

입지특성을 고려한 토지가격의 차등적 산정방안 - 산업시설용지 공급가격을 중심으로 -

A Differential Pricing Model for Industrial Land based on Locational Characteristics

심재헌*

Shim, Jae Heon

Abstract

This paper proposes a differential pricing model for industrial land based on locational characteristics, using Support Vector Regression (SVR) as a land pricing methodology. The initial selling price of industrial land is set based on the total cost of site development that comprises the land acquisition cost and tax, land development expense, infrastructure installation cost, labor cost, migration expense, selling and administrative expense, capital cost, and so on. However, the current industrial land pricing method unreasonably applies the same price per square meter to all parcels within an industrial complex without considering differences in price depending on the location of each parcel. Therefore, this paper proposes an empirical land pricing model to solve this irrationality and verifies its validity and applicability.

Keywords : differential pricing model, industrial land, locational characteristics, support vector regression

요 지

본 연구는 지가산정을 위한 방법론으로 서포트 벡터 회귀분석(SVR)을 이용하여 입지특성에 근거한 산업시설용지의 차등적 가격산정모형을 제시하고자 하였다. 기업의 생산활동을 위한 산업시설용지의 공급가격은 용지부담금, 조성비, 판매 및 관리비, 자본비용 등으로 구성된 조성원가를 바탕으로 책정된다. 그러나 현행 산업시설용지 공급가격 산정방식은 개별적인 용지의 입지여건을 고려하지 않은 채 동일한 가격을 단지 내 모든 용지에 획일적으로 적용하는 불합리성을 내포하고 있다. 따라서 본 연구는 이를 지양하기 위한 방안으로 산업시설용지의 차등적 가격산정모형을 제시하고, 모형의 적정성 및 실무적인 적용가능성을 평가하는 과정으로 이루어진다.

핵심용어 : 차등적 가격산정모형, 산업용지, 입지특성, 서포트 벡터 회귀분석

1. 서 론

1.1 연구배경 및 목적

토지는 여타의 재화와는 달리 고정성과 개별성을 지닌 재화이므로 입지가 재화의 질적인 수준을 결정하는 대표적인 이질재이다. 토지가격은 토지를 둘러싸고 있는 수많은 요인에 의해서 형성되고 영향을 받지만, 공간을 연구하는 학자와 부동산 전문가들 사이에서는 입지가 토지가격에 영향을 미치는 가장 중요한 결정요인이라는 보편적인 동의가 존재하여 왔다. 일반적으로 토지 및 주택의 가격에 영향을 미치는 입지성 요인은 접근성(accessibility), 주변지역의 특성(neighborhood characteristics), 구조적 특성(structural characteristics) 등으로 분류될 수 있다(Olmo, 1995). 이와 유사하게, 실무적 차원에서도 공시지가 산정의 경우와 같이 부지특성, 주위환경, 접근성 등의 성격을 지닌 토지특성항목¹⁾에 근거하

여 토지가격을 산정한다.

기업의 생산활동을 위한 토지인 산업시설용지의 공급가격은 「산업입지 및 개발에 관한 법률 시행령」 제40조(분양가격의 결정 등)에 규정된 바와 같이 용지비, 용지부담금, 조성비, 기반시설 설치비, 이주대책비, 판매비, 일반관리비, 자본비용 등 10개의 항목으로 이루어진 조성원가에 근거하여 구성된다. 그러나 현행 산업시설용지 공급가격 산정방식은 조성원가를 바탕으로 한 가격책정에 있어서 개별적인 용지의 입지여건을 고려하지 않은 채 동일한 단위면적당 가격을 단지 내 모든 산업시설용지에 획일적으로 적용하는 불합리성을 내포하고 있다. 즉, 동일한 산업단지 내의 용지일지라

1) 토지특성항목이란 토지가격형성에 중요한 요인으로 작용하고 토지관련자료의 정보요인으로 가치가 있는 것 중 토지특성조사표에 기재된 항목을 말한다.

*정희원 · Research Associate · Department of Urban & Regional Planning, University of Illinois at Urbana-Champaign (E-mail : cogito128@gmail.com)

도 주변 토지이용상황, 용도지역, 지형적 특성, 접근성 등의 입지특성에 따라서 기업의 생산활동에 대한 상이한 여건이 형성되지만, 현행 산정방식에 있어서 이와 같은 요인이 전혀 고려되지 않고 있다. 이와 관련하여 국가경쟁력강화위원회의 제3차 회의(2008)에서는 현재 모든 산업시설용지가 입지여건 등에 관계없이 조성원가 기준에 의한 추첨방식으로 일률적으로 공급됨에 따라 발생하는 용지위치에 따른 가격차이 및 수요자의 선호가 반영되지 않는 불합리한 문제점을 검토한 바 있다.

따라서 본 연구는 현행 산업시설용지 공급가격 산정방식의 불합리성을 지양하기 위한 방안으로 입지특성에 근거한 산업시설용지 공급가격의 차등적 산정모형을 제시하는 데 그 목적이 있다.

1.2 연구범위 및 방법

입지특성에 근거한 산업시설용지 공급가격의 차등적 산정모형을 구축하기 위하여 총 297필지의 산업용지에 대한 가격데이터(2004년도 기준) 및 입지특성데이터로 구성된 데이터를 사용한다²⁾. 정밀한 지가 산정모형을 구축하기 위해서는 신뢰할 수 있는 가격데이터가 요구되는데, 실거래가격으로 간주할 수 있는 법인체 간의 거래가격 및 법인체와 개인 간의 거래가격만을 사용하며, 토지와 건물이 결합되어 구성된 복합부동산의 경우에는 거래가격에서 토지가격만을 추출하는 과정을 수행한다. 또한, 산업시설용지의 가격형성과 유의미한 관련성을 가지는 입지특성요인을 선정하기 위하여, 지가형성요인에 관한 선행연구를 바탕으로 산업단지와 유관한 전문가 집단을 대상으로 하는 델파이 조사(delphi method)를 시행한다. 델파이 조사는 전문가들에게 선행연구를 참고하여 입지특성요인의 항목을 예측하도록 요구하는 개방형 면담조사와 구조화된 설문(델파이 폐쇄형 문항)조사로 구성된다.

지가 산정모형은 Vapnik, Golowich 및 Smola(1997)가 제안한 서포트 벡터 회귀분석(Support Vector Regression, 이하 SVR)을 이용하여 구축한다. 모형구축을 위하여 학습을 통한 모형의 최적화 과정이 시행되고, 이와 같은 과정에서 산출된 오차를 기준으로 최적적인 모형이 선정된다. 이어서 선정된 모형에 의한 실증적인 지가예측이 이루어지며, 예측결과를 바탕으로 모형의 적정성을 검증하는 과정이 수행된다. 마지막으로, 산업시설용지의 특수성과 공공성, 현행 법률의 취지를 고려하여 모형의 실무적인 적용을 위한 예측가격의 변환과정이 제시된다.

이상에서 언급한 본 연구의 흐름을 간략하게 도식화하면 그림 1과 같다.

2. 예비적 고찰

2.1 산업용지의 가격형성요인

산업용지의 가격형성요인을 규명하는 국외의 연구는 토지의 물리적 특성, 주변환경적 특성, 접근성, 거시적·미시적

2) 본 연구에서는 입지를 토지가격에 영향을 미치는 가장 중요한 결정요인으로 간주하는 기존의 연구에 기초하여, 토지의 시장 거래가격은 입지특성에 근거하여 형성된다는 점을 전제한다.

