

SOM과 LSTM을 활용한 지역기반의 부동산 가격 예측*

신은경** · 김은미*** · 홍태호****

〈 목 차 〉

I. 서론	IV. 실험 및 결과
II. 이론적 배경	4.1 데이터
2.1 부동산의 지역성	4.2 HP 필터
2.2 지역분석을 위한 클러스터링	4.3 군집분석 결과
2.3 부동산 가격 예측 연구	4.4 지역기반 예측모형 결과
2.4 시계열 예측을 위한 딥러닝 기법	V. 결론 및 향후 연구과제
III. 연구 프레임워크	참고문헌
	<Abstract>

I. 서론

주택은 가구의 총자산에서 많은 부분을 차지하는 중요한 자산 중 하나이며, 주택의 가격변동은 중대한 관심사 중 하나이다. 최근에는 코로나19의 장기화로 재택근무와 원격수업이 확대됨에 따라 주택에 대한 수요가 급증하고 있으며 주택가격 또한 계속해서 최고가를 갱신하며 전 세계적으로 급등세를 보이고 있다. 주택가격은 경제분야의 지표 역할을 한다고 볼 수 있으며(Leamer, 2007), 주택가격 조정은 경기

주기의 결정에도 중요한 역할을 하고 있다(Vargas-Silva, 2008). 경제가 호황을 누리게 되면 주택분야의 건설과 고용이 과잉 수요에 대응하기 위해 빠르게 확장되고 주택가격은 상향 조정 된다. 이렇듯 부동산 시장은 경제발전과도 밀접한 관계를 가지기 때문에 주택가격은 정책 입안자와 투자자에게도 중요한 관심사이며 주택가격의 변동으로 인해 발생하는 부정적인 요인을 예방하기 위해서도 주택가격에 대한 정확한 예측이 필요하다(황세진, 2021). 부동산 가격 예측 정보는 주택 공급 및 인허가 등에 대한

* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임.
(NRF-2019S1A3A2098438)

** 부산대학교 경영대학, myleo@hanmail.net(주저자)

*** 경희대학교 스마트관광연구소, keunmi100@gmail.com

**** 부산대학교 경영대학, hongth@pusan.ac.kr(교신저자)

검토 자료로도 활용이 가능하며, 나아가 금융시장과 연계된 안정적인 주거 정책에 관한 대응 체계의 토대를 마련하는데도 유용한 정보를 제공할 수 있을 것이다.

전국부동산협회에서는 주택가격을 예측하기 위해 과거 데이터에 의존하였으나 주택가격에 영향을 미치는 요인은 다양하게 나타날 수 있으며 Wu and Brynjoflsson(2015)는 빅데이터를 통한 예측성고가 보다 우수하다는 것을 입증하였다. 효과적인 의사결정과 예측성고 향상을 위하여 빅데이터에 대한 중요성이 강조되고 있으며(Chen et al., 2012), 부동산 시장도 예외는 아니다. 한국에서도 부동산 중개 및 거래신고에 관한 법률에 따라 2006년 1월부터 부동산 실거래 신고가 등록이 의무화되었다. 이는 연간 95만 건 이상의 부동산 실거래 자료가 축적되고 있으며 부동산 분야에서도 빅데이터 분석의 적용을 가능하게 하였다. 전통적인 부동산 가격 예측에서는 부동산이 지니는 부동산의 특성을 반영하여 부동산 가격을 예측하고 있다. 즉, 대상 부동산이 속한 지역의 영향을 받는다고 보았으며 주택가격 변동은 지역간 높은 상관관계를 가진다고 보았다(Oikarinen, 2004). 이에 따라 특정 지역의 주택 가격을 예측하기 위해 지역성을 중시하며 타지역의 주택가격을 고려하였다(Clapp et al., 1995). 그러나 최근 빅데이터를 활용한 부동산 가격 예측연구에서는 부동산의 본질인 지역성은 고려하지 않고 예측성고의 향상에만 중점을 두고 예측성고를 비교하는 연구가 주를 이루고 있다(Kuşan et al., 2010; Wang et al., 2014; Park and Bae, 2015). 지역성은 대상 부동산이 속한 지역의 특성을 파악하고 그 특성이 부동산의 가격형성에 미치

는 영향을 분석할 수 있기 때문에 부동산의 가격을 결정하는 중요한 요인이다(경응수, 2010). 따라서 본 연구에서는 부동산이 가지는 부동산성에 대한 특성을 반영한 지역 간의 가격 영향을 고려하기 위해 지역성을 반영하여 부동산 가격을 예측하며 딥러닝을 적용하고자 한다.

이를 위해 본 연구에서는 국토교통부에서 제공하는 전국 아파트에 대한 실거래가를 활용한다. 부동산의 가격 데이터는 시계열 데이터로 계절성을 가지고 있으며 추세와 주기적 구성요소가 혼재되어 있어 전처리 과정이 중요하다. 시계열 데이터의 추세를 추출하기 위해 HP filter(Hodrick-Prescott filter)기법을 적용하여 고품질의 데이터로 향상시키고 적절한 대상 시장 지역을 확정하기 위하여 데이터마이닝 기법 중 하나로 인공신경망에 기초한 군집분석인 SOM(Self-organizing map) 기법을 적용하여 유사한 가격 방향을 나타내는 지역을 클러스터링한다. 유사한 군집 지역을 토대로 입력변수에 대한 설명력을 제시하고, 분석대상에 대한 미래의 가격 동향에 대한 예측모형을 구축하기 위해 SVR(Support Vector Regression) 기법과 딥러닝 기법 중 시계열 데이터 예측에 적합한 LSTM(Long Short-Term Memory) 기법을 활용한다. LSTM은 장기 시계열 데이터 예측에 적합하며 시계열 데이터의 예측에 많이 사용되는 기법으로 부동산 가격 예측에서도 우수한 성과를 보이는지 예측성고를 비교하고자 한다.

II. 이론적 배경

2.1 부동산의 지역성

부동산은 대상 부동산이 속한 지역의 구성요소로 그 지역 및 다른 부동산과 상호 의존, 보완, 협동, 대체, 경쟁의 관계에 있다. 이러한 상호관계를 통하여 사회적, 경제적, 행정적 위치가 결정되는 것을 지역성이라 한다. 경응수(2010)는 지역성을 부동산 자체에서 발생하는 자연적 특성으로 볼 수도 있으나, 부동산과 인간과의 관계에서 발생하는 인문적 특성을 지니는 것으로 보는 것이 타당하다고 하였다. 부동산의 지역성으로 인해 지역 내의 부동산들은 상호 유사한 이용 패턴을 보일 뿐 아니라 가격 수준도 일정한 범위 내에서 같이 움직이는 동조화 현상을 보이게 된다. 아울러 지역성은 고정되어 있지 않고 끊임없이 변화하기 때문에 사회적, 경제적, 행정적 위치를 동태적으로 파악할 필요가 있다.

