다. 머신러닝 기법을 활용한 기존의 국외 연구사례로 Chan, Chan and Yeh(2001)는 MLP(Multi-Layer Perception), LVQ(Learning Vector Quantization), DTC(Decision Tree Classifier), MLC(Maximum Likelihood Classifier) 등 4종류의 머신러닝 알고리즘을 이용하여 토지이용 변화의 예측 정확도를 비교하였다. 해당 연구는 세부적인 토지이용 변화 현상을 예측하기보다 토지이용 변화 유무만을 예측하였다.

Dixon and Candade(2008)은 토지이용 분류를 ANN(Artificial Neural Network) 기법과 SVM 기법, MLC 기법을 적용하여 비교 연구하였다. 연구결과 ANN과 SVM 기법이 MLC 기법보다 정확한 결과를 나타냈으며, SVM과 ANN 기법을 비교하면 SVM은 커널변수, ANN은 은닉노드와 같은 요소에 따라서 예측 정확도가 달라질 수 있다고 밝혔다. Lantz(2013)는 SVM 알고리즘은 범주나 수치예측 연구에 적합하며 ANN 알고리즘은 분류나 수치예측 연구에 적합하다고 밝혔다. 따라서 본 논문에서는 범주형의 종속변수(토지이용 변화)가 사용되기 때문에 ANN 기법보다는 SVM 기법이 보다 효과적이라고 판단하였다.

Srivastava, Han, Rico-Ramirez and Bray et al.(2012)은 머신러닝 알고리즘 중 MLC, SVM, ANN을 이용하여 토지이용 변화 예측을 비교하였는데, 분석 데이터로 위성영상 자료만 활용하여 토지피복의 현황으로만 변화를 예측하였다.

머신러닝 모형을 이용한 국내 연구사례는 개별공시지가 선정 연구(김종윤, 박수홍 2004), 자연재해 분석 연구(박선영, 장은나, 임정호, 최재원 2013; 김영민, 황미녕, 김태홍, 정창후 외 2015), 교통사고 심각도 요인분석 연구(강영옥, 손세린, 조나혜 2017) 등이 있으며 토지피복 분류에 대한 연구는 강톨가(2014) 등의 연구가 진행되었으나, 구체적인 미래의 토지이용 변화 추정에 관한 연구는 진행되지 않았다.

시뮬레이션 분석을 위해 다수의 기존 연구는 단일의 방법론을 선택했거나 혹은 다양한 방법론을 활용하여 분석한 이후 각각의 추정 정확도를 비교하였다. 하지만 본 논문은 통계(다항로지스틱 회귀분석) 기반의 시뮬레이션 기법, 머신러닝(DT, SVM) 기반의 시뮬레이션 기법을 다양하게 적용한 후, 가장 적합한 세부 모형을 선택하여 각각의 모형을 결합한 통합 모형을 구축하였다.

III. 연구의 범위 및 방법

1. 연구범위

연구의 공간적 범위인 북한 지역은 접근불능지역이며 폐쇄적인 사회구조로서 연구를 위한 각종 데이터를 확보하는 데 있어 매우 제한적이다.

본 논문을 진행하기 위해 선행연구로 북한 도시 정보를 바탕으로 남한의 도시와 도시 유사도 분석을 통해 공간적으로 유사한 남한의 도시와 북한의 도시를 추출하는 연구를 진행하였다. UN(2014)과 통계청의 북한의 주요통계지표 자료(통계청 2016)에 의하면 2015년 기준 북한의 도시화율과 1980~1990년 남한의 도시화율이 유사하였다. 따라서 2015년 기준 북한의 주요 9개 도시와 1980~1990년 남한의 주요 9개 도시를 대상으로 구축한 공간변수를 활용하여 군집분석 기반의 도시 유사도 분석을 진행하였다(원석환, 황철수 2018). 해당 분석결과, 2015년 기준 북한의 남포특별시와 공간적으로 가장 유사한 남한의 도시는 1980~1990년의 인천직할시로 분석되었다. 또한 이상준, 박세훈, 남기찬, 이건민(2011)의 연구에서도 남한의 인천직할시와 북한의 남포특별시는 도시 기능적으로 유사한 도시임을 밝히고 있다.

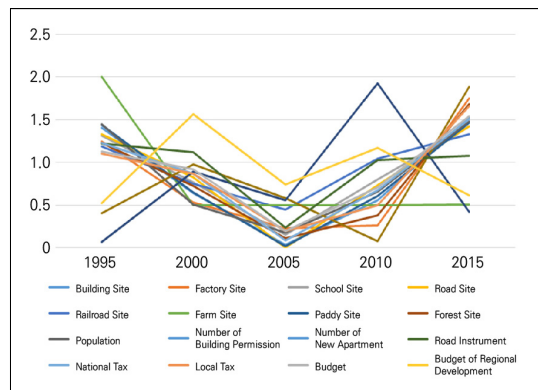
이를 바탕으로 1995~2015년 인천광역시의 토지 이용 변화 규칙을 기반으로 시물레이션 모델을 구축하여 통일을 가정한 2015년 이후 남포특별시의 토지 이용 변화 현상을 추정하는 연구를 진행하였다. 기준 연도를 1995년, 2005년, 2015년으로 설정하였는데, 이는 연구지역의 특징 때문이다. 인천광역시는 1995년에 직할시에서 광역시로 승격되어 많은 사회·공간적 변화가 이루어졌다. 모델을 구축하는 시작점이 되는 시기를 1995년으로 설정한 근거이기도 하다. 향후 진행될 북한 남포특별시의 토지이용 변화는 통일이라

는 이벤트가 발생함으로써 추정이 어려운 각종 변화가 발생할 것이다. 이러한 점에서 동일할 수 없겠지만 통일과 광역시 승격이라는 이벤트를 해당 도시 변화에 지대한 영향을 미치는 특별한 시기로 가정하여 시물레이션의 시작점으로 설정하였다. 단계별 10년의 간격을 설정한 것은 도시 토지이용 변화 현상은 다양한 현상의 복합적인 결과물로 나타나는 것으로 특정 이벤트에 단기간에 반응하지 않는 경우가 다수 존재하기 때문이다.

<Figure 1>은 1995년부터 2015년까지 지역 개발과 밀접할 것으로 판단되는 각종 통계지표(각 용지별 면적, 인구수, 건축허가수, 아파트건립동수, 도로연장 길이, 각종 세금 등)의 표준화 결과를 표현한 그래프이다. 그래프는 2005년을 기준으로 연구지역의 현상이 구분되는 패턴을 나타내고 있다. 또한 시물레이션 모델링을 위해 실험테스트를 수행한 결과 2005년 이후 연구지역의 도시화 현상이 급속하게 진행되었음을 확인할 수 있었다.

따라서 본 논문에서의 모델은 크게 2단계로 구분하여 구축하였다. 1단계는 1995년에서 2005년이며, 2단계는 2005년에서 2015년의 시간적 범위이다. 북한 도

Figure 1 _ Indicators Associated with Local Development in Incheon (1995~2015)



Source: <http://kostat.go.kr> (accessed August 5, 2017).