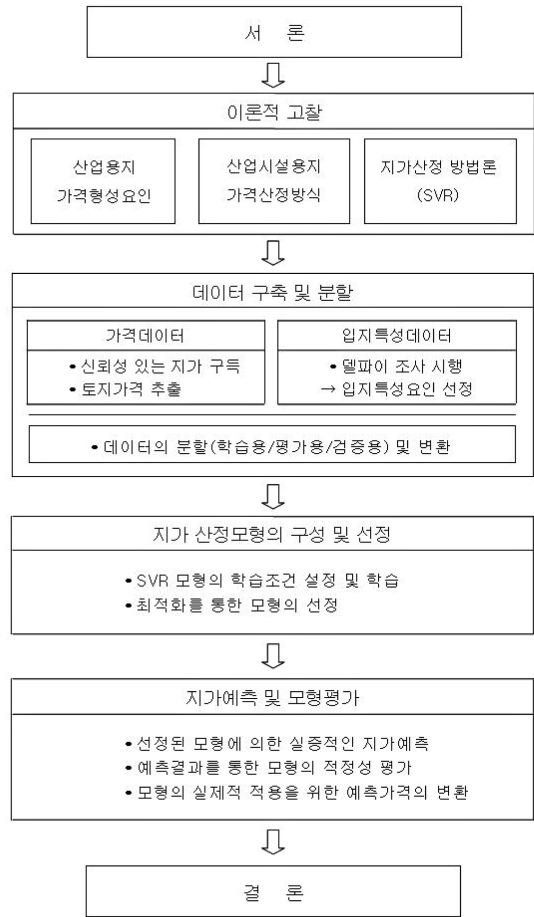


그림 1. 연구의 흐름

시장영향 등의 요인을 살펴보는 데에서 나아가 기반시설 확충 및 도시개발사업을 위한 재원조달 기법, 산업용지의 공급주체, 환경오염 등이 갖는 영향력을 분석하는 수준에 이르고 있다. 반면, 국내에서는 아직까지 이와 관련된 연구가 전무한 실정이므로 앞으로 폭넓은 연구가 수행될 필요성이 제기된다.

산업용지의 가격형성요인과 관련된 국외의 연구를 살펴보면, Saz-Salazar와 García-Menéndez(2005)는 Valencia 지역을 대상으로 공공부문과 민간부문에 의해서 공급되는 산업용지 가격 간의 차이를 유발하는 원인을 규명하고자 하였다. 헤도닉 가격기법(Hedonic Price Function Model)을 이용한 분석의 결과, 고속도로 거리, 도심 거리, 주도(州都) 거리가 산업용지의 가격에 중요한 영향을 미치며, 이와 같은 영향은 민간부문이 제공하는 산업용지에서 더 큰 것으로 나타났다. 또한, 산업용지의 공급주체에 따라서 산업용지의 가격은 큰 영향을 받는 것으로 나타났다. Weber 등(2003)은 Chicago 지역을 대상으로 TIF(Tax Increment Financing, 조세담보금융)가 산업용지의 가격에 미치는 영향을 중점적으로 살펴보았다. TIF 관련 요인, 부지요인, 위치요인, 주변환경적 요인을 고려하여 분석을 수행한 결과, 상업용 및 주거용 부동산이 분포하고 있는 복합용도의 TIF지역에 위치한 산업용지는 TIF지역 외에 위치한 산업용지에 비해서 가격이 높은 것으로 나타났다. 그러나 산업용 부동산이 동질적으로 분포한 TIF지역 내 산업용지는 TIF지역 외의 산업용지와 비교하여

가격이 높지 않거나, 때로는 오히려 낮은 것으로 파악되었다. Jackson(2002)은 Southern California 지역 내 산업용 부동산 거래자료를 이용하여 환경오염이 산업용 부동산 가격에 미치는 영향을 중점적으로 규명하고자 하였다. 그 결과, 환경오염은 산업용 부동산 가격에 통계적으로 유의한 영향을 미치나 환경이 개선된 이후에는 그 영향이 사라지는 것으로 나타났다. Lockwood와 Rutherford(1996)는 Dallas 및 Fort Worth 지역 내 산업용 부동산의 가격결정요인을 살펴보았다. 선형구조관계(LISREL) 모형을 이용하여 지가 영향력을 분석한 결과, 물리적 특성(부지면적, 산업·업무용 면적)과 지역적 시장요인(지역 고용률, 지역소득, 지역 총생산액), 위치요인(도심 거리, 공항 거리, 주요도로 거리, 철도 거리 등)이 주요한 가격결정요인으로 파악되었다. Guntermann(1995)은 폐기물 매립지(landfills)가 주변의 산업용지에 미치는 영향을 중점적으로 살펴보았다. 필지면적, 지역구분, 하천 인접성, 산업단지 내 포함 여부, 거래시기, 철도시설 이용 가능 여부, 고속도로·공항 접근성, 폐기물 매립지 인접성, 매립지 유형 등의 요인을 고려하여 분석한 결과, 밀폐되지 않은 매립지의 입지는 주변 산업용지의 가격을 떨어뜨리는 것으로 나타난 반면에 밀폐된 매립지의 입지는 가격에 영향을 미치지 못하는 것으로 나타났다. Kowalski와 Paraskevopoulos(1990)는 Detroit 지역 근교에 위치한 산업용지의 거래자료를 바탕으로 산업용지의 가격에 대한 입지 영향력을 분석하였다. 그 결과, 토지의 가시성, 산업단지 포함여부, 필지의 전면 폭, 거래시기, 매립지 조성가능성 등은 산업용지의 가격에 유의한 정(+)의 영향을, 필지면적, 경기침체 시기 등은 유의한 부(-)의 영향을 미치는 것으로 나타났다. 마지막으로 Asabere와 Huffman(1991)은 Philadelphia 지역을 대상으로 용도지역제(zoning)가 산업용지의 가격에 미치는 영향을 연구하였다. 용도지역 요인(주거·상업·공업·기타 지역)과 접근성 요인(도심 거리, 주요도로 거리), 사회·경제적 요인(가구소득, 폐가 비율) 등을 고려한 분석의 결과, 공업지역으로 지정된 토지는 상업지역의 토지와 비교하여 대략 58% 정도로 가격이 절하되는 것으로 나타났다.

이상에서 살펴본 바와 같이, 본 연구의 공간적 대상인 산업용지의 가격형성요인에 관한 국외의 연구는 산업용지의 공급주체, TIF, 고속도로 가시성, 산업단지 포함여부, 폐기물 매립지 등의 요인을 고려하는 점에서 다른 용도의 토지를 대상으로 하는 연구와 차별성이 있다고 할 수 있다. 특히 근래의 연구는 환경적인 측면(환경오염 및 폐기물 매립지 등)의 영향력을 비중있게 고려하는데, 이는 과거 주택시장을 대상으로 중점적으로 이루어졌던 기피요인(대기오염, 소음 등) 및 기피시설(발전소, 폐기물 처리시설 등)의 가격 영향력에 관한 연구와 유사한 양상을 보인다.

토지가격은 지속적으로 변화하고 수많은 요인의 상호작용에 의해서 영향을 받으므로 지가형성에 영향을 미치는 다양한 요인을 완전히 고려하기에는 현실적인 한계가 존재한다. 따라서 본 연구에서는 앞서 살펴본 국외의 선행연구를 토대로 산업단지와 유관한 전문가 집단을 대상으로 하는 델파이 조사(delphi method)를 시행하여 산업시설용지의 가격형성과 유의미한 관련성을 가지는 입지특성요인을 선정하였다.

2.2 산업시설용지 공급가격

2.2.1 산업시설용지 공급가격의 구성과 현황

지역 및 기업의 경쟁력을 제고하기 위해서는 저렴하고 기반시설이 잘 갖추어진 산업시설용지의 공급이 필수적이므로, 정부는 산업시설용지의 공급가격을 조성원가에 근거하여 책정하고 있으며 기반시설 일부와 환경시설 등을 지원하는 정책을 추진하고 있다. 이 중에서 산업단지 내 기업의 생산활동을 위한 산업시설용지의 공급가격은 일반적으로 「산업입지 및 개발에 관한 법률 시행령」 제40조(분양가격의 결정 등)에 의한 조성원가와 산업단지의 공급면적³⁾에 의해서 결정된다.

$$\text{공급가격(분양가)} = \frac{\text{사업비(조성원가)}}{\text{공급면적}} \quad (1)$$

산업단지의 조성에 투입된 사업비인 조성원가는 표 1에 제시된 바와 같이 용지비, 용지부담금, 조성비, 기반시설 설치비, 이주대책비, 판매비, 일반관리비, 자본비용 등 10개의 항목으로 구성된다. 이와 같은 조성원가의 구성항목은 크게 용지비, 조성비, 간접비, 자본비용으로 재분류할 수 있는데, 이 중에서 용지비가 가장 큰 비중을 차지하며 차위(次位)의 조성비는 산업단지의 유치를 위해서 지자체가 간선시설 설치비를 부담하는 점 등으로 인하여 택지개발비용의 약 70%