부동산은 부동산성의 특성을 가지며 부동산 가격은 해당 지역의 영향을 많이 받는다. 따라서 부동산의 가격을 평가할 때 대상 부동산 자체에 대한 분석만으로는 정확한 가격을 구하는데 한계가 있으므로 먼저 대상 부동산이 속한 지역에 대한 분석이 필요하다. 미국의 Appraisal Institute는 neighborhood(근린)를 ‘상호보완적인 토지이용이 이루어지는 집단’으로 정의하였다(Appraisal Institute). Clapp et al.(1995)은 주택가격의 변동이 해당 내 지역뿐 아니라 이웃 지역에도 영향을 미치는 주택가격 변동에 대한 확산 증거를 발견하였으며 Pollakowski and Ray(1997)는 여러 국가의 지역 주택시장들이 서로 강한 상호 관련성을 보이고 있음을 확인하였다. Tu(2000)은 주택시장에서 세분화된 지

역 주택 시장 분석에 대한 중요성을 강조하였으며, 주택가격 변동은 지역 시장 간에 상당한 상관관계가 있음을 확인하였다(Oikarinen, 2004). 따라서 부동산 가격을 평가하기 위해서는 대상 부동산 자체에 대한 분석만으로는 정확한 가격을 구하는 데 한계가 있으며 대상 부동산이 속한 지역에 대한 분석이 먼저 이루어져야 한다.

2.2 지역분석을 위한 클러스터링

클러스터링이란 집단에 대한 사전 지식 없이 유사한 데이터를 동일한 집단으로 분류하는 기법이다(Rai and Singh, 2010). 클러스터는 집단 내에서는 최대 유사성을 가지며, 다른 집단 간에는 최소 유사성을 가지도록 집단을 구성한다. 클러스터링은 판매, 주식, 환율, 날씨, 생의학 측정 등 다양한 분야에서 광범위하게 활용되고 있다.

시계열 데이터에서의 클러스터링은 복잡하고 거대한 데이터셋으로부터 일정한 패턴을 감지하여 의미 있는 정보를 추출하기 위해 사용된다. 시계열 데이터에는 데이터 자체에 잡음, 이상치, 변화 등이 포함되어 있고, 시계열 각 지점의 값이 연대순으로 이루어진 하나 이상의 관측치를 의미하는 다차원 클러스터링 알고리즘으로 다루기가 어렵다(Lin et al., 2004). Nassirtoussi et al.(2014)은 시계열 데이터에 대한 클러스터링을 전체 시계열 클러스터링, 하위 시퀀스 클러스터링, 그리고 시점 클러스터링의 3가지 범주로 분류하였다. 전체 시계열 클러스터링은 유사성과 관련하여 개별 시계열 집합의 클러스터링으로 간주하며 시계열인 개별 객체

에 기존 클러스터링을 적용하는 것을 의미한다. 하위 시퀀스 클러스터링은 하나의 긴 시계열에서 세그먼트 클러스터링인 슬라이딩 윈도우를 통해 추출된 시계열의 서브 시퀀스의 집합을 클러스터링하는 방법이다. 시점 클러스터링은 시간 포인트의 시간적 근접성과 해당 값에 대한 유사성의 조합을 기반으로 하며 시계열 세분화와 유사하다. Keogh and Lin(2005)은 하위 시퀀스 클러스터링은 단일 시계열에 수행되기 때문에 클러스터링의 의미가 없다고 하였다. 시점 클러스터링은 시계열 데이터의 시점 포인트의 클러스터를 찾는 것이 목적이기 때문에 본 연구에서는 전체 시계열 클러스터링을 적용하고자 한다. 클러스터링 기법으로는 k-means 알고리즘과 SOM이 널리 사용되고 있으며, SOM은 신경망 기반의 클러스터링 알고리즘으로 주어진 문제를 세분화하는데 많이 적용되고 있다 (이건창, 정남호, 2002).

2.3 부동산 가격 예측 연구

부동산 가격은 경제지표의 역할을 하기도 하며 주택가격 조정은 경기 주기의 결정에도 중요하기 때문에(Leamer, 2007; Vargas-Silva, 2008), 정확한 부동산 가격 예측은 중요하다.

기존 전통적인 연구에서는 <표 1>과 같이 부동산 가격 예측을 위해 횡단면 분석과 시계열 분석 중심의 통계적 기법에 의하여 이루어졌다. 횡단면 분석에는 회귀분석과 헤도닉 모형이 적용되었으며 시계열 분석에서는 단변량과 다변량 시계열 분석으로 크게 분류할 수 있다.

부동산 실거래 신고가 의무화됨에 따라 방대한 양의 부동산 데이터가 축적되었으며 최근에는 보다 정확한 예측성과를 위해 기계학습 알고리즘을 적용한 연구가 많이 이루어지고 있다. Kauko et al.(2002)은 핀란드 헬싱키의 주택 시장에 신경망을 적용하여 데이터셋에서 패턴을 발견함으로써 주택 하위 시장 형성의 다양한 차원을 식별하였다. Fan et al.(2006)은 다양한 트리 기반 접근 방식을 제안하여 주택가격과 주택 특성 간의 관계를 조사하기 위한 다양한 패턴을 파악하였다. Liu et al.(2006)은 새로운 부동산에 대한 적절한 가격수준을 추정하기 위해 헤도닉적 가격 이론에 기반한 퍼지 신경망 예측모델을 제안하였으며 제안한 방법론이 부동산 가격예측에 적합하다는 것을 보여주었다. Selim(2009)은 헤도닉 회귀 모델과 인공신경망 모델 간의 예측 성능을 비교하였으며 인공신경망이 주택가격 예측을 위한 개선된 대안이 될 수 있음을 보여주었다. Kuşan et al.(2010)은 주

<표 1> 부동산 가격 예측 연구

구 분	분석방법	연구자
횡단면 분석	회귀분석	Bailey et al.(1963) Benjamin et al.(2004)
	헤도닉 모형	Meese and Wallace(2003) Bin(2004)
시계열 기법	단변량 분석	Patrick et al.(2000)
	다변량 분석	Das et al.(2009) Forni et al.(2003)

택 건설 판매 가격 예측을 위해 퍼지 논리 모델을 제안하여 주택 판매가격을 예측하였으며 Azadeh et al.(2012)은 주택 시장 변동을 예측하고 최적화 문제를 위해 퍼지 선형 회귀 및 퍼지인지 맵을 기반으로 하는 하이브리드 알고리즘을 도입하였다.