시 토지이용 변화 시뮬레이션 분석의 관점에서도 마찬가지로 북한은 통일 이후 시간이 경과할수록 도시 개발 속도가 빨라질 것이기 때문에, 본 논문의 모델 또한 단계별로 구분하여 구축하는 것이 타당하다고 판단하였다. Alghais and Pullar(2018)의 연구 또한 쿠웨이트 지역의 시기별 도시 발전 정도를 고려한 시뮬레이션 모델을 구축하였다. 따라서 2단계로 구분한 모델을 구축한 후, 북한 남포특별시의 시뮬레이션 분석 또한 2단계로 구분하여 분석을 진행하였다.

본 논문에서 가정하는 통일의 방법은 현재 남한 체제와 동일한 자유민주주의 시장경제 체제로의 통일을 의미하며, 통일 이후 북한은 남한의 과거 도시 토지이용 변화 패턴과 유사할 것으로 가정한다.

2. 연구자료

1) 시뮬레이션 모델 구현을 위한 데이터 구축

인천광역시의 연속수치지적도를 기준으로 1995년, 2005년, 2015년의 토지이용 자료(<http://www.nsd.go.kr>)와 DEM자료 등을 결합한 데이터를 구축하였다. 수치지적도와 각종 데이터의 결합은 위치기반의 벡터격자 형태의 데이터로 구축하였다. 래스터격자의 형태는 격자당 단일의 정보만 입력이 가능한데, 본 논문에서의 격자는 다양한 공간적 정보를 포함해야 하기 때문에 벡터격자의 형태로 데이터를 구축하였다. 격자의 크기는 10m×10m 단위로 설정하였다.

본 논문의 모델 구축을 위해서 설정한 독립변수(도로와의 거리 변화량, 하천과의 거리 변화량, 도심지와 거리 변화량, 철도역과의 거리 변화량, 고도, 경사)를 ArcGIS를 활용해서 각 격자별 데이터로 생성하였다. 각각의 독립변수는 북한 도시 토지이용 변화 시뮬레이션 분석을 위한 데이터 구축 시 위성영상 자료

및 북한 수치지형도를 통해 추출할 수 있는 토지이용 변화와 연관된 변수들이다. 접근불능지역이라는 북한 도시의 특수성으로 자료 획득의 한계가 존재하기 때문에 제한된 독립변수의 설정이 진행되었다. 도로와의 거리 변화량, 하천과의 거리 변화량, 도심지와 거리 변화량, 철도역과의 거리 변화량은 1995년, 2005년, 2015년을 대상으로 각 격자와의 거리를 유클리디안 거리를 활용하여 계산 후, 변화량을 추출하기 위해 각 연도 간 거리값을 차감하는 방식으로 산출하였다. 고도 및 경사는 DEM(<https://earthexplorer.usgs.gov>) 자료를 활용하여 각각의 격자에 정보가 부여될 수 있도록 데이터를 구축하였다.

중속변수(토지이용 현황) 역시 격자별 데이터로 결합 구축하였다. 토지이용 코드는 토지이용 현황별 중분류되어 있는 원자료를 재설정하였다. 토지이용 현황별 중분류되어 있는 원자료와 본 논문에서 설정한 분류 체계는 <Table 1>과 같다. 모델이 구축되면 북한 위성영상 자료를 적용하여 북한 도시의 토지이용 변화 시뮬레이션을 실행하게 되는데, 위성영상을 분류하는 데 있어 세밀한 토지분류체계를 활용한다는 것은 물리적으로 제한사항이 존재한다. 따라서 주거용지, 상업업무용지, 주상복합용지, 공업용지의 시가화 건조지역은 도심지역으로 그룹화하였다. 밭, 논,

Table 1 _ Land Use Classification Scheme

Raw Data	Classifications Set Up in this Study
Residential	Urban
Commercial	
Mix-Use	
Industrial	
Farm	Farm
Paddy	Paddy
Forest	Forest
Special	Etc.
Public	

산림지역은 위성영상으로 분류가 가능하기 때문에 그대로 적용하였으며, 특수용지 및 공공용지는 기타로 그룹화하였다. 본 논문의 모델 구축 단계에서 특수용지 및 공공용지는 토지이용 변화 모델링의 범주에서 제외하고 진행하였다. 용지의 특성상 자유롭게 토지이용 변화가 이뤄질 수 없기 때문이다.

2) 시뮬레이션 데이터 구축

통일 이후 북한의 도시인 남포특별시를 대상으로 시뮬레이션 분석을 위해 1995~2015년 인천광역시의 토지이용 변화 규칙을 바탕으로 모델이 구축되었기 때문에 모델 구축 데이터와 동일한 형태로 남포특별시의 데이터를 구축하였다.

모델 구축단계에서는 미시적인 분석을 위해 연속 수치지적도를 기반으로 토지이용 자료와 DEM 자료 등을 결합한 데이터를 구축하였지만, 남포특별시를 대상으로는 자료 획득의 한계 때문에 국토지리정보원으로부터 획득한 북한 위성영상 자료를 활용하였다. 해당 위성영상 자료는 2015년 아리랑위성 2호로 촬영된 1:25,000 축척의 정사영상 자료이다.

모델 구축단계와 동일하게 시뮬레이션 분석 또한 시기별 2단계로 구분하여 진행하였다. 1995년 이후 인천광역시와 2015년 이후 남포특별시가 동일한 수준과 패턴으로 변화한다는 가정은 다양한 제한사항이 존재하지만, 선행연구인 도시 유사도 분석 결과 공간적으로 가장 유사한 도시임이 밝혀졌기 때문에 이를 통해 남포특별시의 성장 모델로 적용시킨 것이다. 본 논문은 도시 변화에 있어 정적인 변수는 배제한 모델을 적용하였기 때문에 교통 접근성의 변화와 공간의 변화 측면에서 토지이용 변화 시뮬레이션 분석을 진행하였다.

모델 구축단계와 동일하게 토지피복 분류(도심지

역, 밭, 논, 산림지역)를 위해 남포특별시의 위성영상 자료를 대상으로 무감독분류를 진행하였다. 추가적으로 하천 또한 분류하여 독립변수 중 하나인 하천과의 거리를 산출할 수 있도록 하였다. 무감독분류는 ISODATA 기법을 활용하여 토지피복을 분류하였다.

토지피복 분류자료는 벡터격자의 형태로 전환하였다. 각 격자별 다양한 공간적 정보(독립변수: 도로와의 거리, 철도역과의 거리, 하천과의 거리, 도심지와 의 거리, 고도, 경사)를 포함해야 하기 때문이며, 공간적 해상도를 50m×50m 단위로 낮추어 시뮬레이션 분석을 진행하였다. 독립변수를 구축하기 위해서 도로 및 철도역, 하천정보는 국토지리정보원으로부터 획득한 북한 수치지형도 1:5,000 및 1:25,000 축척의 자료를 활용하였으며, 일부 데이터의 유실이 있는 경우 국토지리정보원 웹사이트의 수치지형도를 참고하여 해당 독립변수를 구축하였다. 고도 및 경사는 DEM 자료를 활용하여 해당 자료를 추출하였다.

