

# 딥 러닝을 이용한 부동산가격지수 예측

Predicting the Real Estate Price Index Using Deep Learning

배성완\* · 유정석\*\*

Bae, Seong Wan · Yu, Jung Suk

## Abstract

The purpose of this study was to apply the deep running method to real estate price index predicting and to compare it with the time series analysis method to test the possibility of its application to real estate market forecasting. Various real estate price indices were predicted using the DNN (deep neural networks) and LSTM (long short term memory networks) models, both of which draw on the deep learning method, and the ARIMA (autoregressive integrated moving average) model, which is based on the time series analysis method. The results of the study showed the following. First, the predictive power of the deep learning method is superior to that of the time series analysis method. Second, among the deep learning models, the predictability of the DNN model is slightly superior to that of the LSTM model. Third, the deep learning method and the ARIMA model are the least reliable tools for predicting the housing sales prices index among the real estate price indices. Drawing on the deep learning method, it is hoped that this study will help enhance the accuracy in predicting the real estate market dynamics.

**Keywords** ▶ deep learning, real estate price index, predicting, DNN, LSTM

## 초 록

본 연구의 목적은 딥 러닝 방법을 부동산가격지수 예측에 적용해보고, 기존의 시계열분석 방법과의 비교를 통해 부동산 시장 예측의 새로운 방법으로서 활용가능성을 확인하는 것이다. 딥 러닝(deep learning)방법인 DNN(Deep Neural Networks)모형 및 LSTM(Long Shot Term Memory networks)모형과 시계열분석 방법인 ARIMA(autoregressive integrated moving average)모형을 이용하여 여러 가지 부동산가격지수에 대한 예측을 시도하였다. 연구결과 첫째, 딥 러닝 방법의 예측력이 시계열분석 방법보다 우수한 것으로 나타났다. 둘째, 딥 러닝 방법 중에서는 DNN모형의 예측력이 LSTM모형의 예측력보다 우수하나 그 정도는 미미한 수준인 것으로 나타났다. 셋째, 딥 러닝 방법과 ARIMA모형은 부동산 가격지수(real estate price index) 중 아파트 실거래가격지수(housing sales price index)에 대한 예측력이 가장 부족한 것으로 나타났다. 향후 딥 러닝 방법을 활용함으로써 부동산 시장에 대한 예측의 정확성을 제고할 수 있을 것으로 기대된다.

**주요 단어** ▶ 딥 러닝, 부동산가격지수, 예측, DNN, LSTM

\* 단국대학교 일반대학원 도시계획및부동산학과 박사수료, 제1저자(swbae618@gmail.com)

\*\* 단국대학교 사회과학대학 도시계획부동산학부 부교수, 교신저자(jsyu@dankook.ac.kr)

## I. 서 론

부동산가격지수는 부동산 시장을 분석하고 예측하기 위한 자료로 활용되며, 부동산가격지수 예측에는 자기회귀이동평균모형(autoressive integrated moving average model, ARIMA), 벡터자기회귀모형(vector autoregression model, VAR), 벡터오차수정모형(vector error correction model, VECM) 등과 같은 시계열분석 방법이 주로 이용되고 있다. 시계열분석방법은 다소 현실과 맞지 않는 선형 모형을 가정하기 때문에 예측효율성이 떨어진다는 문제점이 있으며, 이로 인해 비선형 추정기법과 같은 새로운 분석방법 적용의 필요성이 제기되고 있다. 최근 인공지능(Artificial Intelligence, AI) 관련 기술로서 복잡한 문제해결이나 예측에 탁월한 성능을 보이는 것으로 알려진 딥 러닝(Deep Learning, DL) 방법은 비선형추정기법 중 하나인 인공신경망(Artificial Neural Network, ANN)에서 발전한 방법으로 문자 인식, 영상인식, 음성인식, 날씨예측 등 다양한 분야에서 활용되고 있으며, 주가지수 예측에 있어서 활발한 연구와 함께 좋은 성과를 보여주고 있기 때문에 부동산가격지수 예측의 새로운 방법으로서의 활용가능성을 확인해 볼 필요가 있다.

본 연구의 목적은 딥 러닝 방법을 부동산가격지수 예측에 적용해보고, 기존의 시계열분석 방법과의 비교를 통해 부동산 시장 예측의 새로운 방법으로서 활용가능성을 확인하는 것이다.

분석을 위한 자료인 부동산 가격지수는 한국감정원에서 발표하고 있는 아파트 매매실거래가격지수, 아파트 매매가격지수, 아파트 전세가격지수, 지가지수를 이용하였다. 아파트 매매실거래가격지수가 2006년 1월부터 발표된 점을 감안하여 분석기간은 2006년 1월부터 2016년 12월까지로 설정하였으며, 분석지역은 서울지역으로 한정하였다. 분석방법으로 딥 러닝 방법은 심층 신경망(Deep Neural Networks, DNN)모형과 LSTM(Long Short Term Memory networks)모형을 적용하였고, 시계열분석 모형은 분석자료가 단일 시계열자료임을 감안하여 ARIMA모형을 적용하였다.

본 연구는 기존 시계열분석 방법인 ARIMA모형과

딥 러닝 방법인 DNN과 LSTM의 예측력을 서로 비교하여 부동산가격지수 예측을 위한 새로운 예측 방법으로서 딥 러닝 방법의 적용가능성을 확인하였다는 점에서 의의가 있다.

본 연구의 구성은 다음과 같다. 2장은 이론적 고찰 및 선행연구 검토로서 딥 러닝에 대한 개념과 관련 선행연구를 검토하고, 3장에서는 본 연구에 적용될 분석모형, 분석자료 및 분석방법에 대해 고찰하며, 4장은 실증분석으로 시계열분석과 딥러닝분석 결과를 비교 설명한다. 5장에서는 실증분석 결과를 바탕으로 결론과 시사점, 한계점 및 향후 과제에 대해 고찰한다.

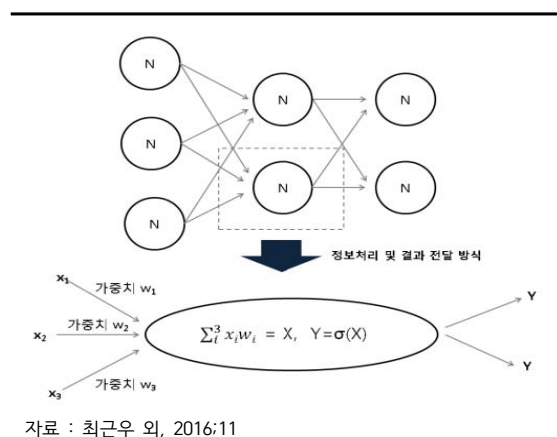
## II. 이론적 고찰 및 선행연구 검토

### 1. 딥 러닝(Deep Learning)이란?

딥 러닝은 인공지능의 한 분야인 머신 러닝(Machine Learning)의 한 분야이다(최근우 외, 2016:3).

딥 러닝은 McCulloch and Pitts(1943)에 의해 고안된 인공신경망에 바탕을 둔 기술로서, 인간의 신경망 구조를 모방한 것이다.

뇌는 서로 연결된 신경세포의 집합으로 이 때 신경세포인 뉴런(neuron)은 정보처리의 단위로서 만 개에서 십만 개의 다른 뉴런과 연결되어 있으며, 연



<그림 1> 인공신경망 모델의 개념도

결되어 있는 뉴런들끼리 전기적 신호를 주고받으며 사물의 인식, 추상적 분류 등의 정보처리를 할 수 있다(최근우 외, 2016: 10).

인간 뉴런을 모방한 인공신경망 정보처리의 최소 단위인 노드(node)는 연결된 이전 노드들로부터 데이터를 받아 처리 후 다음 노드로 전달하게 되는데 이는 <그림 1>과 같다. 데이터 처리는 개별 링크로부터의 입력 값과 가중치 곱의 합인  $X$ 를 활성화 함수인  $\sigma()$ 에 적용하여 결과 값인  $Y$ 를 산출하게 된다(최근우 외, 2016: 10).<sup>1)</sup>

인공신경망 기술은 2차레의 붐(boom)과 침체를 반복하였으며, 최근 3번째 붐을 맞이하고 있다. 인공신경망 기술이 소개된 이후 1960년대까지 이어지던 인공신경망의 첫 번째 붐은 다수 은닉층에서의 신경망 학습이 불가능하다는 Minsky(1969)의 주장 이후 1980년대까지 침체기를 갖는다.