한편으로는 보다 우수한 예측모델을 식별하기 위해 기계학습 알고리즘 간의 성능을 비교하였다. 조보근 등(2020)은 전국 3개 도시에 대한 아파트 실거래 가격지수를 활용하여 ARIMA모형, 랜덤 포레스트, LSTM 등을 활용한 다양한 예측모형을 구축하여 예측성능을 비교하였다. Gerek(2014)은 주택 판매 가격을 추정하기 위해 두 가지의 적응형 신경 퍼지 (ANFIS) 접근법을 제안하였다. 그리드 파티션 모델을 사용하는 ANFIS가 하위 클러스터링 모델을 사용하는 ANFIS보다 우수한 성능을 보였으며 파티션 기술을 사용한 ANFIS가 주택가격 산정에 성공적으로 사용되었다. Wang et al.(2014)은 PSO(Particle Swarm Optimization) 및 SVM(Support Vector Machine)을 기반으로

부동산 가격 예측 모델을 제안하였으며 제안된 PSO-SVM은 부동산 가격 예측에서 그리드 및 유전자 알고리즘에 비해 우수한 예측성과를 보여주었다. Gu et al.(2011)은 유전자 알고리즘과 SVM의 하이브리드 기법으로 G-SVM을 적용하여 주택가격을 예측하였으며 G-SVM의 예측성과가 기존 방법보다 우수하다는 것을 입증하였다. <표 2>는 부동산 가격 예측을 위해 기계학습을 적용한 연구를 정리하였다.

2.4 시계열 예측을 위한 딥러닝 기법

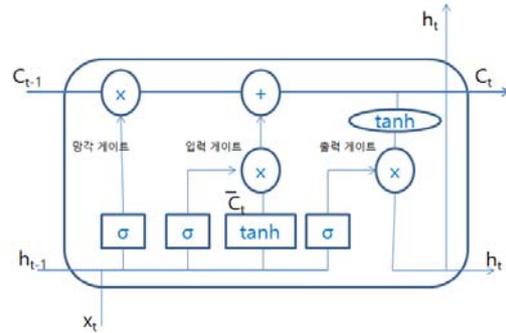
인공신경망(Neural networks)은 데이터 생성과 관련된 가정 없이 모집단의 새로운 관측치에 대한 예측치 수행과 샘플간 복잡한 관계를 수행한다. 인공신경망은 단일 시계열 데이터 예측을 위한 강력한 도구로 지지받아왔으며 내부적인 문제를 보완하기 위해 다양한 기법들이 개발되었다. MLP를 비롯하여 RNN(Recurrent Neural Networks), ESN(Echo State Networks), GRNN(Generalized Regression Neural

<표 2> 기계학습을 적용한 부동산 가격 예측연구

주제	데이터	적용기법	연구자
부동산 가격 예측	2000~2010년 중국 중경 부동산 평균 판매가격	SVM, BP neural network	Wang et al. (2014)
부동산 평가	2005~2006년 리투아니아특정 도시 주택 판매거래	SVM, MLP	Kontrimas and Verikas(2011)
아파트 평가	2010년 러시아 상트페테르부르크2인실 아파트	회귀분석, CHAID, Exhaustive CHAID, CART, KNN, Boosted Trees, Random Forest, MLP	Antipov and Pokryshevskaya(2012)
거래가격 예측	2004~2007년 중개광고 사이트	C4,5, RIPPER, Naive Bayesian, AdaBoost	Park and Bae(2015)
미국 주택가격 지수 예측	1890~2012 연례 관찰 지수	VAR, Bayesian VAR, SVR	Plakandaras et al.(2015)

Networks), LSTM(Long Short-Term Memory) 등이 대표적이며 최근에는 딥러닝 기법인 순환 신경망(RNN)이 시계열 데이터 예측에 많이 적용되고 있다(Lipton et al., 2015). LSTM은 RNN의 한 형태로 바닐라 RNN의 한계점을 보완해주는 특별한 형태이다(Lee et al., 2015). LSTM 네트워크는 Hochreiter and Schmidhuber(1997)에 의해 소개되었으며 바닐라 RNN의 장기 메모리를 보완하면서 RNN에서 정보 흐름을 규제하는 메모리 셀과 게이팅 메커니즘 개념을 확장한 것이다. 게이팅 메커니즘은 입력, 망각, 출력 게이트를 의미하며 이는 메모리 셀로부터 정보가 읽혀지면서부터 얼마나 오랫동안 유지할 것인지 결정한다. 따라서 LSTM은 출력 신호의 주요 정보를 유지하거나, 중요하지 않은 정보를 무시하는 능력을 가지며 메모리 셀은 장기상태에 대한 의존상태를 유지할 것인지 결정한다. 즉, LSTM은 정보를 저장하고 네트워크상 더 긴 기간 동안 오류를 감소시키는 작업을 통하여 RNN의 단점을 보완하게 된다. <그림 1>은 LSTM의 기본구조로 네트워크가 장기 상태에 저장할 것, 버릴 것, 그리고 읽어 들일 것을 학습하는 것이다.

LSTM은 원 시계열 데이터의 순차와 관련이 있고 학습기간 동안 기억해야 할 필요성이 있



<그림 1> LSTM 구조

다. 특정한 시계열 데이터와 관련 있는 학습 배치동안 한 기간이 읽어진다. 그러므로 LSTM 상태는 각 기간에 대해서 초기화될 필요가 있다. 전형적으로 영벡터가 사용되어지나, 고정된 가중치 벡터의 세트로 학습된 LSTM 네트워크는 각 시계열 기간 동안 다른 예측 행동패턴을 보여줄 수 있다. 학습된 네트워크 즉, 특정한 시계열 기간은 그에 맞는 가중치 조합과 이에 일치하는 내적 상태에 의해서 표현된다(Smyl and Kuber, 2016). LSTM은 <표 3>과 같이 다양한 분야에서 활용되고 있으며 특히 시계열 분야에서 각광받고 있다. 이현상과 오세환(2020)은 KIS 신용평점의 시계열 예측을 위해 LSTM을 적용하였으며 타기법에 비해 예측성도가 2배 정도 높아진 것을 확인하였다.