표 1. 산업단지 조성원가 산정표

항 목	내 역
용지비	용지매입비·지장물 등 보상비·조사비·등기비 및 그 부대비용
용지부담금	토지 등의 취득과 관련하여 부담하는 각종 부담금
조성비	해당 산업단지 조성에 소요된 직접비로서 조성공사비·설계비 및 그 부대비용
기반시설 설치비	해당 산업단지 조성에 필요한 기반시설 설치비용(다른 법령이나 인·허가조건에 따라 국가 또는 지방자치단체에 납부하는 부담금 및 공공시설설치비 등을 포함)
직접인건비	해당 사업을 직접 수행하거나 지원하는 직원의 인건비 및 복리후생비
이주대책비	이주대책의 시행에 따른 비용 및 손실액
판매비	광고선전비 및 그 밖의 판매에 소요된 비용
일반관리비	인건비, 임차료, 연구개발비, 훈련비, 그 밖에 일반 관리에 소요된 비용(직접인건비에 포함된 금액은 제외하되, 일반관리비율은 「국가를 당사자로 하는 계약에 관한 법률 시행령」 제9조에 따른 공사에 관한 비율을 초과할 수 없음)
자본비용	산업단지개발사업의 시행을 위하여 필요한 사업비의 조달에 소요되는 비용으로서 최초 실시계획에서 정하여진 사업기간(정부지원계획의 차질 기타 국토해양부령으로 정하는 불가피한 사유로 실시계획기간을 연장한 경우 동 기간을 포함한다)까지의 비용
그 밖의 비용	「산업재해보상보험법」에 따른 보험료 및 천재지변으로 인하여 발생하는 피해액 등 산업단지개발사업과 관련하여 발생하는 비용으로서 위의 항목에 포함되지 아니하는 비용

자료: 「산업입지 및 개발에 관한 법률 시행령」 <개정 2008.2.29>

3) 산업단지의 공급면적은 전체 면적에서 도로·녹지 등의 무상으로 공급되는 공공시설용지(평균 40% 수준)를 제외하고 결정된다.

표 2. 산업단지 조성원가 항목별 추이

(천원/3.3m² 또는 %)

항 목	95~99년		00~04년		05~09년		연평균증가율 (97~07년)
	금액	비율	금액	비율	금액	비율	
① 용 지 비	79	29	124	35	296	48	14.1
② 조 성 비	119	43	145	41	204	34	5.5
③ 자본비용	63	23	54	15	84	14	2.9
④ 간 접 비	13	5	33	9	27	4	7.6
총 계	275	100	356	100	611	100	8.3

비고 : 원가과약이 가능한 토지공사 시행 산업단지 대상(19개소)

자료 : 국가경쟁력강화위원회 제3차 회의자료, 2008. 5.

수준에 달한다.

조성원가의 항목별 추이를 살펴보면, 용지비의 비중이 지속적으로 상승하고 있는 반면에 조성비와 간접비 및 자본비용의 비중은 낮아지고 있는 추세이다. 표 2에서 알 수 있듯이 1997~2007년의 기간 중 용지비의 상승률은 14.1%로서 이는 같은 기간 전체 비용의 상승률인 8.3%의 1.7배의 수준에 달하며, 용지비의 비중은 90년대 후반의 29%에서 현재 48%(수도권은 60%)의 수준까지 급증하였다⁴⁾.

2.2.1 현행 가격산정방식의 문제점

앞서 살펴본 바와 같이, 기업의 생산활동을 위한 산업시설용지의 공급가격은 산업단지의 조성에 투입된 사업비인 조성원가에 근거하여 책정된다. 그러나 현행 산업시설용지 공급가격 산정방식은 조성원가를 바탕으로 한 공급가격의 책정에 있어서 개별적인 용지의 입지여건을 고려하지 않은 채 동일한 단위면적당 가격을 단지 내 모든 산업시설용지에 획일적으로 적용하는 불합리성을 내포하고 있다. 즉, 동일한 산업단지 내의 용지일지라도 주변 토지이용상황, 용도지역, 지형적 특성, 접근성 등의 입지특성에 따라서 기업의 생산활동에 대한 상이한 여건이 형성되지만, 현행의 가격산정에 있어서 이와 같은 요인이 전혀 고려되지 않고 있다.

획일적인 현행 산정방식은 산업시설용지의 수요자이자 소비자인 기업의 선택의 폭을 좁히고 기업 간의 불필요한 경쟁을 조장한다⁵⁾. 즉, 서로 다른 입지여건을 가진 산업시설용지에 대하여 다양한 가격조건을 제시하지 못하므로 기업의 선택범위를 제한하며, 또한 기업은 동일한 공급가격하에서 생산활동에 좀 더 바람직하고 상대적으로 높은 시세차익을 얻을 수 있는 산업시설용지로의 입주를 원하기 때문에 산업단지 내 일부 용지에 대한 과도한 경쟁을 유발한다. 더불어 산업시설용지가 공급된 이후에 매매행위가 이루어지는 시점의 거래가격은 개개의 용지에 따라 적지 않은 차이가 나타나고 있는데, 이는 동일한 가격으로 용지가 공급되었지만 현실적인 거래에 있어서는 입지특성이 반영된 가격으로 거래

4) 동일한 기간 동안 전국 공시지가 상승률(8.0% 수준)과 비교하여 용지비는 큰 폭으로 상승하였으며, 이와 같은 용지비의 급등은 산업시설용지 공급가격 상승의 가장 주요한 요인으로 작용하였다.

5) 일반적으로 산업시설용지의 입주대상 선정은 입주신청 마감 후 업종별로 입주우선 순위에 따르며, 우선순위가 경합되는 경우에는 추첨방식에 의한다.

가 형성되는 점에 기인한다고 볼 수 있다.

이와 관련하여 국가경쟁력강화위원회의 제3차 회의(2008)에서는 현재 모든 산업시설용지가 입지여건 등에 관계없이 조성원가 기준에 의한 추첨방식으로 일률적으로 공급됨에 따라 발생하는 용지위치에 따른 가격차이 및 수요자의 선호가 반영되지 않는 불합리한 문제점을 검토한 바 있다. 또한, 그 개선방안으로 전체 산업시설용지의 가격책정은 조성원가 방식에 근거하되 토지이용여건에 따라서 ±10% 범위 내에서 차등화하는 방안이 제시되었다.

따라서 현행 산업시설용지 공급가격 산정방식은 입지여건에 관계없이 동일한 가격을 단지 내 모든 산업시설용지에 획일적으로 적용하는 불합리성을 내포하고 있으므로, 이를 지양하기 위하여 산업시설용지의 입지특성을 고려한 차등적인 공급가격 산정방안을 모색할 필요가 있다.

2.3 SVR의 특징 및 연구동향

Din, Hoesli 및 Bender(2001)가 언급한 것처럼 부동산 가격은 그 형성요인, 즉 내부의 물리적인 특성 및 외부의 환경적인 특성 등과 반드시 선형적인 관계를 나타내는 것은 아니다. 국외의 연구에서는 데이터의 분포에 대한 가정에 무관하고, 비선형 함수관계를 포착하는데 있어 적합한 것으로 알려진 신경망모형(Artificial Neural Networks)이 회귀모형과의 상호보완적 측면에서 부동산 가격예측의 문제에 종종 사용되어왔다. 비교적 최근에 와서는 경험적 위험 최소화(empirical risk minimization)⁶⁾를 구현하는 기존의 학습 알고리즘과는 달리 구조적 위험 최소화(structural risk minimization) 원리에 기반한 SVR이 기존의 신경망 모형이 가지고 있는 학습데이터에 대한 과대적합에 따른 일반화 능력(generalization) 저하와 모형구축을 위한 시행착오적인 절차 등의 문제점을 해결하기 위한 방안으로 검토되고 있다.

Vapnik(1995)이 제안한 SVR을 사용한 국내의 연구들을 살펴보면, 박우열, 김광희(2007)는 SVR을 공동주택 건설프로젝트의 초기단계에서 공사비를 예측할 수 있는 모형구축에 사용하고 그 적용가능성을 검토하였다. 박찬규(2006)는 SVR 모형을 사용하여 소프트웨어 개발비를 예측하였으며, 김유일 등(2004)은 주가지수의 예측에 신경망 및 SVR 모형

6) 경험적 위험 최소화 학습원리는 관찰 및 경험을 통해서 수집할 수 있는 데이터양의 유한성으로 인하여 모형의 일반화가 상대적으로 어렵고, 과대적합의 문제를 야기할 수 있다.