3. 연구방법

본 논문에서 구축한 모델은 크게 통계 기반 모델과 머신러닝 기반 모델로 구분된다.

통계 기반의 모델은 다항로지스틱 회귀 알고리즘을 바탕으로 모델을 구현하였다. 머신러닝 기반의 모델은 다시 DT 알고리즘을 바탕으로 구현된 모델과 SVM 알고리즘을 바탕으로 구현된 모델로 구분된다. DT 모델은 C5.0 알고리즘을 적용하였다. SVM 모델은 선형커널, 다항커널, RBF(Radial Basis Function)커널을 각각 적용하여 모델을 구축하였다. 선형커널은 데이터의 구별 단계에서 단순히 내적 속성을 이용하고 데이터를 전혀 변환하지 않는 방법이며, 다항커널은 차원을 추가하여 데이터의 비선형 변환을 하는 방법이다. RBF커널은 가우시안 방사형 알고리즘을 활용하여

데이터의 비선형 변환을 하는 방법을 의미한다.

DT 알고리즘은 순환적 분할방식을 이용하여 나무를 구축하는 기법이다. 순환적 분할이란 하향식 귀납법으로 분류과정에서 서로 유사한 데이터들로 세분화하고, 이러한 세분화 작업을 반복 진행하는 방식을 의미한다(배정수 2014; 신은주, 장남식 1999). 해당 알고리즘은 나무구조에 의해 모형이 표현되기 때문에 해석이 용이하며 선형성이나 정규성을 가정하지 않는 비모수적 방법이라는 점에서 본 논문에서 채택하였다.

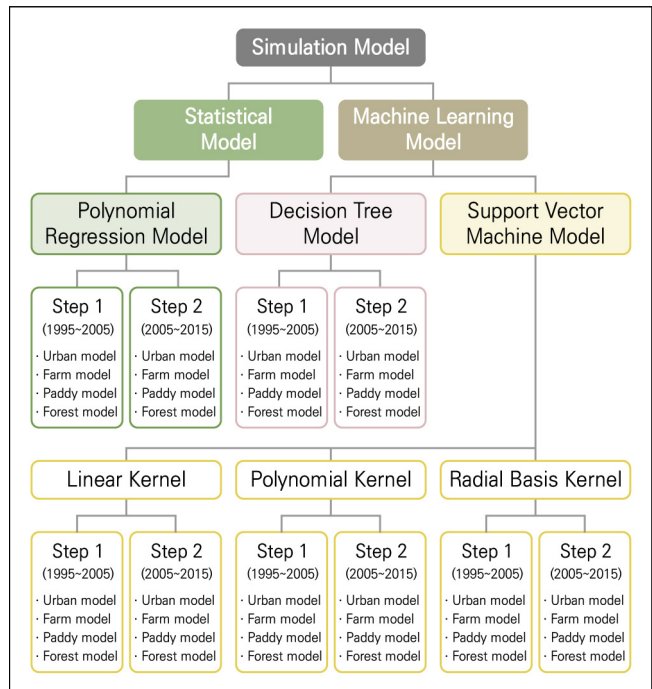
SVM 알고리즘은 분류율을 최소화하는 데 목표를 둔 기존의 분류 방법과 상이하게 마진(Margin)을 최대화하여 일반화 능력을 극대화시키는 방법이다(전주원 2011). 주요 특징은 커널 함수를 적용함으로써 분석을 고차원 공간으로 적용할 수 있다는 것이다. 본 논문에서와 같이 독립변수의 차원이 2차원으로 한정되지 않는 비선형적인 관계일 경우 효과적인 분류가 가능한 기법이다.

총 5개의 알고리즘을 바탕으로 단계별로 각각 모델을 구현하여 <Figure 2>와 같이 총 40개의 세부 모델을 구축하였다. 해당 40개의 세부 모델 중 추정 정확도를 평가하여 가장 적합한 모델 8개를 최종적으로 선정하여 시뮬레이션 분석을 위한 모델을 구축하였다.

통계 기반의 모델은 IBM SPSS Statistics를 사용하여 구축하였으며, 머신러닝 기반의 모델은 R Programming을 사용하여 구축하였다. DT 알고리즘은 'C5.0'을, SVM 알고리즘은 'kernlab'을 채택하였다. kernlab 알고리즘은 비선형 SVM 분석 방법이 가능한 선형커널, 다항커널, RBF커널 방법을 포함하고 있다.

각각의 모델은 데이터 중 무작위로 선정된 70%의 훈련데이터와 30%의 테스트데이터로 구분하여 모델 구축과 추정 및 유효성 검증을 동시에 진행하였다. 70% 수준의 훈련데이터를 사용해서 모델을 구축하고, 구축된 모델에 30% 수준의 테스트데이터를 적용하여 모델의 유효성을 검증하였다. 분석결과에서 다를 각 모델의 추정 정확도 결과는 해당 시기의 전체 지역을 대상으로 하여 추정된 결과가 아닌, 전체 지역 중 70%의 훈련데이터를 기반으로 30%에 해당하는 지역에 대한 추정 정확도를 나타낸다. 이러한 과정을 통해 모델의 논리적 결함 여부를 확인하고, 향후 북한 도시 토지이용 변화 시뮬레이션 분석에 적용이 가능한지 검증할 수 있었다. 해당 분석은 단계별로 진행하기 위해 해당 단계 기간 동안 도로 포장화 및 철도역의 신설, 일부 도심지의 변화 등을 시나리오별로 적용하였다.

Figure 2_ The Simulation Model Diagram



본 논문에서의 시뮬레이션 분석은 이상준, 김찬규, 박세훈, 신혜원(2011)의 연구에서 제안한 남포특별시의 개발가능 후보지 중 일부 지역을 중심으로 도로 포장화 및 철도역의 신설, 일부 도심지의 변화 등의 시나리오를 적용하였다.

IV. 분석결과

1. 시뮬레이션 모델 구축

1) 다항로지스틱 회귀 모델

1단계를 대상으로 다항로지스틱 회귀 모델을 적용하여 토지이용 변화를 추정한 분석 중 토지이용 기준이 밭과 논인 경우에 대한 정확도 결과는 <Table 2, 3>과 같다. 각 테이블의 숫자는 실제 토지이용 대비 추

Table 2_ Estimation Accuracy of Polynomial Regression Model (Farm/Step 1)

Actual Land Uses	Estimate Land Uses				Accuracy
	Urban	Farm	Paddy	Forest	
Urban	20,187	29,291	-	61	40.7%
Farm	6,478	112,488	20	399	94.2%
Paddy	214	5,268	524	55	8.6%
Forest	29	8,917	-	1,544	14.7%
Total	26,908	155,964	544	2,059	72.6%

Table 3_ Estimation Accuracy of Polynomial Regression Model (Paddy/Step 1)

Actual Land Uses	Estimate Land Uses				Accuracy
	Urban	Farm	Paddy	Forest	
Urban	3	1,241	18,166	-	0.0%
Farm	19	7,938	41,979	-	15.9%
Paddy	42	5,882	119,396	-	95.3%
Forest	2	121	866	-	0.0%
Total	66	15,182	180,407	-	65.1%

정한 토지이용 격자의 개수를 의미하며, 이를 근거로 하여 정확도를 나타낸다.