<표 3> LSTM 활용 분야

분야	주요연구
금융분야	Kim and Won(2019), Baek and Kim(2018)
자연어 처리	Lee et al.(2015), Sangeetha and Prabha(2020), Wang et al.(2020)
시계열 예측 분야	Sagheer and Kotb(2019), Karevan and Suykens(2020), Zhao et al.(2017), 이현상, 오세환(2020)

Ⅲ. 연구 프레임워크

본 연구의 프레임워크는 <그림 2>와 같다. 먼저 국토교통부에서 제공하는 공공데이터 API를 통해 지역별 월별 부동산 가격정보를 수집하고 시계열 데이터에 대한 전처리를 수행한다. 부동산 데이터는 시계열 데이터로 추세와 주기적 구성요소가 혼재되어 있어 고품질의 데이터를 위해 HP 필터(Hodrick-Prescott filter)를 적용한다. 다음으로 부동산의 지역성을 반영하기 위해 지역기반의 군집분석을 수행한다. 지역기반의 군집분석은 각 지역별로 SOM을 적용하여 유사한 지역 군집을 형성하며 유사한 군집으로 형성된 군집 내 지역들에 대하여 가격 변동에 대한 인과관계가 있는지 파악한다. 본 연구에서는 가격변동이 유사한 군집들 중에 하나의 지역을 선정하여 부동산 가격예측모형을 구축하였으며 기계학습 기법인 SVR과 딥러닝 기법인 LSTM 기법을 적용하여 부동산 가격 예측모형을 구축하고 예측성과를 비교하여 부동산 가격 예측에서도 딥러닝 기법이 우수한 성과를 보이는지 살펴본다.

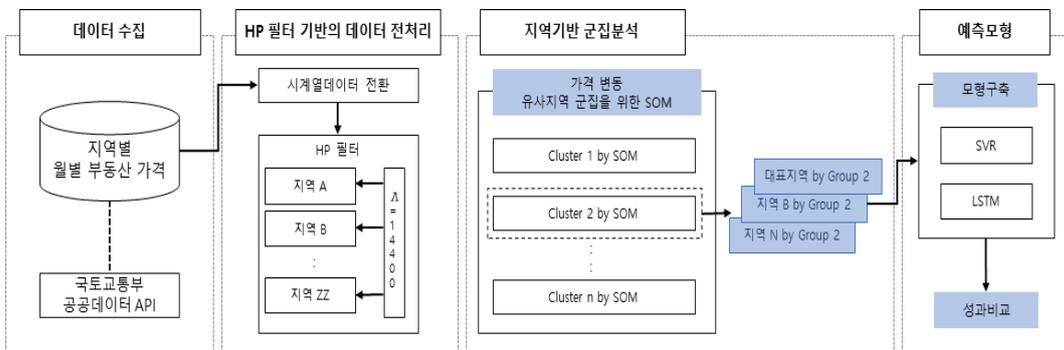
Ⅳ. 실험 및 결과

4.1 데이터

본 연구에서는 국토교통부에서 제공하는 공동주택 실거래가를 활용하였다. 국토교통부의 공공데이터 API를 통해 2006년 1월부터 2019년 12월까지 198개의 자치구에 대한 아파트 실거래가의 월 가격지수를 수집하였다. 그 중 2006년 1월부터 실거래가가 제공되지 않는 자치구는 제외하고 192개의 지역에 대해 총 168개의 데이터를 분석에 사용하였다. 수집된 데이터는 월별로 각 자치구의 평균 가격으로 나타내었으며 모형 구축을 위한 학습용 데이터와 모형 검증을 위한 검증용 데이터를 4:1로 분류하여 학습용 데이터 135개, 검증용 데이터 33개로 사용하였다.

4.2 HP 필터

시계열 데이터에는 추세와 잡음이 혼재되어 있으며 이는 추세와 잡음으로 분해가 가능하다. 추세분해는 자의성, 후행성을 내포하고 있는데



<그림 2> 연구 프레임워크

이러한 한계를 어느 정도 극복하기 위해 HP 필터를 적용하였다. HP 필터는 Whittaker(1922)가 도입하고 Hodrick and Prescott(1997)에 의해 특정 형태로 경제학에서 널리 활용되는 데이터 필터링 기법이다(Peter and Sainan, 2015). Hodrick and Prescott(1997)에서 제안한 바와 같이 시계열은 선형을 나타내는 추세 성분과 비선형을 나타내는 순환성분으로 구성된 것으로 간주 할 수 있다.

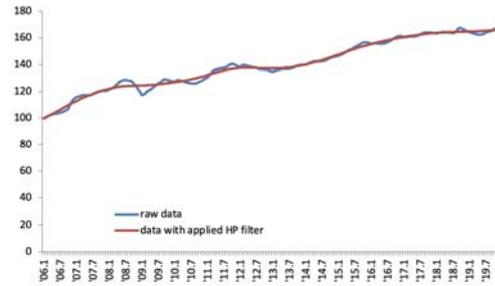
시계열을 $y = g_t + c_t$ 와 같이 장기 평균이 0인 사이클 부분 c_t 와 추세부분 g_t 로 구분하는데 $g_t - g_{t-1}$ 로 정의된 성장률의 차이인 $[(g_t - g_{t-1}) + (g_{t-1} - g_{t-2})]$ 를 적당한 λ 에 대해 식(1)과 같이 값을 최소화하도록 시계열을 분석한다.

$$\min_{g_t} \sum_{t=1}^T c_t^2 + \lambda \sum_{t=1}^T [(g_t - g_{t-1}) + (g_{t-1} - g_{t-2})]^2$$

식(1)

본 연구에서는 부동산 시계열 데이터를 그대로 사용하지 않고 순환 구성을 분해하는 전처리 과정을 거쳐 시스템의 전반적인 정확도를 향상시키고자 하였다. HP 필터를 적용하면 <그

림 3>과 같이 데이터의 잡음이 제거된 상태로 추세선을 좀 더 명확히 할 수 있다.