등을 구축하고 모형의 예측성과를 비교분석하였다. 또한 전 성해 등(2004)은 데이터 품질과 관련된 결측치(missing value) 문제를 해결하기 위해서 SVR 모형을 제안하였다.

SVR은 함수근사에 있어 특이값(outlier)에 대한 둔감성, 복잡한 비선형(nonlinear) 함수관계를 포착할 수 있는 점, 벌칙항(penalty)을 이용한 과대적합의 해소, 모형구축의 용이성⁷⁾ 등의 특징을 가진다. 그러나 다른 한편으로 SVR은 예측결과에 대한 해석이 용이하지 않으므로, 모형에 대한 분석 및 해석의 측면에서 우수한 기존의 회귀모형과 상호보완적인 관점에서 접근하는 것이 바람직하다고 판단된다. 본 연구에서는 SVR을 지가산정을 위한 모형구축에 사용하고, 실증적인 지가예측을 통하여 모형의 적정성을 평가하고자 한다.

3. SVR(Support Vector Regression)

3.1 선형 SVR의 기본원리

학습데이터(training data)를 $\{(x_i, y_i), \dots, (x_N, y_N)\}$, $x_i \in R^n$, $y_i \in R$ 라고 가정할 때, x 는 입력벡터, y 는 출력벡터, R^n 은 입력공간, N 은 학습데이터의 크기를 나타낸다. Vapnik이 제안한 가장 단순한 형태의 ϵ -SVR은 모든 학습데이터에 대해서 실제 목표값 y_i 로부터 최고 ϵ (insensitive parameter)만큼의 편차 내에 있으며, 가능한 작은 크기의 w 를 가지는 함수 $f(x)$ 를 찾는 것이다. 즉, 이것은 ϵ 보다 큰 편차를 허용하지 않으면서 ϵ 보다 작은 학습데이터의 오차는 무시하는 것이다. 이와 같은 조건을 만족하는 선형함수 f 는 식 (2)와 같은 형태를 가진다.

$$f(x) = \langle w, x \rangle + bias$$

with $w \in R^n$, $bias \in R$ (2)

가장 작은 크기의 w 를 찾기 위해 다음과 같은 볼록 최적화 문제(convex optimization problem)를 구성할 수 있다.

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 \\ & \text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - bias \leq \epsilon \\ \langle w, x_i \rangle + bias - y_i \leq \epsilon \end{cases} \end{aligned} \quad (3)$$

만약 ϵ -tube 밖에 학습데이터가 존재한다면 식 (3)은 성립하지 않는다. 이를 해결하기 위해서 어느 정도의 오차를 허용하는 슬랙변수(slack variable) ξ_i, ξ_i^* 을 도입하여 ϵ -tube 밖에 학습데이터가 존재하여도 볼록 최적화의 문제가 성립하도록 할 필요가 있다. 따라서 볼록 최적화의 문제를 다시 구성하면 다음의 식 (4)와 같다.

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ & \text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - bias \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + bias - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (4)$$

7) 신경망 모형과 비교하여 SVR은 학습에 필요한 모수(parameter)의 일부가 자동적으로 결정됨으로써 상대적으로 모형구축이 용이하다고 할 수 있다.

여기서 0보다 큰 값을 가지는 상수 C 는 함수의 ϵ -tube 밖에 존재하는 학습데이터에 대한 오차의 허용범위를 결정한다. 즉, 상수 C 는 추정오차의 패널티(penalty)로서 모형의 복잡도(model complexity)를 결정하는 모수이다. C 값이 크면 오차에 대하여 큰 패널티를 할당하여 회귀가 낮은 일반화 수준으로 오차를 최소화시키는 반면, C 값이 작으면 오차에 대하여 작은 패널티를 할당하여 높은 일반화 수준을 갖게 된다.

앞서 서술한 볼록 최적화 문제에서는 ϵ -무감각(ϵ -insensitive) 손실함수가 사용되었다⁸⁾. 손실함수는 예측값과 실제값 사이의 오차가 있을 경우에 이를 구하고자 하는 함수식에 반영시키는 방법을 제시하는 함수를 의미한다(Cortes, Vapnik, 1995). Vapnik의 ϵ -무감각 손실함수는 식 (5)와 같이 나타낼 수 있다.

$$|\xi|_{\epsilon} := \begin{cases} 0 & \text{if } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{otherwise} \end{cases} \quad (5)$$

식 (5)를 표현한 그림 2에서 ϵ -tube 내에 있는 데이터는 무시되고, ϵ -tube 밖에 존재하는 데이터의 오차는 슬랙변수인 ξ_i, ξ_i^* 로 측정된다. 또한, 식 (4)의 최적화 문제에서 라그랑지 승수(lagrange multiplier)를 도입하면 목적함수(objective function)와 제약조건(constraint)으로 구성되는 식 (6)과 같은 라그랑지 함수로 나타낼 수 있다.

$$\begin{aligned} L = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ & - \sum_{i=1}^N a_i (\epsilon + \xi_i - y_i + \langle w, x_i \rangle + bias) \\ & - \sum_{i=1}^N a_i^* (\epsilon + \xi_i^* + y_i - \langle w, x_i \rangle - bias) \\ & - \sum_{i=1}^N (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (6)$$

식 (6)을 쌍대 최적화 문제(dual optimization problem)로 도출한 후, 해를 구하면 다음과 같은 SVR 추정함수를 구할

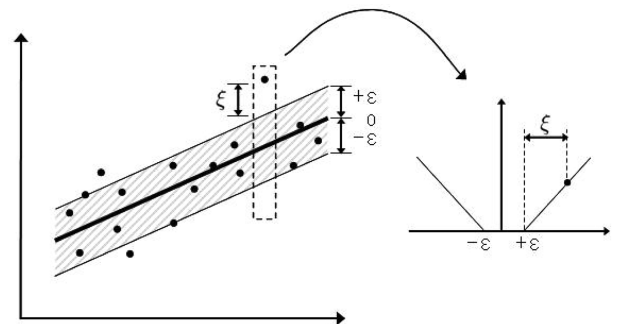


그림 2. ϵ -무감각(ϵ -insensitive) 손실함수

8) 대표적인 손실함수로서 2차다항, Laplacian, robust, ϵ -무감각 손실함수가 있다. 그러나 ϵ -무감각 손실함수를 제외한 나머지 함수들은 희소 데이터(sparse data)에 대해서는 적합하지 않은 것으로 알려져 있다.

수 있다.

$$w = \sum_{i=1}^N (a_i - a_i^*) x_i \quad (7)$$

$$f(x) = \sum_{i=1}^N (a_i - a_i^*) \langle x_i, x \rangle + bias \quad (8)$$

3.2 SVR의 비선형 확장

학습데이터의 선형 예측이 불가능한 경우에 입력공간의 학습데이터를 비선형 사상함수 Φ 를 이용하여 고차원의 특성값 공간(feature space)으로 사상($\Phi: X \rightarrow F$)시켜 선형 함수근사하고, 다시 입력공간으로 비선형 사상을 하게 되면 비선형 함수근사를 할 수 있다. 즉, 선형 SVR 알고리즘은 커널함수(kernel function)⁹⁾라고 불리는 비선형 사상함수를 이용하여 비선형 확장이 가능하다. 그림 3은 커널함수를 이용한 고차원의 특성값 공간으로의 사상과 비선형 함수근사의 관계를 보여준다.

커널함수를 이용하여 비선형 함수근사를 하게 되면 식 (8)의 회귀함수는 식 (9)와 같이 표현될 수 있다.