Rumelhart et al.(1986)에 의해 제시된 역전파 알고리즘(backpropagation algorithm)을 통해 다층 인공신경망의 학습이 가능하다는 것이 입증되면서 인공신경망 연구는 두 번째 붐을 맞이하게 된다. 하지만 여전히 깊은 은닉층을 가지고 있는 인공신경망의 경우 학습이 잘 되지 않았으며, 랜덤포레스트(Random Forest), 서포트 벡터 머신(Support Vector Machine, SVM) 등과 같은 인공신경망보다 늦게 소개된 기계학습 방법보다 성과가 떨어진다는 이유로 인공신경망 연구는 2번째 침체기를 맞게 된다.<sup>2)</sup>

2000년 초반에 다수의 은닉층을 가지고 있는 신경망을 학습시킬 수 있는 심층 신뢰 신경망(Deep Belief Network, DBN), 오토인코더(Autoencoder) 등과 같은 방법이 제시되었고, 이러한 방법이 적용된 인공신경망은 딥 러닝이라는 새로운 이름으로 세 번째 붐을 맞이하게 되었다.

딥 러닝 모형은 알고리즘의 구조에 따라 다양하게 분류되며, 대표적으로 합성곱 신경망(convolutional neural network, CNN), 순환신경망(recurrent neural network, RNN), DNN, DBN 등이 있다.

## 2. 선행연구 검토

김근용(1998)은 ARIMA모형과 State-space모형을 통해 주택매매 및 전세가격지수에 대한 단기적인 예측을 시도하였으며, 매매가격지수에 대한 적합도는 ARIMA모형이, 전세가격지수에 대한 적합도는 State-space모형이 더 높게 나타났다.

손정식 외(2002)는 ARIMA모형과 VAR모형을 이용하여 주택매매 및 전세가격 변동률과 지가변동률에 대한 예측을 시도하여 ARIMA모형보다 VAR모형의 예측력이 더 우수하다는 것을 확인하였다.

임성식(2014)은 자기회귀오차모형, ARIMA모형, 개입모형을 이용하여 국민은행에서 발표한 주택가격지수에 대한 예측을 시도하여 개입모형, ARIMA모형, 자기회귀오차모형 순으로 예측력이 우수하다는 것을 확인하였다.

김성환 외(2016)는 기존의 VAR모형 및 VEC모형의 한계로 지적되고 있는 '차원의 저주(curse of dimensionality)'를 극복하기 위해 베이지언 개념을 추가하고, 공간적 영향까지 고려하며 변수간 가중치를 상관계수로 투입하는 방식을 반영한 모형을 이용하여 아파트실거래가지수에 대한 예측을 시도하여, Correlate RVEC의 활용가능성을 확인하였다.

남경두(1997)는 신경망(Neural Network)과 더미변수를 가지는 시계열회귀분석모형의 예측력을 비교 분석하여 신경망의 예측력이 더 우수하다는 것을 확인하였다.

1) 입력값과 가중치( $W$ )를 곱한 값을 합하고 바이어스( $B$ )를 더한 값이 노드의 총 입력이 되는데 이에 대한 출력 값은 활성화 함수의 값에 의해 결정된다. 활성화 함수로 주로 이용되던  $f(u) = \frac{1}{1 + e^{-u}}$  형태의 시그모이드(sigmoid)함수는 기울기 소실(vanishing) 또는

폭발(exploding) 문제를 야기할 수 있기 때문에 최근에는 학습결과와 속도에서 시그모이드 함수보다는 우수한 평가를 받고 있는  $f(u) = \max(u, 0)$  형태의 렐루함수(Rectifier leaner function, Relu function)가 주로 이용되고 있다.

2) 역전파알고리즘은 일차편도함수를 효율적으로 구하는 방법으로 신경망을 학습시키는 기본 알고리즘이다. 학습은 훈련데이터로부터 계산되는 오차함수를 최소화하여 신경망의 가중치와 바이어스를 최소화하는 과정인데, 이는 각 층의 결합 가중치와 바이어스에 대한 편미분을 계산하는 과정이며, 역전파알고리즘에 대한 자세한 설명은 안성만(2016)의 연구를 참고하기 바란다.

정원구.이상엽(2007)은 인공신경망을 이용하여 공동주택가격지수를 예측하였으며, 1개의 입력층, 2개의 은닉층, 1개의 출력층으로 구성하였고, 입력 뉴런은 거시경제변수와 공동주택가격지수로서 총 60개이며, 은닉층1은 52개, 은닉층2는 50개의 뉴런으로 구성되고, 출력층은 3개의 출력뉴런으로 구성하였다.

이형욱.이호병(2009)는 ARIMA모형과 인공신경망 모형의 주택가격지수 예측력을 비교.분석하여 ARIMA모형과 인공신경망 모형의 예측력 간에는 통계적으로 유의미한 차이는 없지만 인공신경망 모형의 예측력이 다소 우수한 것을 확인하였다.

민성욱(2017)은 국내 연구 중에서 딥 러닝 방법을 부동산가격지수 예측에 최초로 적용하였다. 인공신경망과 DNN을 이용하여 지가지수, 아파트전세가격지수, 경제심리지수 등 65개 거시경제변수를 투입변수로 2016년 3월부터 2018년 2월까지 서울아파트실거래가지수를 예측을 시도하였다. 먼저 선형회귀모형, SVM, 랜덤포레스트, 인공신경망을 이용한 단일시계열분석을 통해 인공신경망의 예측력이 가장 우수하다는 것을 확인하였으며, 인공신경망을 이용하여 아파트실거래가지수를 예측하기 위한 투입변수 65개의 값을 예측하였다. 그리고 인공신경망과 DNN의 은닉층 개수, 노드 개수, 학습 횟수 등을 변화시키면서 예측력을 비교하여 DNN보다 은닉층이 2개로 구성된 인공신경망인 다층퍼셉트론의 예측력이 가장 우수하다는 것을 확인하였다. 그리고 선택된 인공신경망 모형에 예측된 투입변수를 적용하여 아파트실거래가지수를 예측하였다.

부동산가격지수 예측과 관련된 선행연구는 ARIMA 모형과 VAR모형의 예측력을 비교하는 연구가 주를 이루고 있는 가운데, ARIMA모형과 VAR모형의 한계점을 보완하여 개선된 방법론을 제시하는 연구 그리고 인공신경망과 같은 기존의 시계열분석 모형 이외의 모형을 적용한 연구로 구분할 수 있다. 또한 경제선행지수모형, 회귀분석모형, 연립방정식모형 등도 부동산 가격지수 예측에 활용되고 있다. 대체적인 연구결과는 ARIMA모형보다는 VAR모형의 예측력이 우수하며, 개입모형이나 Correlate RVEC가 ARIMA

모형이나 VEC 또는 VAR모형보다 개선된 연구 결과를 보고하고 있다. 인공신경망은 ARIMA모형, 선형회귀모형과 같은 전통적인 모형보다 예측력이 우수하며, 일부 연구에서는 SVM, 랜덤 포레스트와 같은 머신러닝방법보다 예측력이 뛰어난 것으로 보고되고 있다.

추가적으로 딥 러닝 방법을 이용하여 시계열데이터를 분석한 선행연구를 살펴보면 다음과 같다. 하지훈 외(2016)는 DBN과 기존 인공신경망을 이용하여 강수 예측을 하였으며, DBN의 예측력이 더 우수하다는 것을 확인하였다.

송사광 외(2016)는 순환신경망(RNN)의 일종인 LSTM을 이용하여 날씨변화에 따른 소상공인의 매출을 예측하였다.

이우식.전희주(2016)는 GARCH모형과 인공신경망 및 딥러닝을 결합한 MLP-GARCH모형 및 DL-GARCH모형을 이용하여 중국 역내.외 위안화의 변동성을 예측하여, DL-GARCH모형의 예측력이 더 우수하다는 것을 확인하였다.

이우식(2017)은 기술적 분석(technical analysis) 지표와 딥 러닝의 결합 모형, 의사결정나무모형, SVM의 코스피지수에 대한 예측력을 비교하여 세 모형의 예측력은 비슷하다는 결과를 확인하였다.