<그림 3> HP 필터 적용 데이터

4.3 군집분석 결과

192개의 지역을 가격 변동성이 유사한 지역으로 분류하기 위해 SOM을 적용하였으며 MATLAB을 활용하였다. 월별 실거래 가격 지수를 기반으로 SOM을 수행한 결과, 지역별로 가격 변동성이 유사한 8개의 군집으로 분류되었다. 각 군집에 대한 기초 통계량은 <표 4>와 같이 군집3으로 분류된 지역이 가장 많았으며 군집7로 분류된 지역이 가장 적었다. 또한 실거래 가격 지수의 평균이 가장 높은 군집은 군집4로 나타났으며 가격 지수 평균이 가장 낮은 군

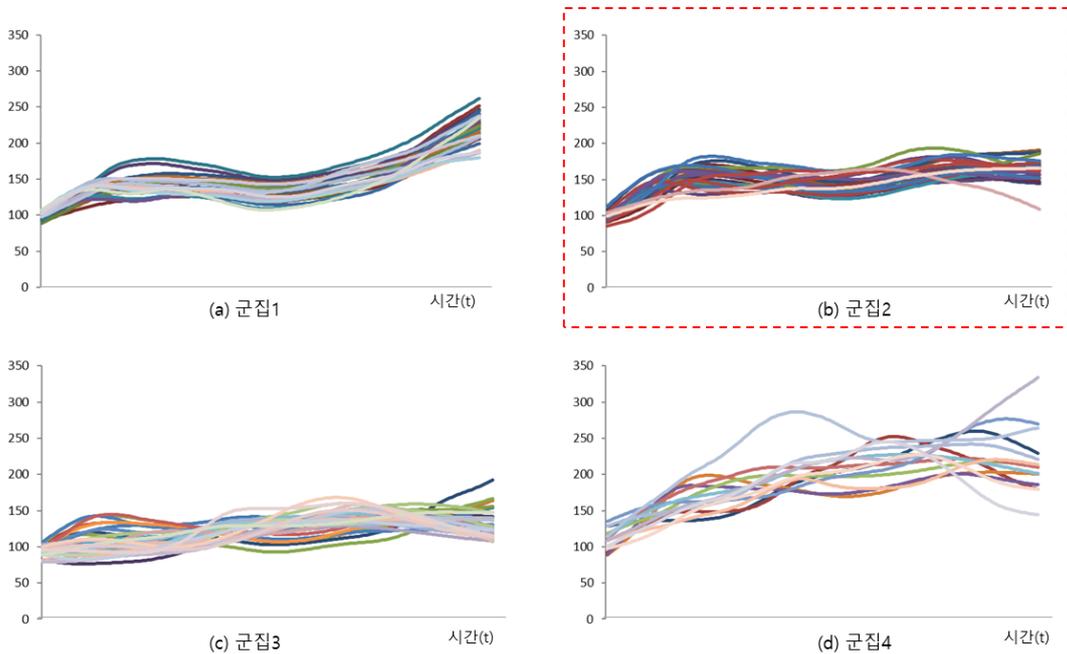
<표 4> 군집별 기초 통계량

군집	군집 내 지역수	가격 지수 평균	표준편차
1	33	148.06	9.19
2	25	148.92	8.28
3	36	122.64	5.90
4	14	193.63	15.80
5	32	152.90	8.22
6	23	94.62	9.59
7	11	165.53	10.30
8	18	136.10	8.18
계	192	145.30	9.43

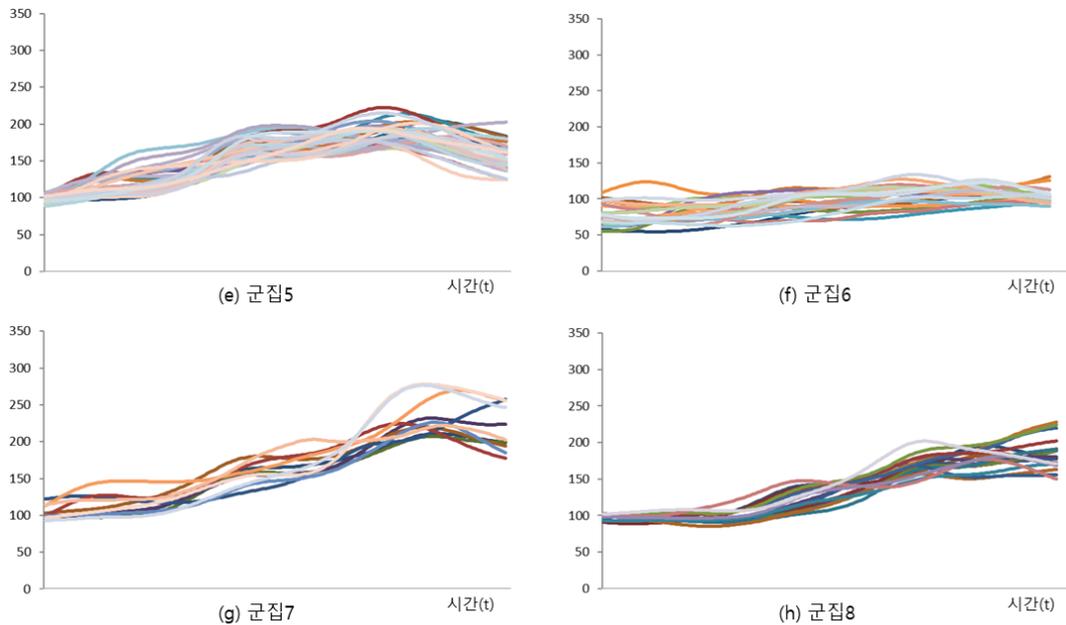
집은 군집6으로 나타났다.

군집분석 결과는 <그림 4>와 <그림 5>와 같다. 동일 군집 내는 지리적으로 가까운 지역이 있으나, 가격수준 및 유사 산업군이 있는 경우에는 동일 군집으로 묶이는 특성을 가지고 있다. 군집1은 서울, 경기 지역과 지방 중 거주 선호도가 높은 지역들로 가격대가 높은 편이며, 지속적인 가격 상승세를 나타내고 있다. 특히 군집2는 전국 아파트 가격이 속한 그룹으로 인천, 경기도내 지역이 동일 군집으로 나타난 특징이 있다. 인천, 경기도가 수도권이면서 전국 아파트 가격을 나타내는 대표 지역이라고 보여진다. 군집3은 지방 광역 및 중소도시, 수도권

신도시가 많은 지역으로 안정적인 가격세를 보이고 있다. 군집4는 가격변동성이 가장 큰 특징을 보이고 있으며, 조선 및 철강을 중심으로 한 산업기반 지역을 두고 있는 군집이다. 군집5는 지방중소 지역이 분포되어 있으며 일정 시점 이후 가격 하락세를 보이는 특징이 있다. 군집6은 지방중소도시 중심으로 가격 보합세를 보이는 특징을 보인다. 군집7은 지방 광역도시가 분포되어 있으며, 가격 상승세를 보이지만 다소 가격 변동성을 나타내는 특징이 있다. 군집8도 지방 광역도시가 분포되어 있으나, 군집7보다는 가격 변동성은 적으면서 가격 상승세를 보이는 특징을 보인다.