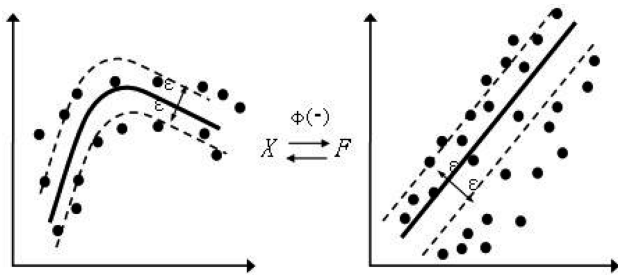


그림 3. 커널함수를 이용한 특성값 공간으로의 사상

$$f(x) = \sum_{i=1}^N (a_i - a_i^*) k(x_i, x) + bias \quad (9)$$

4. 입지특성을 고려한 산업시설용지 공급가격의 차등적 산정

4.1 데이터의 구축 및 분할

4.1.1 가격 및 입지특성 데이터의 구축

본 연구에서는 산업시설용지 공급가격의 차등적 산정모형을 구축하기 위하여, 그림 4에 제시된 명지녹산국가산업단지, 창원국가산업단지, 신평·장림일반산업단지, 양산일반산업단지, 양산 어곡일반산업단지, 진주상평일반산업단지, 김해 덕암일반산업단지, 사상공업단지 내 297필지의 산업용지에 대한 가격데이터와 입지특성데이터로 구성된 데이터를 사용하였다. 이 가운데 부산광역시 사상구 낙동강 하류부 연안 일대의 공업지역을 지칭하는 사상공업단지는 도심에 산재한 용도



그림 4. 데이터의 공간적 범위

표 3. 대상지의 현황

구분	명지녹산	창원	신평·장림	양산	양산어곡	진주상평	김해덕암	사상공단
유형	국가산단	국가산단	일반산단	일반산단	일반산단	일반산단	일반산단	공업지역
지형	해면 및 해안 매립 평지	소구릉 분포된 분지	임야, 농경지 및 하구언 매립지	평탄지	동저서고형 구릉지	평탄지	임야 및 농경지	저습지 매립지
총면적 (천m ²)	6,972	25,302	2,602	1,529	1,270	2,135	156	8,543
산업시설 용지면적 (천m ²)	4,317	17,245	1,761	1,271	615	1,514	106	.
조성기간	1990.1 ~2003.2	1974.2 ~2008.12	1981.1 ~1990.6	1978.3 ~1981.9	1992.12 ~2003.12	1977.10 ~2001.12	1999.8 ~2001.5	1968.11 ~1975.3
사업체수 (개사)	1,339	1,633	613	96	122	500	25	2,847
종사자수(명)	25,968	80,080	16,699	8,726	3,795	8,769	518	36,281
공급가격 (천원/m ²)	185	173 (2001년)	61	150	197	151 (1992년)	151	.

자료: 한국산업단지공단, 사상구청, 한국토지정보시스템(KLIS), 2008. 9.

9) 비선형 함수근사를 가능하게 하는 커널함수로는 다항(Polynomial) kernel, Radial Basis kernel, Hyperbolic Tangent kernel 등이 일반적으로 사용된다.

지역 위배공장 및 공해공장을 이전·집단화시킬 목적으로 개발되었으며, 그 조성방식에 있어 여타의 산업단지와 크게 다르지 않다. 가격데이터의 시간적 범위는 2004년도를 기준으로 하였으며, 신뢰할 수 있는 가격데이터를 구하기 위해서 산업용지의 실거래가격으로 간주할 수 있는 법인체 간의 거래가격 및 법인체와 개인 간의 거래가격만을 사용하였다. 또한, 토지 및 건물이 결합되어 구성된 복합부동산의 경우에는 거래가격에서 건물에 대한 매매시점의 복성가격을 차감하거나, 토지가격 구성비를 알 수 있는 경우에는 그 구성비를 승하여 토지가격만을 추출하였다¹⁰⁾.

산업시설용지의 가격형성에 영향을 미치는 수많은 요인을 전부 고려하기에는 현실적인 한계가 존재하므로, 산업입지개발 정책수립 및 산업단지 조성·관리 등의 실무를 담당하여 산업입지와 관련된 전문성을 확보한 산업단지 전문가를 대상으로 하는 델파이 조사를 시행하여 산업시설용지의 가격형성과 유의미한 관련성을 가지는 입지특성요인을 선정하였다. 따라서 본 연구에서는 건설교통부 산업입지팀, 산업자원부 지역투자입지 담당관실, 한국산업단지공단 산업단지처 개발팀, 한국토지공사, 부산진해경제자유구역청 등의 산업단지 담당자 24명을 패널(panel)로 선정하여, 2005년 7월 4일부터 9월 30일까지의 기간에 델파이 조사(delphi method)를 시행하였다. 패널의 수에 대해서는 적게는 10명 미만도 가능하며, 그 수가 적을수록 효과적이라는 연구와 패널의 수가 커질수록 신뢰도가 커진다는 연구로 의견이 나뉘지만, 델파이 기법을 제안한 Dalkey의 연구에 따르면 그룹의 크기가 15명 정도이면 그룹 간의 중위수에 차이가 나지 않는다.

델파이 조사는 패널들에게 선행연구를 참고하여 입지특성요인의 항목을 예측하도록 요구하는 한 차례의 개방형 면담 조사와 두 차례의 구조화된 설문(델파이 폐쇄형 문항)조사로

구성되었다. 또한, 델파이 반응척도로는 리커트 5점 척도가 사용되었고, 조사의 중단시점을 의미하는 안정도(stability)의 평가기준으로는 변이계수(coefficient of variation, C.V.)가 사용되었다.

1·2차 라운드의 결과로서 토지형상, 지목, 도로접면, 주변 토지이용상황, 간선도로 접근성, 단지면적, 물류여건, 세제혜택, 산업단지 준공연수, 폐기물 처리용량 등 21개의 요인이 도출되었으며, 이어진 3차 라운드를 통해서 12개의 입지특성요인이 최종적으로 선정되었다. 여기서 변이계수의 값은 모두 0.5 이하로 나타나 전문가들의 합의가 이루어진 것으로 볼 수 있었으며¹¹⁾, 전문가들은 산업용지 수요, 제조업 부가가치, 제조업 종사자수, 고속도로 접근성, 주변 토지이용상황 등이 산업시설용지의 가격형성에 중요한 영향을 미칠 것으로 예상하였다. 따라서 산업시설용지 공급가격 선정모형을 구축하기 위한 변수는 표 4와 같으며, 선정된 입지특성요인은 그 속성에 따라서 지역적 요인, 산업단지 관련요인, 필지 자체의 요인으로 분류될 수 있다. 이 중에서 지역적 요인에 속하는 산업용지 수요에 대한 자료값으로는 건설교통부의 「제2차 산업입지 공급계획(2002-2011)」에 제시된 계획기간의 산업용지 수요 추정치(순수요+추가수요)를 사용하였으며, 제조업 종사자수 및 부가가치액은 각 지자체별 통계연보상의 자료를 활용하였다. 또한 필지 자체의 요인 중 간선도로 거리는 도시계획시설의 결정·구조 및 설치기준에 관한 규칙상의 기능별 구분에 따른 도시관리계획상의 간선도로를 기준으로 하여 개별필지로부터 가장 근접한 간선도로까지의 최단 네트워크 거리를 산정하여 분석에 사용하였다. 이들 각 변수값의 분포범위는 표 5의 기초통계량에 제시된 바와 같다.

표 4. 변수의 구성

구분		단위	변수설명	
출력변수	지가	원/m ²	필지가격	
	지역			
입력변수	용지수요	km ²	해당지역의 산업용지 수요	
	제조업 종사자수	인	해당지역의 제조업 종사자 수	
	제조업부가가치	백만원	해당지역 내 제조업체의 평균 부가가치액	
	산업단지	단지면적	천m ²	산업단지의 면적
		고속도로거리	km	산업단지에서 고속도로 진입로까지의 최단 네트워크 거리
		철도거리	km	산업단지에서 철도역까지의 최단 네트워크 거리
		항만거리	km	산업단지에서 항만시설까지의 최단 네트워크 거리
	물류시설거리	km	산업단지에서 물류시설까지의 최단 네트워크 거리	
	필지자체	주변토지이용	dummy	산업시설용지, 지원시설용지 및 기타 용지
		용도지역	dummy	전용공업지역, 일반공업지역, 준공업지역
고도		m	개별필지의 해발고도	
간선도로거리		m	개별필지에서 간선도로까지의 최단 네트워크 거리	

10) 배분법(Allocation Method)
 : {총매매가격-건물가격}*사정보정*시점수정*지역요인
 *개별요인*면적비교
 비율법(Extraction Method)
 : {총매매가격*토지가격 / (토지가격+건물가격)}*사정보정
 *시점수정*지역요인*개별요인*면적비교

11) 변이계수(C.V.)를 이용한 델파이 라운드의 안정도 평가에 있어 $0 < C.V. \leq 0.5$ 일 경우에는 높은 수준의 합의(good degree of consensus)에 도달한 것으로 간주한다.