1단계의 밭과 논을 대상으로 진행한 다항로지스틱 회귀 모델의 추정 정확도 결과 기준이 되는(토지이용 변화가 발생하지 않는) 변수에 대한 추정 정확도만 높게 나타나며, 타 토지이용 변화에 대한 추정 정확도는 현저하게 낮게 나타났다. 모델이 케이스별 균형 있는 추정 정확도를 나타내야 한다는 측면에서 다항로지스틱 회귀 모델은 본 논문에서 적합한 분석 방법이 아닌 것으로 판단되었다.

2) DT 모델

1단계를 대상으로 DT 모델을 적용하여 토지이용 변화를 추정한 분석 중 토지이용 기준이 밭과 논인 경우에 대한 정확도 결과는 <Table 4, 5>와 같다.

Table 4_ Estimation Accuracy of Decision Tree Model (Farm/Step 1)

Actual Land Uses	Estimate Land Uses				Accuracy
	Urban	Farm	Paddy	Forest	
Urban	13,195	1,562	70	118	88.29%
Farm	1,310	33,858	227	371	94.67%
Paddy	104	544	1,165	20	63.56%
Forest	209	831	18	2,040	65.85%
Total	14,818	36,795	1,480	2,549	90.32%

Table 5_ Estimation Accuracy of Decision Tree Model (Paddy/Step 1)

Actual Land Uses	Estimate Land Uses				Accuracy
	Urban	Farm	Paddy	Forest	
Urban	5,812	597	1,307	96	74.40%
Farm	572	17,285	4,755	66	76.22%
Paddy	861	3,346	58,547	82	93.17%
Forest	122	143	173	1,131	72.08%
Total	7,367	21,371	64,782	1,375	87.23%

Figure 3 _Actual Map in Incheon (2005)

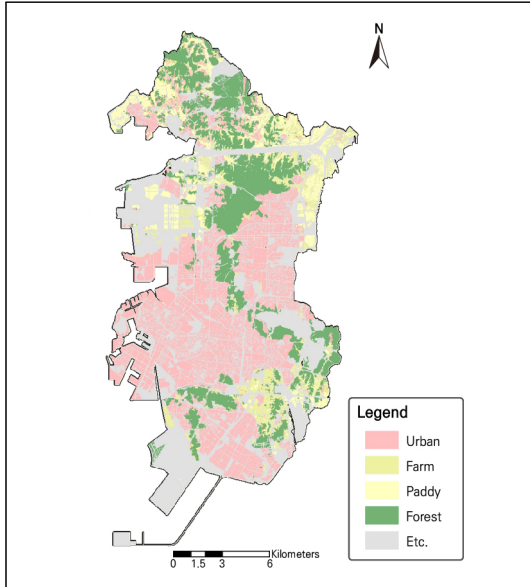
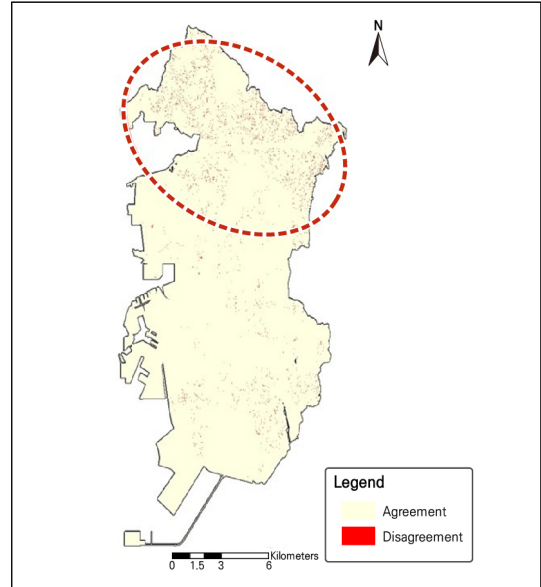


Figure 4 _Compare Map (DT Model, 2005)



1단계 DT 모델의 추정 결과는 앞선 다항로지스틱 회귀 모델과 상이하게 높은 추정 정확도와 함께 상대적으로 각각의 케이스별로 균형 있는 추정 정확도 결과를 나타냈다. <Figure 3>은 2005년 기준 연구지역의 실제 토지이용 현황이며, <Figure 4>는 DT 모델을 통해 추정한 토지이용과 실제의 토지이용 간 일치 여부를 나타낸 지도이다.

전반적으로 도심지역은 높은 추정 정확도를 나타냈으며, 상대적으로 북부 지역인 비도심지역의 추정 정확도는 낮게 나타났다. 전체 격자 326만 8,991개 중 특수용지와 공공용지의 132만 908개를 제외하고 본문에서 실제 모델링하는 데 적용한 격자인 194만 8,083개를 대상으로 하여 추정 결과와 실제 결과가 일치하지 않은 케이스를 살펴본 결과, 3만 4,450건이었다. 이는 전체 케이스를 기준으로 1.77%만 해당하는 높은 정확도 추정 결과이다. 해당 결과는 훈련데이터로 추정한 결과와 실험데이터로 추정한 결과까지 포함하여 나타낸 것이다. 도심지역과 가까울수록 높은 추

정 정확도를 나타내고, 비도심지역과 가까울수록 상대적으로 낮은 추정 정확도 결과를 나타냈다.

2단계를 대상으로 DT 모델을 적용하여 토지이용 변화를 추정한 분석 중 토지이용 기준이 발인 경우에 대한 정확도 결과는 <Table 6>과 같다.

2단계 DT 모델의 추정 결과는 역시 높은 추정 정확도와 함께 상대적으로 각각의 케이스별로 균형 있는 추정 정확도 결과를 나타냈다.

<Figure 5>는 2015년 기준 연구지역의 실제 토지

Table 6 _ Estimation Accuracy of Decision Tree Model (Farm/Step 2)

Actual Land Uses	Estimate Land Uses				Accuracy
	Urban	Farm	Paddy	Forest	
Urban	11,440	2,626	45	42	80.83%
Farm	1,298	48,553	134	157	96.83%
Paddy	48	577	575	4	47.76%
Forest	92	870	9	692	41.61%
Total	12,878	52,626	763	895	91.21%

Figure 5 _Actual Map in Incheon (2005)