<그림 4> SOM에 의한 군집별 아파트 가격 추이(군집1~군집4)



<그림 5> SOM에 의한 군집별 아파트 가격 추이(군집5~군집8)

4.4 지역기반 예측모형 결과

전국 아파트 가격과 유사한 가격변동을 나타내는 군집2 내에서 25개 지역의 월평균 가격을 입력변수로 두고 전국 아파트 가격 예측을 위해 SVR과 LSTM을 적용하였으며 예측성과를 비교하였다. SVR은 R을 사용하였으며 Liner 커널함수를 사용하고 파라미터인 C와 σ 는 그리드 탐색기법을 적용하여 최적의 값을 선정하였다. LSTM은 python의 keras 패키지를 활용하였으며 ReLu함수를 적용하였다. 부동산 가격은 주식과 다르게 변동성이 적은 시장이기 때문에 일간 가격이 아니라 월간 가격을 분석의 대상으로 삼았다. 금융시장 내 안정적인 가격을 나타내는 부동산 가격으로 1개월 후, 2개월 후, 3개월 후의 가격을 각각 예측하였다. 가격 예측 정확도는 RMSE(Root Mean Squared Error)로

나타내었으며 <표 5>와 같다. 1개월 후는 SVR의 RMSE가 1.77%이며 LSTM의 RMSE가 1.42%로 LSTM이 SVR보다 우수한 것으로 나타났다. 2개월 후의 예측성과에서도 SVR(1.64%)보다 LSTM(1.18%)의 예측성과가 우수한 것으로 나타났으며 3개월 후의 예측성과에서도 SVR(1.64%)보다 LSTM(1.38%)에서 우수한 예측성과를 보여주었다. 시계열 데이터를 적용한 예측모형에서 딥러닝 기법인 LSTM이 SVR보다 우수함을 확인할 수 있었다. SVR과 LSTM 결과 모두, 1개월 후의 예측보다 2개월 후와 3개월 후의 예측성과가 우수한 것으로 나타났다. 이는 부동산은 장기적으로 우상향의 경향을 나타내며 예측 시점이 미래일수록 변동성이 많이 반영되지 않으며 가까운 시점의 예측에서는 미래에 대한 예측보다 높은 변동성의 특성에 기인한 것이라 할 수 있다. 또한 LSTM

은 장기 기억 의존성에 대한 문제를 해결하며 시계열 예측에 적합하다는 것을 SVR의 예측성 과를 통해 확인할 수 있다.

<표 5> SVR과 LSTM의 예측 결과(RMSE)

구분	1개월	2개월	3개월
SVR	1.77%	1.64%	1.64%
LSTM	1.42%	1.18%	1.38%

V. 결론 및 향후 연구과제

본 연구는 전국 부동산의 가격 예측을 위하여 지역분석을 기반으로 하였으며, 시계열 데이터의 잡음을 제거하기 위하여 HP 필터를 적용하고 예측을 위한 유사 가격 변동 지역을 SOM을 통하여 선정하였다. 부동산은 부동산의 특성을 가지기 때문에 해당 지역의 특성도 함께 고려되어야 하지만 기계학습을 적용한 기존의 부동산 가격 예측 연구에서는 부동산 가격에 중요한 영향을 미치는 지역성의 중요성을 간과하였다. 본 연구에서는 지역성을 기반으로 신경망 기반의 SOM을 적용하여 193개의 지역을 8개의 군집으로 분류하였다. 동일한 군집의 지역은 지리적으로 가까운 지역이었으나, 가격수준 및 유사 산업군이 있는 경우에도 동일군집으로 분류되는 특성을 확인할 수 있었다. 부동산 가격 예측모형은 기계학습 기법인 SVR과 딥러닝 기법인 LSTM으로 구축하였으며 SOM을 통해 유사 가격 변동지역으로 선정된 군집 중 군집2를 대상으로 하였다. 아파트 가격에 영향을 미치는 군집2에 속한 25개 지역의 아파트 월평균 가격을 입력변수로 선정하고, 1개월, 2개월, 3개월

후의 가격을 각각 예측하였다. 예측 결과 1개월, 2개월, 3개월 후의 가격 모두 SVR보다 LSTM에서 우수한 예측성과를 보였으며 LSTM은 1개월 후의 단기 예측보다 3개월 후의 장기 예측에서 보다 우수한 예측성과를 보였다.

본 연구의 시사점은 다음과 같다. 첫째, 부동산 가격 예측을 위해 지역성을 반영하였다. 부동산은 유사 지역간 유기적으로 상호작용하기 때문에 지역분석이 반드시 필요하지만 기존의 기계학습을 적용한 부동산 가격 예측에 지역성이 반영되어 있지 않은 문제를 해결하였다. 둘째, 지역분석 및 부동산 가격 예측을 병행하였다. 본 연구에서 제안하는 유사한 특성을 가진 군집을 대상으로 구축하는 가격 예측모형은 지역과 가격을 동시에 고려하여 부동산 정책 마련시 기초자료로 활용할 수 있을 것이다. 이는 주택 공급 및 인허가 등에 대한 검토 자료로도 활용이 가능하며, 부동산 가격 예측을 통한 정보는 금융정책 및 자산 변동성에 대한 정보로도 활용하여 금융시장과 연계된 안정적인 주거 정책에 관한 대응체계의 토대를 마련하는데도 유용한 정보를 제공할 수 있을 것이다. 셋째, 시계열 데이터 예측에 많이 활용되고 있는 딥러닝 기법인 LSTM을 적용하였다. LSTM은 시계열 데이터에 많이 적용되고 있는 기법으로 LSTM을 부동산 시장에 적용해 보았으며 단기 예측과 장기 예측의 성과를 비교하였다.

본 연구의 한계점 및 향후 연구방향은 다음과 같다. 첫째, 부동산 가격 예측모형을 구축하기 위한 입력변수로 전국 아파트의 가격 지수만을 예측 변수로 하였다. 그러나 부동산 가격에 영향을 미치는 요인을 가격과 지역 외에도 거시경제 지표 등과 같은 다양한 요인이 있을

수 있으며 경제뉴스와 같은 텍스트 정보도 반영해 볼 수 있다. 향후 연구에서는 보다 다양한 변수를 사용하여 예측모형을 구축하여 다양한 사회환경을 반영할 필요가 있다. 둘째, SOM을 적용한 군집의 수를 유효한 8개의 군집으로 분류하였으나 최적의 군집 수를 결정하기 위해 유전자 알고리즘과 같은 방법을 적용해 볼 수 있다. 셋째, 부동산 가격 예측모형을 구축하기 위해 8개의 군집 중 전국 아파트 가격을 나타내는 대표지역으로 군집2에 대해 예측모형을 구축하였으나 향후에는 군집별로 예측모형을 구축하여 성과를 비교해 볼 수 있을 것이며 가격의 변화와 지역의 특성을 파악할 수 있을 것으로 기대한다.