Chakraborty et al.(1992)는 다변량 시계열분석을 위한 신경망 접근법을 제안하였고, 신경망 접근법이 ARMA모형보다 더 좋은 예측력을 보이고 있는 것을 확인하였다.

Kuremoto et al.(2014)은 시계열자료인 CATS benchmark data를 2개의 Restricted Boltzmann Machines(RBMs)로 구성된 3-layer DBN을 통해 분석하였다. 분석결과 인공신경망, ARIMA모형, Self-organizing Fuzzy Neural Network(SOFNN)와 비교했을 때 DBN모형의 예측력이 더 우수한 것을 확인하였다.

이상의 딥 러닝 방법을 부동산가격지수 외 시계열데이터 분석에 이용한 선행연구를 정리하면 DBN, RNN 등 딥러닝 계열의 방법이 기존의 인공신경망이나 ARIMA모형과 같은 시계열분석기법보다 더 좋은 예측력을 보여주고 있는 것을 확인할 수 있다. 특히 RNN은 시계열데이터의 분석에 뛰어

난 성능을 보이는 것으로 나타났다(최승현·도명식, 2016: 7).

### 3. 선행연구와의 차별성

최근 딥러닝 방법을 이용한 연구가 다양한 분야에서 이루어지고 있으나, 아직까지 국내에서 딥러닝 방법을 부동산가격지수 예측에 적용한 연구는 민성욱(2017)의 연구 외에 확인되지 않고 있다. 본 연구는 첫째, DNN모델 외에 추가적으로 시계열자료와 같은 연속형 데이터 분석에 적합한 LSTM모델을 적용하였다는 점, 둘째, 투입자료의 종류 및 형태, 모형의 초기화 방법, 가중치에 대한 최적화(optimizer) 방법 등에서 민성욱(2017)의 연구와 차별성을 갖는다.<sup>3)</sup>

딥러닝 방법이 인공지능망에서 발전된 방법인 점을 고려하여, 기존 인공지능망을 이용한 부동산 가격 지수 예측과 관련된 국내 선행연구와의 차이점을 검토하면 다음과 같다. 첫째, 기존 연구는 은닉층(hidden layer)이 2개였던 정원구·이상엽(2007)의 연구를 제외하고 대부분 은닉층이 1개에 불과하였으나, 본 연구는 다수의 은닉층을 적용한 DNN모형과 LSTM모형을 적용하였다는 점, 둘째, 본 연구는 기존 인공지능망이 깊은 은닉층에서의 학습이 불가능했던 문제점을 극복하기 위해 고안된 기법을 적용하였다는 점, 셋째, 이형욱·이호병(2009)은 입력 변수로 부동산 가격 지수 외에 다양한 거시경제변수를 활용한 인공지능망모형과 단일 시계열모형인 ARIMA모형을 비교하였으나, 본 연구는 부동산 가격 지수만을 투입변수로 하여 동일한 조건하에서 딥러닝 모형과 ARIMA모형의 예측력을 비교하였다는 점에서 차이점을 보이고 있다.<sup>4)</sup>

## III. 분석모형, 분석자료 및 분석방법

### 1. 분석모형

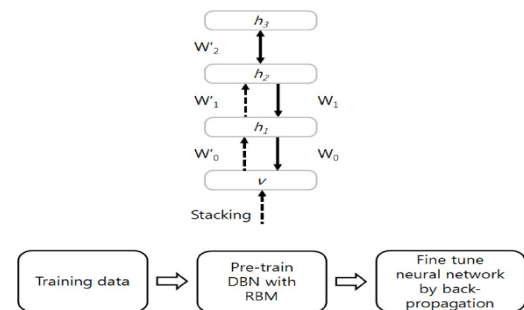
#### 1) 심층 신경망(Deep Neural Networks)

심층 신경망은 다수의 은닉층을 가지고 있는 인공지능망으로서 Hinton et al.(2006)이 제시한 DBN, Vincent et al.(2008)이 제시한 디노이징 오토인코더(denoising autoencoder) 등을 포함한다.

Hinton et al.(2006)이 제시한 DBN은 <그림 2>와 같이 역전파 알고리즘과 RBM이 결합되어 있는 구조이다(하지훈 외, 2016: 94).

학습 자료가 입력되면 먼저 RBM을 통해 비지도 학습(unsupervised learning)으로 가중치 값을 사전학습하여 인공지능망의 초기 가중치 값을 사용한다. 초기 가중치 값이 결정되면 인공지능망에서는 지도학습(supervised learning)으로 오류 역전파 알고리즘을 이용한 미세조정(fine tune)을 하게 된다(하지훈 외, 2016: 94).

RBM은 인공지능망 연구의 두 번째 침체를 야기한 다수의 은닉층을 가지고 있는 인공지능망에서 학습이 되지 않는 문제점을 해결할 수 있는 방법인



자료 : 하지훈 외, 2016:94

<그림 2> Deep Belief Network 개념도

3) 인공지능망의 경우 기존 시계열분석과 같이 투입변수의 안정성을 고려해야 하는지에 대해서는 명확하지 않으며 민성욱(2017)의 경우 원자료를 투입변수로 적용하였으나, 본 연구는 시계열데이터에 선형 요인(linear factors)이 존재하면 인공지능망의 학습능력이 감소하기 때문에 선형 요인을 제거한 차분 변수를 적용한 Kuremoto et al.(2014)의 방법과 동일하게 차분된 자료를 이용하였다. 민성욱(2017)의 연구는 RBM을 지원하는 R패키지를 이용하여 DNN을 분석하였기 때문에 DNN의 초기화방법으로 RBM을 적용한 것으로 추정되며, 최적화 방법으로 이론적 검토에서 경사하강법(Gradient Decent Method)외에 언급이 없어 어떤 방법을 적용하였는지 확인하지 못하였다.

4) <표5>와 같이 활성화함수(activation function), 초기화방법(initialization), 최적화방법(optimizer)에서 기존 인공지능망의 한계를 극복할 수 있는 방법을 적용하였다.

데 이는 신경망 가중치의 초기값 설정과 관련된 것이다. 최근에는 He et al.(2015), Glorot and Bengio (2010)에 의해 RBM보다 성능이 뛰어나면서도 사용하기 쉬운 초기값 설정 방법이 제시되고 있다.<sup>5)</sup>

## 2) LSTM(Long Short Term Memory networks)

RNN은 데이터를 입력하면 입력층에서 은닉층을 거쳐 출력까지 순서대로 노드를 한 번씩만 거쳐 진행되는 일반 인공신경망과 달리, 입력에 대한 은닉층의 값을 신경망 내부의 메모리에 기억하여 다음 입력 값에 대한 출력 시 이를 다음 입력 값과 함께 고려하여 연산하기 때문에 입력의 시계열적인 정보를 효과적으로 모델링하는 특징이 있다(이세희, 이지형, 2016: 46).

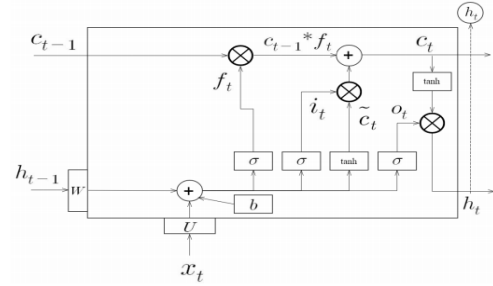
하지만 RNN모델에 의한 예측은 무한한 과거 시점 관측값에 의존하기 때문에 기울기가 소실(vanishing gradient)되거나 기울기가 매우 큰 값(exploding gradient)이 되는 문제를 발생시킬 수 있다(안성만 외, 2017: 75).

Hochreiter and Schmidhuber(1997)에 의해 제안된 LSTM은 RNN의 문제점을 해결하기 위해 제시된 방법으로서, LSTM은 전통적인 RNN구조에서 은닉계층의 유닛들을 LSTM블록으로 대체시킨 형태와 같다(김양훈 외, 2016: 595).

LSTM의 메모리 유닛 하나의 구조는 <그림 3>과 같다. 각각의 LSTM블록 내부는 재귀적 구조를 가지고 있는 기억 소자(Memory Cell)와 입력게이트(Input Gate), 잊기게이트(Forget Gate), 출력게이트(Out Gate)로 이루어져 있다(김양훈 외, 2016: 595).

LSTM 내부에서는 입력게이트, 잊기게이트, 출력게이트를 통해 기억 소자에 어떠한 정보가 반영될지가 결정된다.