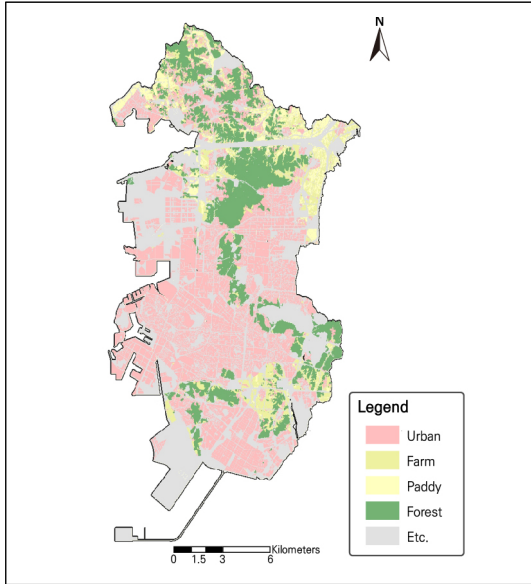
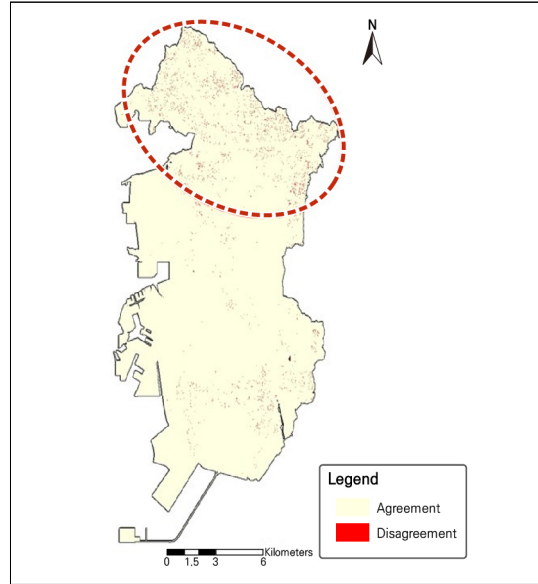


Figure 6 _Compare Map (DT Model, 2015)



이용 현황이며, <Figure 6>은 DT 모델을 통해 추정된 토지이용과 실제의 토지이용 간 일치여부를 나타낸 지도이다. 추정 결과와 실제 결과가 일치하지 않은 케이스는 3만 1,839건으로 전체 케이스 중 1.63%만 해당하는 높은 추정 정확도 결과이다.

3) SVM 모델

1단계를 대상으로 SVM 모델 중 선형커널 알고리즘을 적용하여 토지이용 변화를 추정한 분석 중 토지이용 기준이 발인 경우에 대한 정확도 결과는 <Table 7>과 같으며, 다항커널 알고리즘을 적용하여 토지이용 변화를 추정한 분석 중 토지이용 기준이 논인 경우에 대한 정확도 결과는 <Table 8>과 같다.

<Table 7, 8>의 결과는 앞선 다항로지스틱 회귀 모델의 추정 결과와 유사한 양상을 나타내었다. 기준이 되는 토지 혹은 다량의 변화가 이루어진 케이스만 추정을 하고, 상대적으로 적은 변화가 이루어진 케이스의 추정 정확도는 현저하게 낮게 나타났다. 해당 선

Table 7 _ Estimation Accuracy of Linear-SVM Model (Farm/Step 1)

Actual Land Uses	Estimate Land Uses				Accuracy
	Urban	Farm	Paddy	Forest	
Urban	5,926	9,019	-	-	39.65%
Farm	1,779	33,987	-	-	95.03%
Paddy	41	1,792	-	-	0.00%
Forest	9	3,089	-	-	0.00%
Total	7,755	47,887	-	-	71.73%

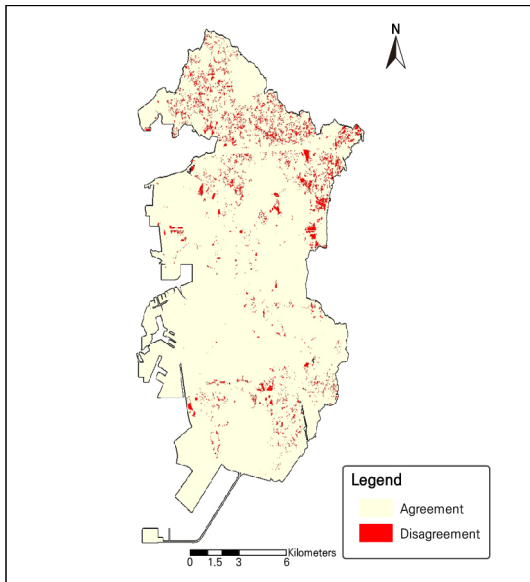
Table 8 _ Estimation Accuracy of Polynomial-SVM Model (Paddy/Step 1)

Actual Land Uses	Estimate Land Uses				Accuracy
	Urban	Farm	Paddy	Forest	
Urban	8,255	-	4,585	-	64.29%
Farm	248	-	15,341	-	0.00%
Paddy	141	-	38,216	-	99.63%
Forest	257	-	391	-	0.00%
Total	8,901	-	58,533	-	68.91%

Table 9 _ Estimation Accuracy of RBF-SVM Model (Farm/Step 1)

Actual Land Uses	Estimate Land Uses				Accuracy
	Urban	Farm	Paddy	Forest	
Urban	9,747	5,142	-	56	65.22%
Farm	1313	34,345	7	101	96.03%
Paddy	51	1,566	212	4	11.57%
Forest	92	2,149	4	853	27.53%
Total	11,203	43,202	223	1,014	81.16%

Figure 7 _ Compare Map (SVM Model, 2005)



형 및 다항커널 기반 SVM 모델은 본 논문에서 적합하지 않은 것으로 판단하였다. 반면 RBF커널 기반 SVM 모델의 추정 결과는 <Table 9>와 같이 높은 추정 정확도와 함께 상대적으로 케이스별 균형 있는 추정 정확도 결과를 나타냈다.

<Figure 7>은 RBF 기반 SVM 모델을 통해 추정된 토지이용과 실제의 토지이용 간 일치 여부를 나타내는 지도로서, 추정 결과와 실제 결과가 일치하지 않은 케이스는 16만 2,015건으로 전체 케이스 중 8.32%만 해당된다.

4) 모델 구축 결과

분석결과 통계분석 기반의 로지스틱 회귀 모델과 머신러닝 기반의 선형커널 SVM 모델과 다항커널 SVM 모델은 본 논문의 모델로 적합하지 않은 것으로 확인되었다. 로지스틱 회귀 모델과 선형 및 다항커널 기반의 SVM 모델은 비선형 분석을 위해 데이터를 전혀 변환하지 않고 단순히 속성 내에서만 분석이 이루어지는 알고리즘이기 때문에, 본 논문과 같이 미시적이고 복잡한 토지이용 변화 패턴을 탐색하는 데 어려움이 있었던 것으로 판단된다.

<Table 10>은 DT 모델과 RBF커널을 활용한 SVM 모델의 추정 정확도를 비교한 것이다. 해당 추정 정확도는 모든 대상(변화한 토지+변화하지 않은 토지이용)의 토지에 대한 결과이며 균형 있는 추정을 위해 가중산술평균 방법을 통해서 산출된 것이다.

해당 분석결과를 바탕으로 본 논문에서 채택한 최종 시뮬레이션 모델은 <Figure 8>과 같다. 본 논문에서 구축한 총 40개의 세부 모델 중 가장 정확도가 높은 모델 8개를 선정하여 모델을 구축하였다. DT 알고리즘과 RBF 기반 SVM 알고리즘으로 구현된 모델 중 높은 정확도 결과를 나타낸 모델을 각각 채택하여 결합하였다.