참고문헌

경응수, “감정평가론,” 나무미디어, 2018.

이건창, 정남호, “SOM 을 이용한 인터넷 주식 거래시장의 시장세분화 전략수립에 관한 연구,” 한국경영과학회지, 제27권, 제3호, 2002, pp. 75-92.

이현상, 오세환, “시계열 예측을 위한 LSTM 기반 딥러닝: 기업 신용평점 예측 사례,” 정보시스템연구, 제29권, 제1호, 2020, pp. 241-265.

조보근, 박경배, 하성호, “기계학습 알고리즘을 활용한 지역 별 아파트 실거래가격지수 예측모델 비교: LIME 해석력 검증,” 정보시스템연구, 제29권, 제3호, 2020, pp. 119-144.

황세진, “주택가격의 단기 예측성과에 관한 연

구,” 부동산학연구, 제27집, 제1호, 2021, pp. 91-108.

Antipov, E. A., and Pokryshevskaya, E. B., “Mass Appraisal of Residential Apartments: An Application of Random Forest for Valuation and A CART-based Approach for Model Diagnostics,” *Expert Systems with Applications*, Vol. 39, No. 2, 2012, pp. 1772-1778.

Appraisal Institute, Available: <https://www.appraisalinstitute.org/>

Azadeh, A., Ziaei, B., and Moghaddam, M., “A Hybrid Fuzzy Regression-fuzzy Cognitive Map Algorithm for Forecasting and Optimization of Housing Market Fluctuations,” *Expert Systems with Applications*, Vol. 39, No. 1, 2012, pp. 298 - 315.

Baek, Y., and Kim, H. Y., “ModAugNet: A New Forecasting Framework for Stock Market Index Value with An Overfitting Prevention LSTM Module and A Prediction LSTM Module,” *Expert Systems with Applications*, Vol. 113, 2018, pp. 457-480.

Bailey, M. J., Muth, R. F., and Nourse, H. O., “A Regression Method for Real Estate Price Index Construction,” *Journal of the American Statistical Association*, Vol. 58, No. 304, 1963, pp. 933-942.

Benjamin, J., Guttery, R., and Sirmans, C. F., “Mass Appraisal: An Introduction to

- Multiple Regression Analysis for Real Estate Valuation,” *Journal of Real Estate Practice and Education*, Vol. 7, No. 1, 2004, pp. 65-77.
- Bin, O., “A Prediction Comparison of Housing Sales Prices by Parametric versus Semi-parametric Regressions,” *Journal of Housing Economics*, Vol. 13, No. 1, 2004, pp. 68-84.
- Chen, H., Chiang, R. H. and Storey, V. C., “Business Intelligence and Analytics: From Big Data to Big Impact,” *MIS Quarterly*, Vol. 36, No. 4, 2012, pp. 1165-1188.
- Clapp, J. M., Dolde, W., and Tirtiroglu, D., “Imperfect Information and Investor Inferences from Housing Price Dynamics,” *Real Estate Economics*, Vol. 23, No. 3, 1995, pp. 239-269.
- Das, S., Gupta, R., and Kabundi, A., “Could We Have Predicted the Recent Downturn in the South African Housing Market?,” *Journal of Housing Economics*, Vol. 18, No. 4, 2009, pp. 325-335.
- Fan, G. Z., Ong, S. E., and Koh, H. C., “Determinants of House Price: A Decision Tree Approach,” *Urban Studies*, Vol. 43, No. 12, 2006, pp. 2301-2315.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L., “Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area?,” *Journal of Monetary Economics*, Vol. 50, No. 6, 2003, pp. 1243-1255.
- Gerek, I. H., “House Selling Price Assessment using Two Different Adaptive Neuro-fuzzy Techniques,” *Automation in Construction*, Vol. 41, 2014, pp. 33 - 39.
- Hochreiter, S., and Schmidhuber, J., “Long Short-term Memory,” *Neural Computation*, Vol. 9, No. 8, 1997, pp. 1735-1780.
- Hodrick, R. J., and Prescott, E. C., “Postwar US Business Cycles: An Empirical Investigation,” *Journal of Money, Credit, and Banking*, 1997, pp. 1-16.
- Karevan, Z., and Suykens, J. A., “Transductive LSTM for Time-series Prediction: An Application to Weather Forecasting,” *Neural Networks*, Vol. 125, 2020, pp. 1-9.
- Kauko, T. O. M., Hooimeijer, P., and Hakfoort, J., “Capturing Housing Market Segmentation: An Alternative Approach based on Neural Network Modelling,” *Housing Studies*, Vol. 17, No. 6, 2002, pp. 875-894.
- Keogh, E., and Lin, J., “Clustering of Time-series Subsequences is Meaningless: Implications for Previous and Future Research,” *Knowledge and Information Systems*, Vol. 8, No. 2, 2005, pp. 154-177.

- Kim, H. Y., and Won, C. H., "Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-type Models," *Expert Systems with Applications*, Vol. 103, 2018, pp. 25-37.
- Kontrimas, V., and Verikas, A., "The Mass Appraisal of the Real Estate by Computational Intelligence," *Applied Soft Computing*, Vol. 11, No. 1, 2011, pp. 443-448.
- Kuşan, H., Aytakin, O., and Özdemir, İ., "The Use of Fuzzy Logic in Predicting House Selling Price," *Expert Systems with Applications*, Vol. 37, No. 3, 2010, pp. 1808-1813.
- Leamer, E. E., "Housing Is the Business Cycle," *National Bureau of Economic Research*, 2007.
- Lee, C. Y., Xie, S., Gallagher, P., Zhang, Z., and Tu, Z., "Deeply-supervised Nets," *Artificial Intelligence and Statistics*, PMLR, 2015, pp. 562-570.
- Lin, J., Keogh, E., Lonardi, S., Lankford, J. P., and Nystrom, D. M., "Visually Mining and Monitoring Massive Time Series," *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 460-469.
- Lipton, Z. C., Kale, D. C., Elkan, C., and Wetzel, R., "Learning to Diagnose with LSTM Recurrent Neural Networks," *arXiv preprint arXiv:1511.03677*, 2015.
- Liu, J. G., Zhang, X. L., and Wu, W. P., "Application of Fuzzy Neural Network for Real Estate Prediction," *In International Symposium on Neural Networks*, Springer, Berlin, Heidelberg, 2006, pp. 1187-1191.
- Meese, R., and Wallace, N., "House Price Dynamics and Market Fundamentals: The Parisian Housing Market," *Urban Studies*, Vol. 40, No. 5-6, 2003, pp. 1027-1045.
- Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., and Ngo, D. C. L., "Text Mining for Market Prediction: A Systematic Review," *Expert Systems with Applications*, Vol. 41, No. 16, 2014, pp. 7653-7670.
- Oikarinen, E., "The Diffusion of Housing Price Movements from Center to Surrounding Areas," *Journal of Housing Research*, Vol. 15, No. 1, 2004, pp. 3-28.
- Park, B., and Bae, J. K., "Using Machine Learning Algorithms for Housing Price Prediction: The Case of Fairfax County, Virginia Housing Data," *Expert Systems with Applications*, Vol. 42, No. 6, 2015, pp. 2928-2934.
- Patrick, J., Okunev, J., Ellis, C., and David, M., "Comparing Univariate Forecasting Techniques in Property Markets," *Journal of Real Estate Portfolio Management*, Vol. 6, No. 3, 2000, pp.