LSTM의 첫 번째 단계는 잊기 게이트를 통해 소자 상태(Cell State)에 어떤 정보를 반영할지 결정하는 것이며 식(1)과 같다.  $h_{t-1}$ 과  $x_t$ 를 입력받아 시그모이드 활성화 함수를 통해 0~1 사이의 값을 출력하며



자료 : 김양훈 외, 2016:595

<그림 3> LSTM의 구조도

1에 가까울수록 많은 정보를 반영하고 0에 가까울수록 해당 정보의 반영을 적게한다.

두 번째 단계는 입력게이트를 통해 어떤 새로운 정보가 소자 상태에 저장될 것인지를 결정하며 이 단계는 두 부분으로 구성된다. 먼저 입력 게이트 층이라고 불리는 시그모이드 층은 어떤 값들을 갱신(update)할 것인지 결정하며 식(2)와 같다. 다음 탄젠트에이치(tanh) 층은 소자 상태에 더해질 새로운 후보 값들의 벡터(vector)인  $\tilde{C}_t$ 를 만들어내며 식(3)과 같다. 다음은 잊기게이트와 입력게이트에 의해 소자 상태가  $C_t$ 로 갱신되게 되는데 이는 식(4)와 같다.

세 번째 단계는 필터링(filtering)된 소자상태를 기반으로 출력값인  $h_t$ 를 결정하는 것으로 식(5)에서 생성된 시그모이드 층을 통한 출력값을 식(6)에서 소자상태( $C_t$ )가 적용된 탄젠트에이치 값과의 곱을 통해 산출하게 된다.

$$f_t = \sigma(W_f [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

5) He et al.(2015)와 Glorot and Bengio(2010)이 제시한 초기값 설정방법은 '노드의 입력값의 숫자(fan\_in)와 출력값의 숫자(fan\_out)'를 '입력값의 숫자(fan\_in) 또는 입력값의 숫자를 2로 나눈값(fan\_in/2)'으로 나눠서 산출된 값의 범위에서 랜덤하게 결정하는 방식이다.

$\sigma, \tanh$ 는 비선형활성화함수,  $x_t$ 는 입력값,  $h_t$ 는  $t$ 시점의 은닉변수,  $o_t$ 는  $t$ 시점의 출력값,  $b$ 는 바이어스(bias),  $U$ 와  $W$ 는 가중치를 의미한다.

## 2. 분석자료

2006년 1월부터 2016년 12월까지 발표된 서울지역의 아파트 매매실거래가격지수(APT\_T), 아파트 매매가격지수(APT\_M), 아파트 전세가격지수(APT\_J), 지가지수(LAND)를 분석대상으로 한다.<sup>6)</sup>

아파트 매매실거래가격지수는 지수산정기간 중 거래신고가 2번 이상 있는 동일 주택(아파트)을 이용한 반복매매모형을 통해 가격변동률과 거래량으로 지수 산출한다.

아파트 매매가격지수와 아파트 전세가격지수는 전국의 재고 아파트 중 공공임대주택이나 사원 임대주택 등과 같이 거래가 이루어지지 않는 임대아파트를 제외한 조사모집단에서 추출한 표본의 가격을 한국감정원에서 매월 조사하여 이를 지수화 한 것으로 제본스 지수(Jevons Index) 방법론에 의해 지수가 산정된다.

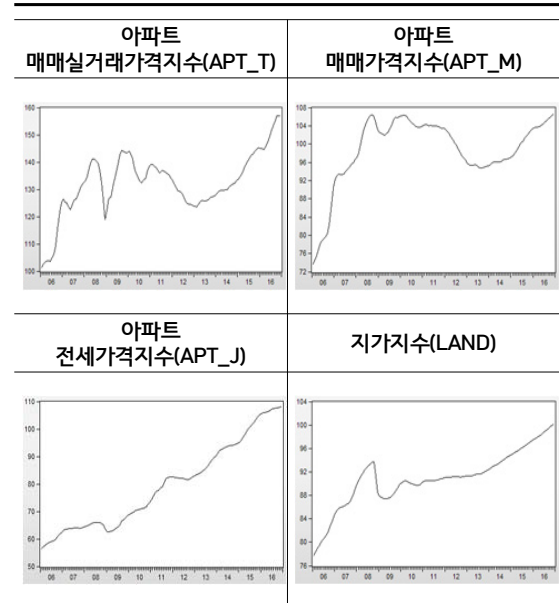
지가지수는 개별공시지가 대상필지 전체 중 실제 이용 상황을 기준으로 ‘도로’, ‘하천’ 등 특수한 필지를 제외한 조사모집단에서 추출한 표본의 가격을 한국감정원에서 매월 조사하여 라스파이레스 수정산식에 의해 산정한다.

각 지수는 조사방법과 지수산정방식에서 차이를 보이고 있으며, 가장 큰 차이점은 아파트 매매 실거래가격지수는 시장 내에서 실제 거래된 자료를 이용하여 산정되는 반면 아파트 매매 및 전세가격지수와 지가지수는 조사자들이 매월 직접 조사한 가격을 기준으로 산정된다는 것이다.

<표 1>는 부동산가격지수의 기초통계량이며 <그림 4>는 2006년 1월부터 2016년 12월까지의 부동산 가격지수의 변화 양상을 보여주고 있다. 전반적으로 우상향하는 추세를 보이고 있으며, 아파트 전세가격지수를 제외한 나머지 지수들은 2008년 금융위기가 발생한 시점에 급락하는 모습을 보여주고

<표 1> 기초통계량

구분	평균	중위수	최대값	최소값	표준편차
APT_T	132.3	132.7	157.6	100.0	11.2
APT_M	98.7	101.1	106.9	73.6	7.4
APT_J	79.3	78.4	108.3	56.3	15.5
LAND	90.9	91.0	100.2	77.4	4.7



주 : 2006년 1월부터 2016년 12월까지의 자료(www.r-one.co.kr)를 그래프로 표시함.

<그림 4> 2006년 1월부터 2016년 12월까지  
부동산 가격지수

있다. 특히 아파트 매매실거래가격지수는 다른 지수들과 비교했을 때 상대적으로 큰 변동성을 보여주고 있다.

## 3. 분석방법

본 연구는 부동산가격지수에 대한 딥 러닝 방법의 예측력을 확인하고 기존 시계열분석 방법과 예측력을 비교하기 위한 것이다. 먼저 기존 시계열분석 방법인 ARIMA모형을 통해 부동산가격지수에 대한 예측을 실시하였고, 딥 러닝 방법을 이용한 부동산 가

6) 한국감정원 부동산통계정보(www.r-one.co.kr)의 자료를 활용하였다.

&lt;표 2&gt; 단위근 검정 결과

구분		수준변수			차분변수		
		None	Con.t	Con.t+trend	None	Con.t	Con.t+trend
APT_T	ADF	0.834	-2.538	-2.959	-4.648***	-4.765***	-4.749***
	PP	1.423	-2.012	-2.233	-4.723***	-4.836***	-4.821***
APT_M	ADF	1.202	-3.302**	-2.999	-4.731***	-4.955***	-5.156***
	PP	1.485	-3.661***	-3.131	-2.889***	-3.072**	-3.175*
APT_J	ADF	2.359	0.172	-2.567	-2.758***	-3.689***	-3.752**
	PP	5.183	0.716	-1.612	-2.773***	-3.917***	-3.989**
Land	ADF	1.785	-1.566	-3.225*	-4.003***	-4.452***	-4.435***
	PP	2.379	-2.024	-2.815	-4.046***	-4.377***	-4.369***

주 : 1) \*\*\*는 1%, \*\* 및 \*는 5% 및 10% 수준에서 유의함을 의미함.

2) Con.t는 상수항만, Con.t+trend는 상수항과 추세를 동시에 고려한 것임.

&lt;표 3&gt; LM-test 결과

	Model	F-statistic	Prob.Chi-Square	AIC
APT_T	ARIMA(1,1,0)	5.511990	0.030**	3.410557
	ARIMA(2,1,0)	2.237986	0.181	3.389627
	ARIMA(1,1,1)	1.524406	0.502	3.365525
APT_M	ARIMA(1,1,0)	22.59243	0.000***	1.128608
	ARIMA(2,1,0)	0.942579	0.819	0.960838
	ARIMA(1,1,1)	1.006522	1.000	0.960763
APT_J	ARIMA(1,1,0)	3.582162	0.059*	0.197634
	ARIMA(2,1,0)	0.123245	0.879	0.190802
	ARIMA(1,1,1)	0.024728	0.873	0.183400
LAND	ARIMA(1,1,0)	12.01396	0.002***	0.423436
	ARIMA(2,1,0)	7.251842	0.001***	0.327989
	ARIMA(1,1,1)	0.663493	1.000	0.205895

주 : \*\*\*는 1%, \*\* 및 \*는 5% 및 10% 수준에서 유의함을 의미함.