표 5. 데이터의 기초통계량

구 분	평균값	표준편차	최소	최대
지가(원/m ²)	520,252.53	259,713.55	110,000.00	1,140,000.00
용지수요(km ²)	6.47	3.69	3.75	11.45
종사자수(인)	129,988.00	60,250.25	9,895.00	171,156.00
부가가치(백만원)	1,749.79	1,475.54	954.83	4,943.24
단지면적(천 m ²)	9,632.03	7,734.14	1,268.00	25,506.00
고속도로거리(km)	6,166.45	4,552.57	1,184.00	14,310.00
철도거리(km)	9,221.00	4,531.35	3,730.00	19,807.00
항만거리(km)	17,324.74	15,212.34	7,001.00	53,610.00
물류시설거리(km)	4,363.65	2,667.27	1,200.00	12,806.00
고도(m)	16.31	21.06	2.00	138.00
간선도로거리(m)	115.32	113.69	0.00	670.00
구 분	빈도	%	누적 %	
용도지역	1: 전용공업지역	50	16.8	16.8
	2: 일반공업지역	125	42.1	58.9
	3: 준공업지역	122	41.1	100.0
	계	297	100	
주변토지 이용	1: 산업시설용지	184	62.0	62.0
	2: 지원시설용지 및 기타용지	113	38.0	100.0
	계	297	100	

4.1.2 데이터의 분할 및 변환

본 연구에서는 구축된 데이터를 학습용(training data), 검증용(test data)으로 분할하였다. 보다 구체적으로, 가격데이터와 입지특성데이터로 구성된 전체의 데이터에서 임의 추출된 약 80%(237필지)의 데이터를 모형구축을 위한 학습용 데이터로, 학습에 사용되지 않은 약 20%(60필지)의 데이터를 선정된 모형의 적정성을 평가하기 위한 검증용 데이터로 분할하였다. 더불어 모형구축의 효율성과 모형의 예측력을 제고하기 위하여 그 특성과 크기가 서로 다른 입지특성데이터의 값(입력변수 값)을 0~1의 범위로 정규화(normalization)하는 데이터의 변환과정을 수행하였다.

표 6. 데이터의 분할

데이터	개수	구분
학습용(training)	237필지(80%)	모형의 구축
검증용(test)	60필지(20%)	모형의 검증

4.2 SVR 모형을 이용한 지가예측

4.2.1 SVR 모형의 학습조건 설정

SVR 모형을 구축하기 위해서는 커널함수(kernel function)와 모수(parameter) 등의 조건이 우선 설정되어야 한다. 즉, SVR 모형의 학습을 위해 표 7에 명시된 커널함수와 추정오차의 패널티에 해당하는 상수 C 및 연산종료를 위한 ε를 포함하여 각 커널함수에 따른 모수들의 범위가 설정되어야 한다. 본 연구에서는 SVR 모형에 일반적으로 사용되는 다양한 커널함수를 고려하였으며, 또한 모수들의 설정범위가 광범위한 경우에 많은 연산시간이 소요되므로 기존의 선행 연구를 참고하여 C는 0~200, ε는 0~1 등으로 설정하여 예비적인 학습을 수행하였다¹²⁾.

4.2.2 SVR 모형의 학습 및 모형선정

예비적인 학습의 결과, 커널함수 중에서는 다항 커널함수가 가장 적합한 것으로 나타났다. 다항함수를 사용하는 비선

표 7. 커널함수와 기본적 모수

구 분	설 명	
커널함수 (kernel functions)	Linear kernel	$k(x,y) = x^T y + c$
	Polynomial kernel	$k(x,y) = (ax^T y + c)^d$
	Radial Basis kernel	$k(x,y) = \exp(-\gamma \ x-y\ ^2)$
	Hyperbolic Tangent kernel	$k(x,y) = \tanh(ax^T y + c)$
모수 (parameters)	C	학습오차와 일반화 수준 사이의 trade-off를 조절하는 모수
	ε (epsilon)	연산종료를 위한 오차의 설정

12) 모형의 학습을 위하여 SVMdark를 이용하였으며, 다항함수의 모수 d, s, c는 0~5, Radial Basis함수의 모수 γ는 -5~5, Hyperbolic Tangent 함수의 모수 s, c는 -5~5의 범위로 설정하였다.

형의 SVR 모형은 모형의 복잡도를 결정하는 C, 연산을 종료하기 위한 오차의 설정값 ϵ , 다항함수의 차수(polynomial degree) d, 기울기(slope)를 나타내는 s 또는 α , 상수항(constant term) c에 따라서 다양한 모형이 구축된다. 따라서 앞서 수행한 예비적인 학습결과를 바탕으로 각 설정범위를 조정하고 반복적인 학습을 통한 모형의 최적화를 수행하였다.

일반적으로 모형의 적합성 평가에 사용되는 오차는 다양한 척도에 의해서 산출될 수 있는데, 본 연구에서는 평균제곱근 오차(root mean square error, RMSE)를 기준으로 하였다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (D_i - A_i)^2} \quad (10)$$

(D_i : 목표값, A_i : 예측값)

표 8에서는 설정조건의 변화에 따른 모형별 오차(RMSE)가 제시되었다. 이를 살펴보면 $C = 148.4863$, $\epsilon = 0.8897$, $d = 5$, $s = 4.8027$, $c = 3.2379$ 의 조건을 갖는 모형의 오차가 71,349.43으로 최소인 것을 알 수 있다. 따라서 본 연구에서는 이와 같은 조건으로 구성된 SVR 모형을 실증적인 지가 예측을 위한 최종의 모형으로 선정하였다.

4.3 SVR 모형에 의한 지가예측 및 모형평가

4.3.1 SVR 모형에 의한 지가예측

선정된 SVR 모형을 이용하여 검증용 데이터(60필지)의 가격을 예측하였으며, 실제가격 및 모형에 의한 예측가격, 두 가격 간의 차이의 분포를 도식화하면 그림 5와 같다. 여기

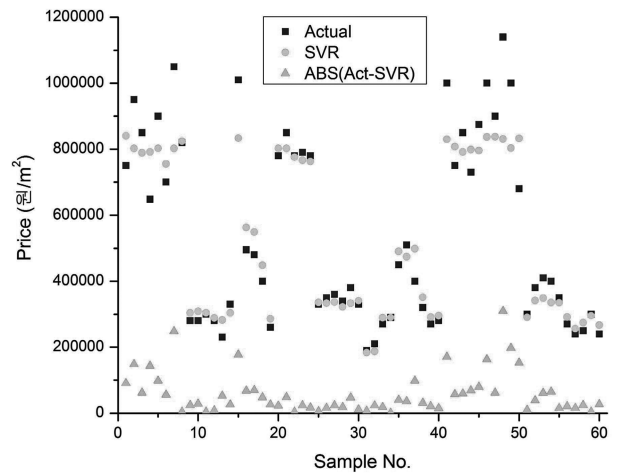


그림 5. 실제가격 및 예측가격(SVR)의 분포

서 예측가격은 최소 183,446원/m²에서 최대 840,404원/m²의 범위를 갖는 것으로 나타나, 최소 190,000원/m², 최대 1,140,000원/m²인 실제가격과 비교적 유사한 분포를 보였다. 보다 구체적으로, 실제가격과 예측가격 간의 오차율(absolute percentage error)¹³⁾은 표 9에 제시된 바와 같으며, 이들 오차율의 평균(mean absolute percentage error, MAPE)은 9.67%로 나타났다. 또한 두 가격 간의 차이(실제가격-예측가격)의 평균은 16,774원/m², 최대값 및 최소값은 각각 309,182원/m², -152,862원/m²으로 나타났다.