Table 10 _ Estimation Accuracy of DT Model and SVM Model

(unit: %)			
Step	Land Uses	DT Model	SVM Model (RBF based)
Step 1	Urban	98.92	98.83
	Farm	91.10	87.69
	Paddy	88.03	90.13
	Forest	98.03	98.17
Step 2	Urban	99.79	99.80
	Farm	92.76	92.28
	Paddy	88.42	90.69
	Forest	97.02	96.39

Figure 8_ The Hybrid Simulation Model

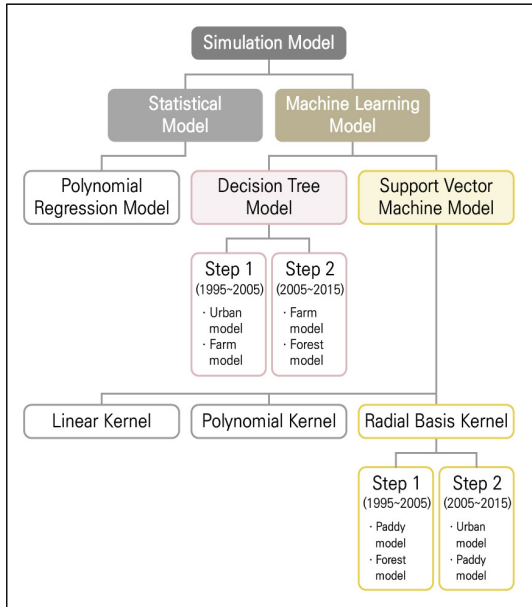


Figure 9_ Actual Map in Nampo (2015)

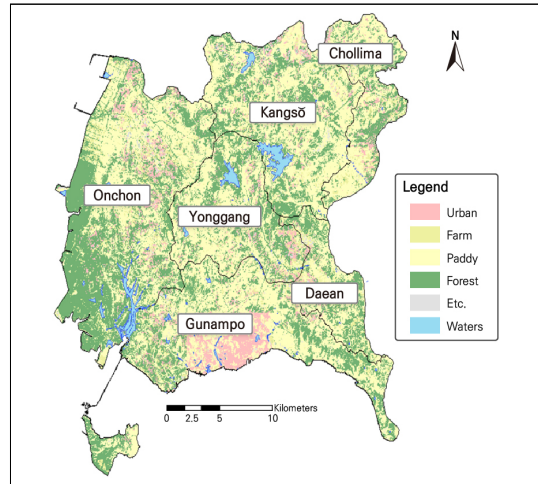
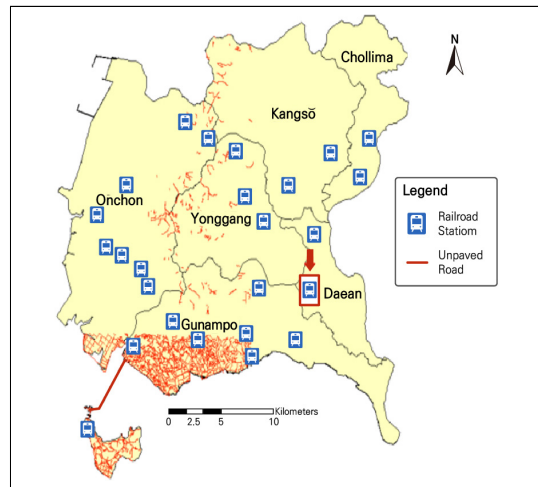


Figure 10_ Simulation Scenario (Step 1)



2. 시뮬레이션 분석

1) 1단계 시뮬레이션 분석

위성영상 자료를 활용하여 무감독분류한 남포특별시의 2015년 기준 토지피복 현황은 <Figure 9>와 같다. 도심지역은 구 남포시를 중심으로 분포되어 있다.

1단계 분석은 <Figure 10>과 같이 2015년 기준 남포특별시의 비포장도로 일부의 포장화와 함께 대안군 일대에 추가 철도역 1개소를 신설하는 가정으로 토지 이용 변화 시뮬레이션 분석을 진행하였다.

시뮬레이션 분석결과는 <Table 11>과 같다. 1단계 분석에서 주로 추정된 케이스는 밭이 도심지역으로 변화한 경우와 논과 산림지역이 밭으로 변화한 경우가 주를 이룬다. 초기 도시화가 진행되는 과정에서 수도 및 도심지와와의 교통 접근성이 향상되면서 도심지의 발달과 함께 논보다 상대적으로 고부가가치를 창출할 수 있는 밭으로 토지이용 변화가 다수 발생하

Table 11_ Estimation of Land Use Change in Nampo (Step 1)

(unit: %, km²)

Present Land Uses	Estimate Land Uses	Estimate Case Ratio	Area Changed
Farm	Urban	12.55	17,900
Paddy		0.30	400
Forest		0.05	70
Paddy	Farm	61.99	88,400
Forest		11.59	16,500
Farm	Paddy	2.93	4,200
Forest		0.06	85

는 현상으로 해석할 수 있다.

2) 2단계 시뮬레이션 분석

2단계 분석은 <Figure 11>과 같이 1단계 분석에서 시행되지 않은 2015년 기준 남포특별시 비포장도로 일부의 포장화와 함께 용강군 일대에 추가 철도역 1개소 신설, 이상준, 김천규, 박세훈, 신혜원(2011)의 연구에서 제시한 남포특별시의 개발가능후보지 중 천리마군 일대에 도심지역을 개발하는 것을 가정하여 토지이용 변화 시뮬레이션 분석을 진행하였다.

2단계 분석 결과 <Table 12>와 <Figure 12>와 같은 결과를 나타냈다. 1단계의 분석에 비해 상대적으로 도심지역으로의 변화가 다수 발생할 것으로 추정된 것이 특징이다. 남포특별시의 주요 도로를 중심으로 하여 도심지역으로의 변화가 이루어질 것으로 추정한 결과와 함께, 논과 산림지역의 밭으로의 변화, 산림지역의 논으로의 변화를 다수 추정한 결과를 확인할 수 있다.