- 283-306.
- Plakandaras, V., Gupta, R., Gogas, P., and Papadimitriou, T., "Forecasting the US Real House Price Index," *Economic Modelling*, Vol. 45, 2015, pp. 259-267.
- Pollakowski, H. O., and Ray, T. S., "Housing Price Diffusion Patterns at Different Aggregation Levels: An Examination of Housing Market Efficiency," *Journal of Housing Research*, 1997, pp. 107-124.
- Rai, P., and Singh, S., "A Survey of Clustering Techniques," *International Journal of Computer Applications*, Vol. 7, No. 12, 2010, pp. 1-5.
- Sagheer, A., and Kotb, M., "Time Series Forecasting of Petroleum Production using Deep LSTM Recurrent Networks," *Neurocomputing*, Vol. 323, 2019, pp. 203-213.
- Sangeetha, K., and Prabha, D., "Sentiment Analysis of Student Feedback using Multi-head Attention Fusion Model of Word and Context Embedding for LSTM," *Journal of Ambient Intelligence and Humanized Computing*, 2020, pp. 1-10.
- Selim, H., "Determinants of House Prices in Turkey: Hedonic Regression versus Artificial Neural Network," *Expert Systems with Applications*, Vol. 36, No. 2, 2009, pp. 2843-2852.
- Smyl, S., and Kuber, K., "Data Preprocessing and Augmentation for Multiple Short Time Series Forecasting with Recurrent Neural Networks," *36th International Symposium on Forecasting*, 2016.
- Tu, Y., "Segmentation of Australian Housing Markets: 1989-98," *Journal of Property Research*, Vol. 17, No. 4, 2000, pp. 311-327.
- Vargas-Silva, C., "Monetary Policy and the US Housing Market: A VAR Analysis Imposing Sign Restrictions," *Journal of Macroeconomics*, Vol. 30, No. 3, 2008, pp. 977-990.
- Wang, J., Yu, L.C, Lai, K. R., and Zhang, X., "Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis," *IEEE/ACM Transactions on Audio Speech and Language Processing*, Vol. 28, 2020, pp. 581-591.
- Wang, X., Wen, J., Zhang, Y., and Wang, Y., "Real Estate Price Forecasting Based on SVM Optimized by PSO," *Optik*, Vol. 125, No. 3, 2014, pp. 1439-1443.
- Whittaker, E. T., "On A New Method of Graduation," *Proceedings of the Edinburgh Mathematical Society*, Vol. 41, 1922, pp. 63-75.
- Wu, L., and Brynjolfsson, E., "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales," *In Economic Analysis of the Digital Economy*, University of Chicago Press, 2015, pp. 89-118.
- Zhao, Z., Chen, W., Wu, X., Chen, P. C., and

Liu, J., “LSTM Network: A Deep Learning Approach for Short-term Traffic Forecast,” *IET Intelligent Transport Systems*, Vol. 11, No. 2, 2017, pp. 68-75.

신 은 경 (Shin, Eun Kyung)



부경대학교 공학학사와 부산대학교 석사를 취득하였다. 현재 부산대학교 경영정보학과 박사수료 하였으며, 주요 관심분야는 애널리틱스, 딥러닝, 빅데이터, 오피니언 마이닝 등이다.

김 은 미 (Kim, Eun Mi)



부산대학교 경영학과에서 석사학위와 박사학위를 취득하였다. 현재 경희대학교 스마트관광연구소에서 학술 연구 교수로 재직하고 있다. 주요 관심분야는 비즈니스 애널리틱스, 딥러닝, 소셜미디어, CRM 등이다.

홍 태 호 (Hong, Tae Ho)



부산대학교 경영학과 교수로 재직하고 있다. KAIST에서 경영정보시스템을 전공하여 공학석사와 공학박사를 취득하였다. 주요 관심분야는 비즈니스 애널리틱스, 딥러닝, 오피니언 마이닝, CRM 등이다.

<Abstract>

Real Estate Price Forecasting by Exploiting the Regional Analysis Based on SOM and LSTM

Shin, Eun Kyung · Kim, Eun Mi · Hong, Tae Ho

Purpose

The study aims to predict real estate prices by utilizing regional characteristics. Since real estate has the characteristic of immobility, the characteristics of a region have a great influence on the price of real estate. In addition, real estate prices are closely related to economic development and are a major concern for policy makers and investors. Accurate house price forecasting is necessary to prepare for the impact of house price fluctuations. To improve the performance of our predictive models, we applied LSTM, a widely used deep learning technique for predicting time series data.

Design/methodology/approach

This study used time series data on real estate prices provided by the Ministry of Land, Infrastructure and Transport. For time series data preprocessing, HP filters were applied to decompose trends and SOM was used to cluster regions with similar price directions. To build a real estate price prediction model, SVR and LSTM were applied, and the prices of regions classified into similar clusters by SOM were used as input variables.

Findings

The clustering results showed that the region of the same cluster was geographically close, and it was possible to confirm the characteristics of being classified as the same cluster even if there was a price level and a similar industry group. As a result of predicting real estate prices in 1, 2, and 3 months, LSTM showed better predictive performance than SVR, and LSTM showed better predictive performance in long-term forecasting 3 months later than in 1-month short-term forecasting.

Keyword: Deep Learning, LSTM, Time Series Data, HP Filter, Regional Analysis

* 이 논문은 2021년 6월 1일 접수, 2021년 6월 7일 1차 심사, 2021년 6월 22일 게재 확정되었습니다.