표 8. SVR 모형의 학습결과

구분	C	ϵ	kernel	d	s	c	RMSE
1	15.8879	0.8846	poly	5	4.8788	2.8970	72,197.83
2	19.9255	0.9467	poly	5	3.9486	3.7205	72,599.44
3	30.1187	0.0953	poly	5	4.9873	3.8235	72,215.68
4	36.0454	0.8303	poly	5	3.4307	4.0362	72,476.08
5	48.1521	0.1437	poly	5	4.3704	1.0042	71,989.65
6	49.5865	0.5844	poly	5	3.6996	2.9789	72,418.86
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
79	128.7286	0.2557	poly	5	2.2137	3.1075	73,615.45
80	130.3201	0.1174	poly	5	2.9644	0.2841	73,252.37
81	145.9089	0.1861	poly	5	3.1907	1.6923	72,126.81
82	148.4863	0.8897	poly	5	4.8027	3.2379	71,349.43
83	149.1958	0.2086	poly	5	4.7040	3.1019	71,780.79
84	149.4873	0.7118	poly	5	4.4188	2.0714	71,805.54
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
95	167.0278	0.8856	poly	4	4.5926	3.8844	73,636.22
96	167.4581	0.5616	poly	5	2.8939	4.2773	72,916.26
97	171.8894	0.1761	poly	5	3.6184	0.8719	72,259.53
98	184.2250	0.9373	poly	5	4.2393	1.8427	71,785.38
99	187.5790	0.7074	poly	5	3.9047	2.1522	72,016.41
100	190.6095	0.8320	poly	5	2.7836	4.3265	72,978.46

13) 오차율(%) = $\frac{|D_i - A_i|}{D_i} \times 100$ (D_i : 목표값(실제가격), A_i : 예측값)

표 9. 실제가격과 예측가격 간의 오차율(%)

오차범위(%)	빈도	%
0-5	15	25.00
5-10	23	38.33
10-15	10	16.67
15-20	6	10.00
20-25	5	8.33
25-30	1	1.67
계	60	100
평균(MAPE)	9.67%	

4.3.2 예측결과의 평가

SVR 모형의 적정성을 평가하기 위해서 실제가격과 예측가격 간의 통계적 유의성을 살펴볼 필요가 있으므로 대응표본 t검정(paired t-test)을 시행하였다. 그 결과, 검정통계량 값은 1.524이고 이에 대한 유의확률(양쪽)은 0.133으로서 유의수준인 0.05보다 크므로, 실제가격과 SVR 모형에 의한 예측가격 간에는 통계적으로 유의한 차이가 있다고 할 수 없었다(표 10). 따라서 SVR 모형은 개별적인 토지의 상이한 입지특성이 반영된 실제가격을 근사한 수준에서 예측할 수 있는 것으로 판단하였다.

표 10. 실제가격과 SVR 모형에 의한 예측가격의 비교(paired t-test)

구분	평균	t	자유도	유의확률
실제가격-예측가격	16,773.78	1.524	59	0.133

4.3.3 기존 연구와의 비교

앞서 실제가격과의 통계적 유의성을 살펴보았던 SVR 모형의 예측결과를 동일한 데이터를 적용한 다른 방법론과 비교할 경우, SVR 모형의 적정성을 다른 각도에서 판단할 수 있을 것으로 판단된다. 따라서 이를 기존의 연구(심재현 외, 2007)에서 활용한 신경망 모형과 비교함으로써 SVR 모형의 적정성을 검토하였다. SVR 모형의 구축과정에서 사용한 동일한 학습데이터(237필지)를 이용하여 신경망 모형을 구축·선정하고 검증용 데이터(60필지)의 가격을 예측한 결과, 두

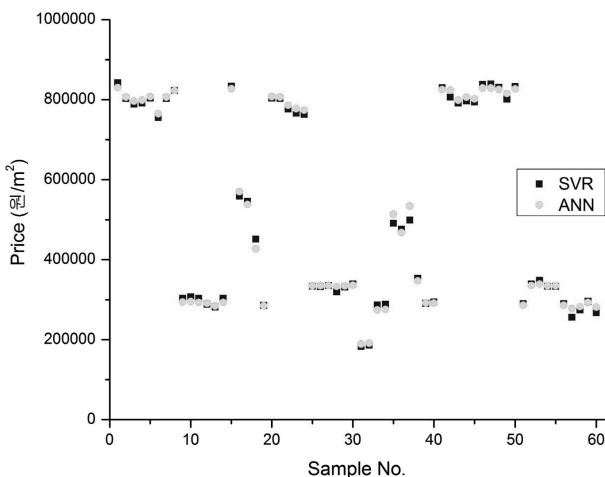


그림 6. SVR 및 신경망 모형의 예측가격 분포

표 11. 검증용 데이터에 대한 SVR 및 신경망 모형의 오차(RMSE) 비교

구분	RMSE	설정조건
SVR	86,213.58	$C = 148.4863, \epsilon = 0.8897, d = 5, s = 4.8027, c = 3.2379,$ polynomial kernel
신경망	87,468.18	$H=1, \eta=15, \alpha=0.9, \beta=0.7^{14}),$ back-propagation learning, sigmoid function

표 12. SVR 및 신경망 모형의 예측가격 비교(paired t-test)

구분	평균	t	자유도	유의확률
SVR-신경망	-981.97	-0.760	59	0.451

모형에 의한 예측가격의 분포는 그림 6에 나타난 바와 같다. 여기서 신경망 모형에 의한 예측가격은 최소 188,468원/m²에서 최대 830,195원/m²의 범위를 보여 SVR 모형에 의한 예측가격과 유사한 분포를 나타냈다.

SVR 및 신경망 모형 간의 예측력을 비교하기 위해서 검증용 데이터에 대한 두 모형의 오차(RMSE)를 확인하였다. 표 11에 제시된 바와 같이, 두 모형의 오차는 각각 86,213.58, 87,468.18으로 SVR 모형의 오차가 신경망 모형에 비해서 다소 낮았으나 그 차이는 크지 않은 것으로 나타났다. 따라서 두 모형에 의한 예측가격 사이에 통계적으로 유의한 차이가 있는지를 검토하기 위하여 대응표본 t검정(paired t-test)을 시행하였다. 검정의 결과, 두 모형에 의한 예측가격 간의 차이(SVR-신경망)의 평균은 -981.97원/m², 검정통계량 값은 -0.760이며 유의확률(양쪽)은 0.451로서 두 모형에 의한 예측가격 사이에는 통계적으로 유의한 차이가 있다고 볼 수 없었다. 그러므로 어떠한 모형의 예측력이 상대적으로 우수하다고 단정할 수 없었다.

더불어 모형의 구축과정과 관련하여, 신경망 모형은 은닉층의 수, 은닉층의 노드수, 학습률(learning rate), 모멘텀(momentum) 등의 여러 가지 학습조건을 적절한 조합을 찾기 위한 시행착오적인 절차로 인하여 연구자의 기교적인 요소 및 적지 않은 시간이 요구되었다. 이와 비교하여, SVR 모형은 학습에 필요한 모수의 일부가 자동적으로 결정됨으로써 상대적으로 그 과정이 용이하고 적은 시간이 소요되었다.

4.4 예측가격의 변환

앞서 언급한 바와 같이 본 연구에서는 토지의 시장거래가격은 보편적으로 입지특성에 근거하여 형성된다는 점을 전제하여 가격 및 입지특성 데이터를 사용하여 지가를 예측하므로, 예측가격은 개별적인 필지의 상이한 입지특성에 기반한 실제가격과 유사한 분포를 보인다. 그러나 산업시설용지의 특수성과 공공성을 고려하고, 기반시설이 구비된 산업시설용지를 저렴하게 공급하려는 취지를 담은 현행 법률을 감

14) H 는 은닉층(hidden layer)의 수, η 은 은닉층의 노드 수, α 는 신경망의 학습과정에서 연결가중치의 변화율을 결정하는 학습률(learning rate), β 는 신경망의 연결가중치 조절식에 관성을 줌으로써 학습성능을 향상시키는 모멘텀(momentum)을 의미한다.

안하여 모형의 실제적인 적용에 있어서는 식 (10)에 제시된 예측가격의 변환과정¹⁵⁾을 수행할 필요가 있다. 즉, 이와 같은 예측가격의 변환을 통하여 산업시설용지의 공급가격 총액은 조성원가로 구성되지만, 단지 내 각각의 산업시설용지는 개별적인 입지특성이 반영된 차등적인 가격으로 책정될 수 있다. 이로써 본 연구의 입지특성을 고려한 산업시설용지 공급가격의 차등적 산정모형이 실무적으로 적용 가능함을 알 수 있다.

$$\text{변환가격} = \text{예측가격} \times \frac{\text{조성원가}}{\text{예측가격 총액}} \quad (10)$$

subject to (예측가격 총액 > 조성원가)

5. 결 론

현행 산업시설용지 공급가격 산정방식은 입지여건에 관계 없이 동일한 가격을 단지 내 모든 산업시설용지에 획일적으로 적용함으로써, 용지위치에 따른 가격차이를 반영하지 못하며 수요자인 기업의 선택범위를 제한하는 불합리한 문제점을 가지고 있다. 국가경쟁력강화위원회 제3차 회의(2008)에서는 이와 관련한 문제점을 검토하여 산업시설용지의 가격책정은 조성원가 방식에 근거하되 토지이용여건에 따라서 차등화하는 기본적인 개선방향을 제시한 바 있다. 따라서 본 연구는 현행 산정방식의 불합리성을 지양하기 위한 방안으로 입지특성에 근거한 산업시설용지 공급가격의 차등적 산정모형을 제시하고자 하였다. 본 연구의 수행과정 및 결과를 정리하면 다음과 같다.