격지수에 대한 예측을 실시하였다. 그리고 각 모형에 따라 산출된 RMSE값과 그래프를 통해 예측력을 비교 분석한다. 실증분석시 ARIMA모형은 이뷰즈(Eviews) 패키지, 딥 러닝 모형은 파이썬(Python) 기반의 케라스(Keras) 라이브러리를 이용하였다.

## IV. 실증분석

### 1. ARIMA 모형

#### 1) 단위근 검정

분석에 앞서 시계열의 안정성 여부를 확인하기 위해 ADF(Augmented Dickey-Fuller) 및 PP(Phillips-Perron) 검정 방법을 통해 단위근 검정을 실시하였다. 분석결과 원시계열 자료에서는 단위근이 존재한다는 귀무가설을 대체로 기각하지 못하였으나, 1차 차분 후 단위근 검정을 재실시한 결과 단위근이 존재한다는 귀무가설을 모두 기각하였으며, 분석결과와 <표 2>와 같다. 이하에서는 1차 차분된 자료를 기준으로 분석을 실시하며, 딥 러닝 방법에서도 1차 차분된 자료를 적용하였다.<sup>7)</sup>

7) 원자료의 경우 X-12 ARIMA법으로 계절조정하였다.



## 2) 모형 추정 및 진단

부동산가격지수 별로 자기상관함수 및 편자기상관함수를 추정한 결과, ARIMA(1,1,0), ARIMA(2,1,0)이 식별되었으며 추가적으로 ARIMA(1,1,1)을 포함하여 모수를 추정하였다. 추정후 잔차의 계열상관성 여부를 LM-Test를 통해 검정하였으며 그 결과는 <표 3>과 같다.

지가지수(LAND)는 계열상관성이 없다는 귀무가설을 기각하지 못한 ARIMA(1,1,1)을 최종 모형으로 선정하였다. 아파트 매매실거래가격지수(APT\_T), 아파트 매매가격지수(APT\_M), 아파트 전세가격지수(APT\_J)는 ARIMA(2,1,0)과 ARIMA(1,1,1)이 계열상관성이 없다는 귀무가설을 기각하지 못하였으며, 모수절약의 원칙으로는 어떤 모형을 선정해야 하는지 판단할 수 없는바, AIC(Akaike Information

Criterion)값이 상대적으로 낮은 ARIMA(1,1,1)을 최종 모형으로 선정하여, 모든 부동산 가격지수를 ARIMA(1,1,1)을 통해 추정하게 되었다.

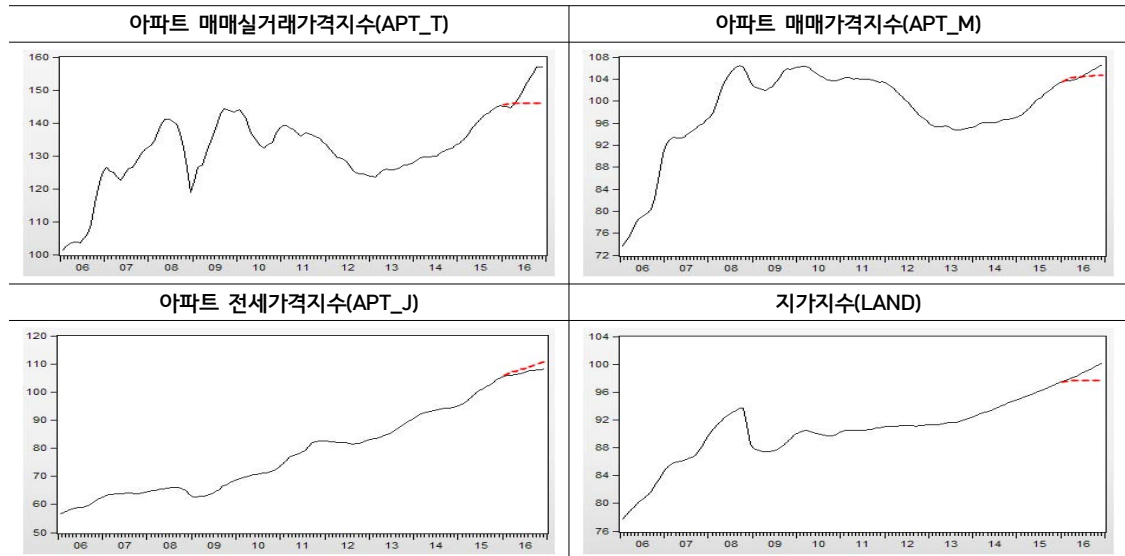
## 3) 예측력

2006년 1월부터 2015년 12월까지의 데이터를 학습 데이터로 2016년 1월부터 2016년 12월까지의 데이터를 검정 데이터로 분석한 결과는 <표 4> 및 <그림 5>와 같다.

RMSE값을 보면 아파트 매매가격지수(APT\_M)에 대한 예측력이 상대적으로 가장 우수하며, 아파트 전세가격지수(APT\_J)와 지가지수(LAND)에 대한 예측력은 비슷하다. 그리고 아파트 매매실거래가격지수(APT\_T)에 대한 예측력은 상대적으로 부족한 것을 확인할 수 있다.

<표 4> ARIMA 모형 RMSE 값

구분	APT_T	APT_M	APT_J	LAND
RMSE	6.746864	0.906315	1.636948	1.433592



주 : 실선은 실제값, 점선은 예측값을 의미함.

<그림 5> ARIMA 모형 예측 결과

## 2. 딥 러닝 모형

### 1) 모형 설계

딥 러닝 모형을 설계하고 최적화하기 위해서는 투입 변수(input), 은닉층(hidden layer)의 수, 은닉층 내 노드(node)의 개수, 활성화 함수(activation function)의 결정, 가중치에 대한 최적화 방법(optimizer)의 결정, 테스트 회수(epochs), 배치(batch), 과적합(overfitting) 방지를 위한 드랍아웃(dropout) 등을 결정해야하며, 이를 초모수(hyperparameter)에 대한 최적화(optimization)라고 한다.<sup>8)</sup>

딥 러닝 모형의 초모수를 최적화하기 위해서 메뉴얼 서치(manual search), 그리드 서치(grid search), 랜덤 서치(random search) 등과 같은 다양한 방법이 제시되고 있으며, Bergstra and Bego(2012)는 랜덤 서치가 가장 효율적이라고 하였으며, Kuremoto et al. (2014)은 DBN모형을 최적화하는데 있어서 Kenedy and Eberhart(1995)가 고안한 입자군집최적화(particle swarm optimization, PSO) 방법을 이용하였고, 하지훈 외(2016)는 경험적 방법을 통해 초모수를 결정하였다. 메뉴얼 서치, 그리드 서치, 랜덤 서치, 입자군집최적화 알고리즘 등은 연구자가 지정한 범위 안에서 컴퓨터가 최적의 값을 갖는 모형을 찾는 방식이라고 할 수 있으며, 이는 결국 시행착오(trial and error)를 거쳐 이루어진다고 할 수 있다.