Figure 11_ Simulation Scenario (Step 2)

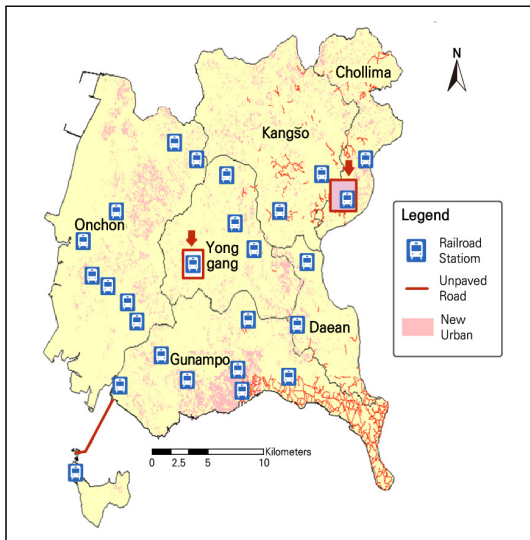
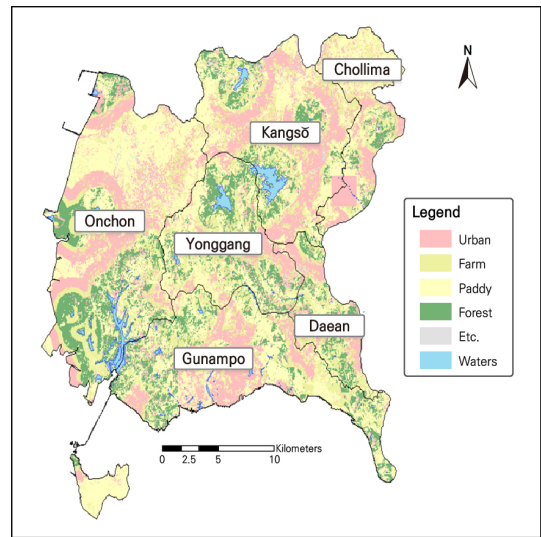


Table 12_ Estimation of Land Use Change in Nampo (Step 2)

(unit: %, km²)

Present Land Uses	Estimate Land Uses	Estimate Case Ratio	Area Changed
Farm	Urban	3.86	4,200
Paddy		20.91	77,100
Forest		27.13	100,000
Paddy	Farm	13.19	48,600
Forest		15.30	56,400
Farm	Paddy	0.06	200
Forest		19.44	71,700

Figure 12_ Simulation Map (Step 2)



V. 결론

본 논문은 시뮬레이션 분석을 통해 통일 이후 북한 지역의 도시 토지이용 변화를 추정하는 것을 목적으로 설정하였다. 분석대상 도시는 남·북한 도시의 도시 유사도 분석을 진행한 선행연구(원석환, 황철수 2018)를 바탕으로 남한 인천광역시와 과거 토지이용 변화 패턴을 기반으로 머신러닝을 활용한 시뮬레이션을 통하여 통일 이후 북한 남포특별시의 토지이용 변화 양

상을 추정하였다.

1단계의 분석결과 기존 도심지역인 구 남포시와, 평양직할시와 접근성이 높은 천리마군, 강서군 중심으로 논이 밭으로 변화할 것으로 추정된 결과가 다수 확인되었다. 구 남포시와 용강군 일대의 교통수용능력을 향상시킴으로써 강서군과 천리마군 또한 주도심지역과 교통접근성이 향상된 결과로 판단된다.

2단계의 분석결과는 도심지역으로 다수의 변화가 발생할 것으로 추정하였다. 토지이용 변화 시뮬레이션 모델링 단계에서 확인한 2단계 기간 동안 급속적인 도시화 현상을 나타낸 인천광역시와 토지이용 패턴이 적용된 것을 확인할 수 있는 결과이다.

지역의 교통접근성 변화와 신규 도심지역의 개발은 해당 지역의 도시화는 물론이며 타 지역까지 연계되어 도시화 현상이 나타나게 되는 것을 확인할 수 있었다. 따라서 교통시설의 신설과 신규 도심지역의 개발을 위한 정책수립 단계에서 해당 지역의 발전과 함께 타 지역과의 도심 연계성을 고려하는 것이 중요할 것임을 고찰할 수 있었다.

본 논문에서는 인구변수를 고려하지 않았으나 도시변화 연구에서 인구변수는 주요한 요소이다. 추가적인 인구이동성의 특징을 포함한 도시변화 추정 연구가 필요할 것이다.

통일은 국토의 지속가능한 발전을 위해 달성해야 할 중요한 과제임을 인식하여, 통일 대비를 위한 다양한 학제 간 연구가 지속적으로 진행되어야 할 것이다. 이러한 관점에서 본 논문은 통일 한국 시대의 북한 도시화 현상을 미시적으로 추정하여, 향후 북한의 도시의 개발정책 수립 시 전략적인 대응을 할 수 있는 측면에서 연구의 의의를 지닌다.

참고문헌 •••••

1. 강영옥, 손세린, 조나혜. 2017. 의사결정나무와 시공간 시각화를 통한 서울시 교통사고 심각도 요인 분석. *지적과 국토정보* 47권, 2호: 233-254.
Kang Youngok, Son Serin and Cho Nahye. 2017. Analysis of traffic accidents injury severity in Seoul using decision trees and spatiotemporal data visualization. *Journal of Cadastre & Land InformatiX* 47, no.2: 233-254.
2. 강톨가. 2014. 고해상도 위성영상 자료로부터 토지피복분류를 위한 다중 SVM 비교 연구. 석사학위논문, 전북대학교.
Gangtulga. 2014. *A Comparative Study on Multi-class SVM for Land Cover Classification Using High Resolution Satellite Image*. M.D. diss., Chonbuk National University.
3. 구유희. 2007. 의사결정나무와 로지스틱 회귀분석을 이용한 태권도 수련생 이탈 예측을 위한 비교 연구. 석사학위논문, 한양대학교.
Koo Youhee. 2007. *Comparative Analysis of Prediction Taekwondo Trainee's Defection Using Decision Tree and Logistic Regression*. M.D. diss., Hanyang University.
4. 국토교통부 국가공간정보센터. <http://www.nsd.go.kr> (2017년 8월 5일 검색)
Korea National Spatial Data Infrastructure Portal. <http://www.nsd.go.kr> (accessed August 5, 2017)
5. 국토지리정보원. 2015. 북한 위성영상 자료. 내부자료.
National Geographic Information Institute. 2015. North Korea satellite image data. raw data.
6. 김영민, 황미영, 김태홍, 정창후, 정도현. 2015. 자연재해 분석을 위한 빅데이터 마이닝 기술. *한국데이터정보과학회지* 26권, 5호: 1105-1115.
Kim Youngmin, Hwang Minyeong, Kim Taehong, Jeong Changhoo and Jeong Doheon. 2015. Big data mining for natural disaster analysis. *Journal of the Korean Data and Information Science Society* 26, no.5: 1105-1115.
7. 김종윤, 박수홍. 2004. 의사결정트리를 이용한 개별 공시지가 비교표준지의 자동 선정. *한국지리정보학회지* 7권, 1호: 9-19.
Kim Jongyun and Park Suhong. 2004. An automatic method for selecting comparative standard land parcels in land price appraisal using a decision tree. *The Korean Association of Geographic Information Studies* 7, no.1: 9-19.
8. 박선영, 장은나, 임정호, 최재원. 2013. 원격탐사와 기계학습