지가 산정모형을 구축하기 위하여 총 297필지의 산업용지에 대한 가격 및 입지특성 데이터로 구성된 데이터를 구축하였다. 이어서 구축된 데이터를 학습용, 검증용으로 분할하고, 그 특성과 크기가 서로 다른 입지특성데이터의 값을 정규화하는 데이터의 변환과정을 수행하였다. 구축된 데이터를 바탕으로 실증적인 자가예측을 위한 SVR 모형을 구축하기 위하여 학습을 통한 모형의 최적화 과정을 시행하였다.

선정된 모형을 이용한 자가예측의 결과를 살펴보면, 모형에 의한 예측가격은 실제가격과 유사한 분포를 보였으며, 모형의 적정성을 평가하기 위해서 실제가격과 예측가격 간의 통계적 유의성을 검증한 결과, 두 가격 사이에는 통계적으로 유의한 차이가 있다고 판단할 수 없었다. 따라서 본 연구에서 제시한 산업시설용지 공급가격의 차등적 산정모형은 개별적인 토지의 상이한 입지특성에 근거하여 형성되는 실제가격을 근사한 수준에서 예측할 수 있는 것으로 판단하였다. 또한 동일한 데이터를 적용한 신경망 모형과 예측결과를 비교한 결과, SVR 모형의 오차가 신경망에 비하여 다소 낮게 나타났으나, 두 모형에 의한 예측가격 사이에는 통계적으로 유의한 차이가 있다고 볼 수 없었다. 더불어 산업시설용지의 특수성과 공공성을 고려하고, 기반시설이 구비된 산업시설용지를 저렴하게 공급하려는 취지를 담은 현행 법률을 감안하

여 모형에 의한 예측가격의 이론적 변환과정이 제시되었다.

방법론적인 측면에서, 경험적 위험 최소화 원리를 바탕으로 하는 SVR은 벌칙항(penalty)을 이용한 과대적합의 해소, 함수근사에 있어서 특이값에 대한 둔감성 및 복잡한 비선형 함수관계를 포착할 수 있는 점 등의 특성을 가지며, 신경망 모형과 비교하여 학습에 필요한 모수의 일부가 자동적으로 결정됨으로써 모형구축이 상대적으로 용이하다고 볼 수 있다. 그러나 SVR은 예측결과에 대한 해석이 용이하지 않다는 단점을 가지므로, 모형해석의 측면에서 우수한 전통적인 통계분석모형과 상호보완적인 관점에서 접근하는 것이 바람직하다고 판단된다.

끝으로 본 연구의 한계로서 산업시설용지의 초기 공급가격을 실제가격 및 예측가격과 더불어 검증하고, 또한 예측가격의 변환에 따른 가격의 차이를 실증적으로 검증하는 추가적인 분석이 필요하나, 대상지에 속하는 각 필지들이 동일한 산업단지에 속할지라도 다양한 시기에 서로 다른 가격으로 공급되거나 혹은 일부 산업단지의 경우 조성이 완료된 이후 상당한 시간이 경과되어 초기 공급가격 자료를 구득할 수 없어 이를 수행하지 못한 점이 노정된다. 그럼에도 불구하고, 본 연구는 향후 입지여건을 고려하여 차등화된 산업시설용지 공급가격을 산정함에 있어 출발점이 될 수 있다는 점에서 그 의의를 찾을 수 있다.

참고문헌

- 국토해양부(2008) 개별공시지가 조사·산정지침.
 국토해양부, 한국감정평가협회(2008) 표준지 공시지가 조사·평가 업무요령.
 기획재정부, 지식경제부, 환경부, 국토해양부, 국가경쟁력강화위원회(2008) 산업단지 공급가격 인하방안, 국가경쟁력강화위원회 제3차 회의자료.
 한국산업단지공단 산업입지연구센터(2009) 한국산업단지총람.
 강영호(1997) 우리나라 암 연구수준의 평가와 암 연구인력 훈련 요구의 예측 : 델파이법의 적용, 석사학위논문, 서울대학교.
 김유일, 신은경, 홍태호(2004) 신경망과 SVM을 이용한 주가지수 예측의 비교, 인터넷전자상거래연구, 한국인터넷전자상거래학회, 제4권 제3호, pp. 221-243.
 박우열, 김광희(2007) 서포트 벡터 회귀분석을 이용한 공동주택 공사비 예측에 관한 연구, 대한건축학회논문집(구조계), 대한건축학회, 제23권 제4호, pp. 165-172.
 박우열, 김재엽(2006) Support Vector Machine을 이용한 흙막이공법 선정모델에 관한 연구, 한국건설관리학회논문집, 한국건설관리학회, 제7권 제2호, pp. 118-126.
 박찬규(2006) Support Vector Regression을 이용한 소프트웨어 개발비 예측, 경영과학, 한국경영과학회, pp. 75-91.
 심재현, 조찬호, 이성호(2007) 인공신경망을 이용한 산업용지의 공급가격 산정, 국토계획, 대한국토·도시계획학회, 제42권 제5호, pp. 223-232.
 전성해, 박정은, 오경환(2004) Support Vector Regression을 이용한 회소 데이터의 전처리, 퍼지 및 지능시스템학회 논문지, 한국지능시스템학회, 제14권 제2호, pp. 789-792.
 Ripley, B. D. (1996) *Pattern Recognition and Neural Networks*, Cambridge, Cambridge University Press.
 Smola, A. J. and Schölkopf, B. (1998) *A tutorial on support vector regression*, Royal Holloway College, London, U.K., Neuro COLT Tech. Rep. TR-1998-030.
 Vapnik, V. (1995) *The Nature of Statistical Learning Theory*, Springer.
 Asabere, P. K. and Huffman, F. E. (1991) *Zoning and Industrial*

15) 이는 산업시설용지의 가격이 공급이 완료된 이후 초기 공급가격보다 상승하는 것을 전제하고 있으며, 여기서 예측가격 총액은 하나의 산업단지 내 입지한 모든 산업시설용지에 대한 예측가격의 합계를 의미한다.

- Land Values: The Case of Philadelphia, *Real Estate Economics*, Vol. 19, No. 2, pp. 154-160.
- Cortes, C. and Vapnik, V. (1995) Support Vector Networks, *Machine Learning*, 20, pp. 273-297.
- Ellis, C. and Wilson, J. P. (2005) Can a Neural Network Property Portfolio Selection Process Outperform the Property Market?, *Journal of Real Estate Portfolio Management*, Vol. 11, No. 2, pp. 105-121.
- Guntermann, K. L. (1995) Sanitary Landfills, Stigma and Industrial Land Values, *Journal of Real Estate Research*, Vol. 10, No. 5, pp. 531-542.
- Jackson, T. O. (2002) Environmental Contamination and Industrial Real Estate Prices, *Journal of Real Estate Research*, Vol. 23, No. 1/2, pp. 179-199.
- Kowalski, J. G. and Paraskevopoulos, C. C. (1990) The Impact of Location on Urban Industrial Land Prices, *Journal of Urban Economics*, Vol. 27, No. 1, pp. 17-24.
- Lockwood, L. J. and Rutherford, R. C. (1996) Determinants of Industrial Property Value, *Real Estate Economics*, Vol. 24, No. 2, pp. 257-272.
- Saz-Salazar, S. D. and García-Menéndez, L. (2005) Public provision versus private provision of industrial land: a hedonic approach, *Land Use Policy*, Vol. 22, No. 3, pp. 215-223.
- Weber, Rachel., Bhatta, S. D., and Merriman, D. (2003) Does Tax Increment Financing Raise Urban Industrial Property Values?, *Urban Studies*, Vol. 40, No. 10, pp. 2001-2021.

(접수일: 2010.5.10/심사일: 2010.6.14/심사완료일: 2011.1.3)