본 연구에서는 <표 5>와 같이 초모수를 결정하였다. 활성화 함수는 DNN의 경우 렐루 함수(relu function), LSTM의 경우 쌍곡탄젠트 함수(hyperbolic tangent function, tanh function), 최적화(optimizer)방법은 아담(ADAM)알고리즘, 초기화(initialization)방법은 He et al.(2015)이 제시한 방법을 적용하였다. 테스트 회수(epochs)는 100회, 데이터가 많지 않음을 감안하

<표 5> 초모수(hyper-parameter)의 설정

Parameter	DNN	LSTM	비고
Input Variables	1~3	1~3	<표 9> 참조
Hidden Layer	1~3	1~3	<표 9> 참조
Hidden Node	5, 15, 25	5, 15, 25	<표 9> 참조
Epochs	100	100	
Batch Size	1	1	
Activation Function	Relu function	tanh function	
Optimizer	ADAM Algorithm	ADAM Algorithm	
Initialization	He et al.	He et al.	
Dropout	0 or 0.1	0 or 0.1	<표 9> 참조

여 배치(batch) 사이즈는 1로 결정하였고, 전체 132개 데이터 중 학습데이터는 90%, 검정데이터는 10%로 설정하였다. 투입변수는 1~3개, 은닉층의 수는 1~3개, 은닉층 내 노드의 개수는 5, 15, 25개, 드랍아웃비율은 0 또는 10%로 변화시키면서 종속변수별 총 468개의 모형 중 RMSE값을 기준으로 예측력이 가장 우수한 모형을 최종모형으로 결정하였다.<sup>9)</sup>

### 2) 모형 결정 및 예측력 비교

<표 6>은 종속변수별로 추정한 468개 모형들의 RMSE값의 기초통계량을 보여주고 있다. <표 7>은 종속변수별 RMSE 최소값을 정리한 것이며, <표 8>은 RMSE값이 최소가 되는 경우의 투입변수, 은닉층의 수, 은닉층 내 노드의 개수, 드랍아웃을 나타내고 있다.<sup>10)</sup>

시계열분석을 효과적으로 모델링 할 수 있는 LSTM모형보다 DNN모형의 예측력이 더 우수한

8) 최적화(optimizer)는 각 노드의 비용함수를 최소화 시키는 가중치를 찾아가는 방법이며, 배치(batch)는 대규모 신경망의 효율적 있는 계산을 위해 자료를 집합으로 묶는 것이고, 드랍아웃(dropout)은 과적합을 방지하기 위해 입력값 중 일부를 제외하고 학습을 하는 것이다.

9) DNN모형에서는 학습이 빠르고 결과도 더 좋은 것으로 평가받고 있어 현재 가장 많이 사용되고 있는 렐루함수(relu function), LSTM모형에서는 일반적으로 이용되는 쌍곡탄젠트 함수를 활성화함수로 적용하였다. 초기값 설정은 간단하면서도 성능이 우수한 He et al.(2015)이 제시한 방법을 적용하였으며, 신경망의 가중치와 바이어스를 최소화를 위한 알고리즘은 성능이 가장 우수한 것으로 나타나고 있는 아담(ADAM) 알고리즘을 적용하였다. 활성화함수인 렐루함수는 Glorot et al.(2011)과 Nair and Hinton(2010), 초기값 설정은 He et al.(2015)와 <https://keras.io>(케라스 홈페이지), 최적화 알고리즘은 Kingma and Ba(2015)를 참조하기 바란다.

10) <표 6>은 차분자료를 기준으로 산출된 종속변수별 468개 모형(DNN:234개, LSTM:234개)의 RMSE값의 기초통계량이다.

&lt;표 6&gt; 모형별 RMSE값의 기초통계량

구분	DNN				LSTM			
	평균	표준편차	최소값	최대값	평균	표준편차	최소값	최대값
APT_T	0.698979	0.852782	0.537154	0.058475	0.711395	1.162883	0.611347	0.070954
APT_M	0.189172	0.304709	0.130779	0.040340	0.173732	0.288800	0.144955	0.023566
APT_J	0.196011	0.250092	0.145866	0.016687	0.218449	0.298230	0.196079	0.017458
LAND	0.078542	0.157461	0.053410	0.017151	0.100165	0.265687	0.056822	0.032517

&lt;표 7&gt; 종속변수별 RMSE 최소값

구분		APT_T	APT_M	APT_J	LAND
DNN	원자료	0.838981	0.198930	0.112699	0.070614
	차분자료	0.537154	0.130779	0.145866	0.053410
LSTM	원자료	0.859259	0.134363	0.195866	0.052440
	차분자료	0.611347	0.144955	0.196079	0.056822

&lt;표 8&gt; 최종모형

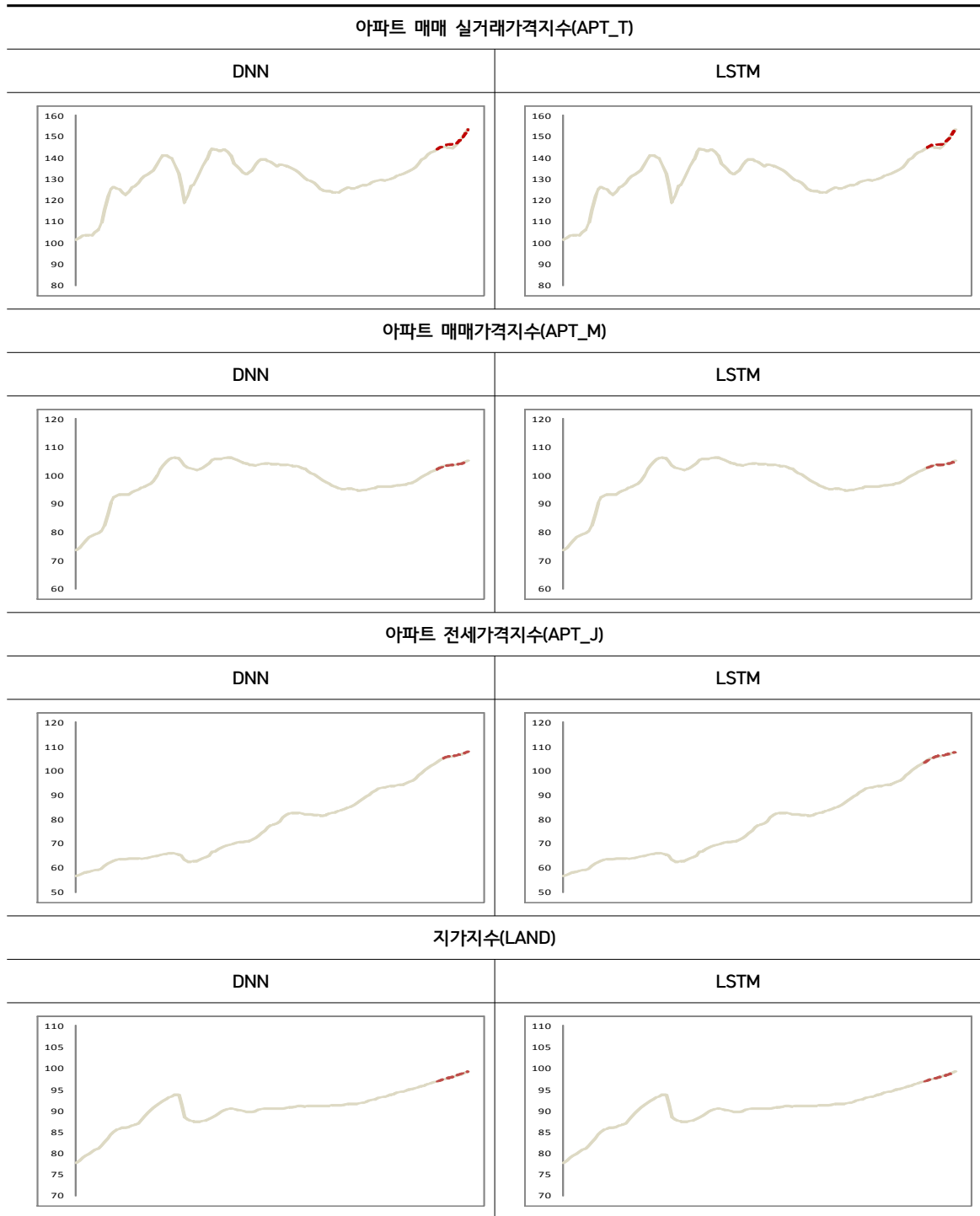
구분		APT_T		APT_M		APT_J		LAND	
		DNN	LSTM	DNN	LSTM	DNN	LSTM	DNN	LSTM
Input		2	1	1	2	3	1	1	1
Hidden Layer	1Layer Node	5	5	25	25	25	15	15	5
	2Layer Node	15	15	25	25	25	0	5	0
	3Layer Node	15	25	15	15	25	0	25	0
Dropout		0	0	0.1	0.1	0.1	0	0.1	0