- 을 이용한 산사태 취약성 평가. 2013 한국측량학회 춘계학술 발표회 자료집 5권: 307-309.
- Park Seonyoung, Jand Eunna, Im Jungho and Chio Jaewon. 2013. Assessment of landslide susceptibility using remote sensing and machine learning approaches. *Proceedings of the 2013 Korean Society of Surveying Geodesy Photogrammetry and Cartography Conference* 5: 307-309.
9. 배정수. 2014. 의사결정나무 분석 기법을 활용한 고등학생 진로결정수준 및 진로준비행동 결정요인 우선순위 탐구. 박사학위논문, 순천향대학교.
- Bai Jungsu. 2014. *A Study on Priority of Determinants of Career Decision Level and Career Preparation Behavior in High School Students based on Decision Tree Analysis*. Ph.D. diss., Soonchunhyang University.
10. 변병설, 윤갑식. 2001. 통일시대에 대비한 국토환경관리 방안. 서울: 한국환경정책·평가연구원.
- Byun Byungseol and Yun Kapsik. 2001. *Strategies on National Territorial Environment for preparing Unification*. Seoul: Korea Environment Institute.
11. 신은주, 장남식. 1999. 데이터마이닝 기법 비교 연구: 단일 및 복수 의사결정나무. 경영정보학회 춘계학술대회 자료집: 361-369.
- Shin Eunjuo and Chang Namsik. 1999. Comparisons of tree-based data mining techniques: Single vs. multi-decision trees. *Proceedings of the The Korea Society of Management Information Systems Conference*: 361-369.
12. 원석환. 2010. 접근성 변동에 따른 도시토지이용의 시·공간적 변화에 관한 연구. 석사학위논문, 경희대학교.
- Won Seokhwan. 2010. *A Study on the Spatiotemporal Change of Urban Land Use according to the Change of Accessibility*. M.D. diss., Kyunghee University.
13. 원석환, 황철수. 2018. 공간적 도시 유사도 분석: 남·북한 도시를 사례로. 국토지리학회지 52권, 1호: 127-137.
- Won Seokhwan and Hwang Chulsue. 2018. Urban similarity analysis concerning the spatial factors: A comparative study of major cities in South Korea and North Korea. *The Geographical Journal of Korea* 52, no.1: 127-137.
14. 이상준, 김천규, 박세훈, 신혜원. 2011. 통일 한반도 시대에 대비한 북한 주요 거점의 개발잠재력과 정책과제(I). 안양: 국토연구원.
- Lee Sangjun, Kim CheonKyu, Park Sehoon and Shin Hyewon. 2011. *Development Issues of the Growth Centers in North Korea for Preparing Korean Unification(I)*. Anyang: Korea Research Institute for Human Settlements.
15. 이상준, 박세훈, 남기찬, 이건민. 2014. 통일기반 강화를 위한 북한 거점도시 발전모형과 남북협력 실천전략 연구. 안양: 국토연구원.
- Lee Sangjun, Park Sehoon, Nam Gichan and Lee Geonmin. 2014. *A Study on Urban Development Models and Implementation Strategies for Selective North Korean Cities*. Anyang: Korea Research Institute for Human Settlements.
16. 전주원. 2011. 벌점화 부분최소제곱과 벌점화 서포트 벡터 머신의 성능 비교. 석사학위논문, 성균관대학교.
- Jeon Juewon. 2011. *Performance Comparison of Penalized Partial Least Squares and Penalized Support Vector Machines*. M.D. diss., Sungkyunkwan University.
17. 정재준. 2001. 수도권권의 도시성장 분석 및 예측을 위한 셀룰라 오토마타 모델링. 박사학위논문, 서울대학교.
- Jeong Jaejoon. 2001. *Cellular Automata Modeling for the Analysis and Prediction of Urban Growth in Seoul Metropolitan Area*. Ph.D. diss., Seoul National University.
18. 통계청. 2016. 북한의 주요통계지표. 대전: 통계청.
- Statistics Korea. 2016. *Major Statistics Indicators of North Korea*. Daejeon: Statistics Korea.
19. _____. <http://kostat.go.kr> (2017년 8월 5일 검색).
- _____. <http://kostat.go.kr> (accessed August 5, 2017)
20. Alghais, N. and Pullar D. 2018. Modelling future impacts of urban development in Kuwait with the use of ABM and GIS. *Transactions in GIS* 22: 20-42.
21. Chan, J. C., Chan, K. and Yeh, A. G. 2001. Detecting the Nature of change in an urban environment: A comparison of machine learning algorithms. *Photogrammetric Engineering & Remote Sensing* 67, no.2: 213-225.
22. Dixon, B. and Candade, N. 2008. Multispectral landuse classification using neural networks and support vector machines. *International Journal of Remote Sensing* 29, no.4: 1185-1206.
23. Jegou, David. 2005. 다중 에이전트의 상호 협력과 경쟁을 통한 호이스트 스케줄링 기법에 관한 연구. 박사학위논문, 한국과학기술원.
- Jegou, David. 2005. *Negotiation and Cooperation in Multi-agent System: Application to the Hoist Scheduling Problem*. Ph.D. diss., Korea Advanced Institute of Science and Technology.

24. Lantz, Brett. 2013. *Machine Learning with R*. Birmingham: Packt Publishing Ltd.

25. Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M. and Islam, T. 2012. Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research* 50, no.9: 1250-1265.

27. USGS. SRTM 30m. <https://earthexplorer.usgs.gov> (accessed August 10, 2017)

28. UN. 2014. *World Urbanization Prospects*. New York, NY: UN.

- 논문 접수일: 2018. 4. 10.
- 심사 시작일: 2018. 5. 3.
- 심사 완료일: 2018. 6. 21.

요약

주제어: 머신러닝, 토지이용/피복 변화, 북한 연구, 도시 시뮬레이션 분석, 접근불능지역 분석

본 논문의 목적은 시뮬레이션 분석을 통한 통일 이후 북한 특정 도시를 대상으로 하여 토지이용 변화를 추정하는 것이다. 시뮬레이션 분석은 도시화 패턴 탐색을 위한 모델 적합도 연구와 시뮬레이션 결과에 따른 유효성 검증, 북한 도시 토지이용 변화 시뮬레이션 분석으로 구분된다. 도시화 패턴 탐색을 위한 모델링에 있어 적합한 방법론을 선정하기 위해서 통계 분석의 일종인 다항로지스틱 회귀분석과 머신러닝의 일종인 의사결정나무, 서포트 벡터 머신 알고리즘을 활용한 분석을 통해 토지이용 변화 추정 정확도를 비교 분석하였다. 이를 통해 시뮬레이션 단계 및 기준 토지이용별 추정 정확도가 가장 높은 모델을 결합한 시

뮬레이션 통합 모델을 구축하였다. 해당 모델을 활용하여 북한 도시(남포특별시)의 시나리오별 토지이용 변화를 추정하였다. 남포특별시의 토지이용 변화 시뮬레이션 분석은 단계별 비포장도로 일부의 포장화, 철도역 신설, 신규 도심지역 개발을 가정한 시나리오로 진행하였다. 분석결과, 지역의 교통 접근성 변화와 신규 도심지역의 개발 영향은 해당 지역의 도시화는 물론이며 타 지역까지 연계되어 도시화 현상이 나타나게 되는 것을 확인할 수 있었다. 본 논문은 통일 한국 시대의 북한 도시화 현상을 미시적으로 추정하여, 향후 북한 도시의 개발정책 수립 시 전략적인 대응을 할 수 있는 측면에서 연구의 의의를 지닌다.