것으로 나타났다. 종속변수별로 보면 차분자료의 경우 DNN모형과 LSTM모형 모두 RMSE값이 ‘아파트 매매실거래가격지수(APT\_T) > 아파트 전세가격지수(APT\_J) > 아파트 매매가격지수(APT\_M) > 지가지수(LAND)’의 순으로 지가지수(LAND)에 대한 예측력이 가장 우수하며, 상대적으로 아파트 매매실거래가격지수(APT\_T)에 대한 예측력이 부족한 것으로 나타났다. 원자료를 기준으로 보면 RMSE값이 DNN모형의 경우 ‘아파트 매매실거래가격지수(APT\_T) > 아파트 매매가격지수(APT\_M) > 아파트 전세가격지수(APT\_J) > 지가지수(LAND)’이며, LSTM모형의 경우 ‘아파트 매매실거래가격지수(APT\_T) > 아파트 전세가격지수(APT\_J) > 아

파트 매매가격지수(APT\_M) > 지가지수(LAND)’의 순으로 나타나, 아파트 매매가격지수(APT\_M)와 아파트 전세가격지수(APT\_J)에 대한 예측력에서 다소 차이를 보이는 것으로 나타났다. 다만, 이를 그래프로 나타낸 <그림 6>을 보면 모든 모형의 예측값이 실제값과 거의 일치하고 있어 모형별 예측력의 차이를 시각적으로 확인하기는 어려워보인다. RMSE값 역시 모두 1미만으로 그 격차는 미미하다고 할 수 있다.<sup>11)</sup>

대부분의 모형에서 은닉층의 개수가 3개인 경우 RMSE값이 최소가 되었으며, 아파트전세가격지수와 지가지수에 대한 LSTM모형은 은닉층의 개수가 1개인 경우 RMSE값이 최소가 되었다.

11) 딥 러닝 모형의 경우 차분자료를 투입변수로 이용하였기 때문에 예측결과가 차분자료로 산출되어 이를 다시 원자료로 변환한 후 RMSE값을 계산하였다. 132개 데이터를 기준으로 10%를 검정데이터(약13개월)로 적용하였으나, 실제 산출된 검정데이터는 DNN(APT\_J)는 2015.12~2016.8까지 9개월, DNN(APT\_T)와 LSTM(APT\_J)은 2015.11~2016.8까지 10개월, 나머지는 2015.10~2016.8까지 11개월이다.



주 : 실선은 실제값, 점선은 예측값을 의미함.

<그림 6> 딥 러닝 모형 예측 결과

아파트 매매실거래가격지수에 대한 DNN모형과 아파트 매매가격지수에 대한 LSTM모형은 투입변수가 2개인 경우, 아파트전세가격지수에 대한 DNN모형의 경우 투입변수가 3개인 경우를 제외하고 모든 모형이 투입변수가 1개인 경우 RMSE 값이 최소가 되었다.

아파트 매매실거래가격지수에 대한 DNN모형 및 LSTM모형, 아파트전세가격지수와 지가지수에 대한 LSTM모형을 제외하고 드랍아웃이 0.1인 경우 RMSE값이 최소가 되었다.

### 3. 검토

시계열분석모형의 RMSE값인 <표 4>와 딥러닝 모형의 RMSE값인 <표 7>을 비교해보면 딥러닝 모형의 RMSE값이 더 낮아, 딥 러닝 모형의 예측력이 더 우수하다는 것을 알 수 있다. 이러한 결과는 Kuremoto et al.(2014) 등 선행연구와 동일한 결과이다. ARIMA모형과 딥 러닝 모형의 예측 결과를 <그림 5>와 <그림 6>을 통해 비교해보면 예측력의 차이는 더욱 명확하게 나타난다. ARIMA모형의 경우 실제 데이터와의 차이가 확실히 나타나고 있지만, 딥 러닝 모형의 경우 실제 데이터와 거의 일치하고 있어 차이를 확인하기 어렵다. ARIMA모형은 기본적으로 선형모형을 가정하고 있기 때문에 비선형 구조를 가지고 있는 시계열분석에서는 한계를 가지고 있는 반면, 딥 러닝모형은 '유연한 비선형 모델링 능력을 가지고 있는 인공지능망(Zhang(2003: 160))'에서 발전하였기 때문에 ARIMA모형보다 높은 예측력을 보이는 것으로 이해된다. 한편, 기존 인공지능망 연구와 비교해보면 분석기간이 상이하여 직접 비교는 어렵지만 RMSE값이 기존 연구보다 상당히 낮은 수준을 유지하고 있어 인공지능망보다 딥 러닝의 예측력이 우수하다고 볼 수 있다.<sup>12)</sup>

민성욱(2017)은 은닉층이 3개이상인 경우보다 2개인 경우 예측력이 가장 우수하다는 것을 보고하고 있으나 본 연구에서는 아파트 전세가격지수와 지가지수를 LSTM으로 추정된 경우를 제외하고 모두 은

닉층이 3개인 경우 예측력이 가장 좋은 것으로 나타났다. 이는 분석 데이터의 차이, 노드의 개수, 활성화 함수, 최적화방법, 드랍아웃 비율과 같은 모형 설계 방식의 차이로 인한 것으로 판단된다.

ARIMA모형에서는 아파트 매매가격지수에 대한 예측력이 가장 우수하며, 아파트 매매실거래가격지수에 대한 예측력이 상대적으로 부족한 것으로 나타났다. 딥 러닝 모형에서는 지가지수에 대한 예측력이 가장 우수하며, 아파트 매매실거래가격지수에 대한 예측력이 가장 부족한 것으로 나타나, 딥러닝 분석모형과 시계열 분석모형 모두 아파트 매매실거래가격지수에 대한 예측력이 가장 부족한 것으로 나타났다. 지수별 예측력의 차이는 지수가 가지고 있는 성격으로 인한 것이 아닌지 의심된다. 이용만(2008)은 평가가격에 기초한 지수는 평활화(smoothing) 현상을 보인다는 문제가 있다고 지적하면서, 평활화는 평가가격의 변동성이 실제 시장가격의 변동성보다 작고, 시장가격의 움직임보다 늦게 움직이는 현상을 말하며 이는 평가자의 평가가 과거의 평가가격에 부분적으로 의존하기 때문에 나타난다고 하였다. 본 연구에 적용된 아파트 매매가격지수와 아파트 전세가격지수 그리고 지가지수는 평가가격에 기초한 지수(appraisal-based index)로서 이러한 평활화 현상으로 인한 미미한 변동성, 그리고 과거값에 의존하는 경향을 가지고 있다는 점에서 분석모형이 과거값에 대한 학습을 기반으로 하기 때문에 과거 값에 대한 의존 경향이 상대적으로 낮은 아파트 매매실거래가격지수보다 상대적으로 예측이 용이한 것으로 이해된다. 반면, 아파트 매매실거래가격지수는 실거래가격에 기초한 지수(transaction-based index)로서 다른 지수와 비교했을 때 변동성이 심하고 과거값에 대한 의존도가 낮기 때문에 분석모형에 의한 예측력은 상대적으로 부족한 것으로 이해된다.

## V. 결 론

본 연구는 복잡한 문제해결 방법으로 주목받고 있는 딥 러닝 방법을 부동산가격지수 예측에 적

12) 정원규, 이상엽(2007)이 인공지능망을 이용하여 2005년 10월부터 2006년 9월까지 예측한 서울 공동주택가격지수에 대한 RMSE값은 2.544이다.

용해보고, 기존의 시계열분석 방법과의 비교를 통해 부동산 시장 예측의 새로운 방법으로서 활용가능성을 확인하였다는데 연구의 의의가 있다.

본 연구 결과를 요약하면 다음과 같다. 첫째, 기존 시계열분석모형인 ARIMA모형보다 딥 러닝 모형의 예측력이 더 우수한 것으로 나타났다. 이는 딥 러닝 모형이 유연한 비선형 모델링이 가능하기 때문에 선형모형을 가정하는 ARIMA모형보다 더 좋은 예측력을 보이는 것으로 이해된다. 둘째, 시계열분석을 효과적으로 모델링 할 수 있는 LSTM모형보다 DNN모형의 예측력이 우수한 것으로 나타났다. 다만, 예측력을 판단하는 RMSE값이 양자 모두 1 미만이고, 특히 그래프를 통해서는 차이를 확인하기 어려울 정도로서 예측력의 차이는 미미한 것으로 판단된다. 셋째, 부동산 가격지수 중 아파트 매매실거래가격지수에 대한 예측력이 상대적으로 가장 부족한 것으로 나타났다. 이는 다른 지수에 비해 상대적으로 변동성이 높고 과거값에 대한 의존도가 낮기 때문인 것으로 이해되며, 결국 지수 산정방식 차이로 인해 발생하는 것으로 판단된다.

미래 부동산 가격의 하락 또는 상승 여부에 대한 예측은 정부·기업·개인의 중요한 관심사항이다. 지금까지는 주로 ARIMA모형, VAR모형 및 VECM모형 등을 활용하여 부동산 가격지수를 예측하였으나, 딥 러닝 방법이 기존의 시계열분석 방법 중 ARIMA모형보다 우수한 예측력을 가지고 있다는 것을 확인함에 따라 딥 러닝 방법은 기존의 시계열분석 방법을 대체·보완하는 역할을 할 수 있을 것으로 기대된다. 또한, 최근 관련기관에서 부동산투자지수에 대한 연구가 활발히 진행되고 있다는 점을 고려해보면, 향후 부동산 가격지수에 대한 예측은 투자자들에게 있어 중요한 관심사항이 될 것이며, 이에 따라 딥 러닝 방법은 투자자들의 의사결정을 위한 분석도구로서의 활용 역시 기대해 볼 수 있을 것이다.<sup>13)</sup>

딥 러닝 방법은 결과 값이 비교적 정확하지만, 투입변수와 출력변수만 확인할 수 있을 뿐 변수들간의 관계, 변수들간의 영향력, 은닉층에서 산출

과정을 확인할 수 없기 때문에 결과가 도출된 이유를 확인할 수 없으며, 이로 인해 모형을 최적화하기 위한 명확한 기준 역시 없다는 한계를 가지고 있다. 이는 인과관계를 생명으로 하는 과학에 있어서 큰 흠결이라고 할 수 있다.

본 연구는 부동산가격지수라는 단일 변수만을 이용한 분석이다. 따라서 동일한 조건하에서 딥 러닝 방법과의 비교를 위해 단변량 시계열모형인 ARIMA모형을 비교 대상으로 선택하였다. 하지만 ARIMA모형이 VAR모형 또는 VECM모형보다 예측력이 낮다는 점을 감안하면, VAR모형 또는 VECM모형과의 비교 없이 딥 러닝 방법의 예측력이 기존 시계열분석 방법보다 우수하다고 단정하기 어려운 측면이 있다. 따라서 향후 거시경제변수를 반영한 딥 러닝 방법과 VAR모형 또는 VECM모형과의 비교를 시도해 볼 필요가 있다. 또한, 딥러닝 방법이 과거 데이터의 학습을 통한 예측이기 때문에 분석기간 및 분석시점이 달라지거나, 장기 예측의 경우 또는 시장의 구조적 변화가 발생하는 경우에는 예측력이 떨어질 가능성이 있는데, 이 경우에도 기존 시계열분석모형보다 딥러닝 방법의 예측력이 뛰어난지 여부에 대해서도 확인해 볼 필요가 있다.

## Ⅱ 참고문헌 Ⅱ

1. 김근용, 1998, “주택가격 예측을 위한 모형설정과 검증”, 국토 197, 국토연구원: 54~61
2. 김성환·김갑성·유예진, 2016, “주택경기 예측 향상을 위한 시계열모형 구축”, 2016년 한국주택학회 상반기 학술대회 발표자료집, 한국주택학회: 33~49
3. 김양훈·황용근·강태관·정교민, 2016, “LSTM 언어모델 기반 한국어 문장 생성”, 한국통신학회논문지 41(5), 한국통신학회: 592~601
4. 남경두, 1997, “부동산 분석을 위한 뉴럴네트워크의 응용”, 상경연구 .22(2), 건국대학교 경제경영연구소: 187~201
5. 민성욱, 2017, “딥러닝을 이용한 주택가격 예측모형 연구”, 강남대학교 박사학위 논문
6. 손정식·김관영·김용순, 2002, “부동산가격 예측모형에 관한 연구”, 주택연구 11(1), 한국주택학회: 49~75
7. 송사광·신성호·선충녕·임형준·이미경·조민희·홍승균

13) 한국감정원 보도자료(2015.12.14)에 따르면 한국감정원과 한국거래소에서는 ‘부동산 금융상품 투자지수’를 공동개발 중에 있다.

- 조민수, 2016, "딥러닝 기반 날씨와 소상공인 매출 연계 빅데이터 분석", 한국기상학회 학술대회 논문집, 한국기상학회: 489-290
8. 안성만, 2016, "딥러닝의 모형과 응용사례", 지능정보연구 22(2), 한국지능정보시스템학회: 127-142
  9. 안성만 · 정여진 · 이재준 · 양지현, 2017, "한국어 음소 단위 LSTM 언어모델을 이용한 문장 생성", 지능정보연구 23(2), 한국지능정보시스템학회: 71-88
  10. 이세희 · 이지형, 2016, "RNN을 이용한 고객 이탈 예측 및 분석", 한국컴퓨터정보학회 학술발표논문집, 24(2), 한국컴퓨터정보학회: 45-48
  11. 이용만 · 이상한, 2008, "국민은행 주택가격지수의 평활화 현상에 관한 연구", 주택연구 16(4), 한국주택학회: 27-47
  12. 이우식, 2017, "딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측", 한국데이터정보과학회지 28(2), 한국데이터정보과학회: 287-295
  13. 이우식 · 전희주, 2016, "딥러닝 분석을 이용한 중국 역내, 외 위안화 변동성 예측", 한국데이터정보과학회지 27(2), 한국데이터정보과학회: 327-335
  14. 이형욱 · 이호병, 2009, "서울시 주택가격지수의 모형별 예측력 비교 분석", 부동산학보 38, 한국부동산학회: 215-235
  15. 임성식, 2014, "주택가격지수 예측모형에 관한 비교연구", 한국데이터정보과학회지 25(1), 한국데이터정보과학회: 65-76
  16. 정원구 · 이상엽, 2007, "인공신경망을 이용한 공동주택 가격지수 예측에 관한 연구 - 서울지역을 중심으로", 주택연구 15(3), 한국주택학회: 39-64
  17. 최근우 · 송기선 · 강요셉, 2016, 딥러닝(Deep Learning)기술의 이해와 연구개발 정책과제, 한국과학기술기획평가원
  18. 최승현 · 도명식, 2016, "딥러닝 기법을 이용한 아스팔트 포장의 공용수명예측", 대한토목학회 학술대회 10, 대한토목학회: 7-8
  19. 하지훈 · 이용희 · 김용혁, 2016, "딥러닝 기법을 이용한 내일강수 예측", 한국지능시스템학회논문지 26(2), 한국지능시스템학회: 93-98
  20. Bergstra, J., Y. Bengio, 2012, "Random Search for Hyper-Parameter Optimization", *Journal of Machine Learning Research* 13, JMLR:281-305
  21. Chakraborty, K., K. Mehrotra, C. K. Mohan, S. Ranka, 1992, "Forecasting the behavior of multivariate time series using neural networks.", *Neural networks* 5(6), Elsevier: 961-970
  22. Glorot, X., Y. Bengio, 2010, "Understanding the difficulty of training deep feedforward neural networks", *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, PMLR 9, 249-256
  23. Glorot, X., A. Bordes, Y. Bengio, 2011, "Deep Sparse Rectifier Neural Networks", In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, 315-323
  24. He, Kaiming, X. Zhang, S. Ren, J. Sun, 2015, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification", *The IEEE International Conference on Computer Vision (ICCV)*, Computer Vision Foundation: 1026-1034
  25. Hinton, G. E., S. Osindero, Y. W. The, 2006 "A Fast Learning Algorithm for Deep Belief Nets", *Neural Computation* 18(7), MIT Press: 1527-1554
  26. Hochreiter, S., J. Schmidhuber, 1997, "Long short-term memory", *Neural Computation* 9(8), MIT Press: 1735-1780
  27. Kaastra, I., M. Boyd, 1996, "Designing a neural network for forecasting financial and economic time series", *Neurocomputing* 10, Elsevier Science B. V.: 215-236
  28. Kennedy, J., R. C. Eberhart, 1995, "Particle swarm optimization", *Proceedings of the IEEE International Conference on Neural Networks*, 1942-1948.
  29. Kingma, D. P., J. L. Ba, 2015 "ADAM: A Method for Stochastic Optimization", *The 3rd International Conference for Learning Representations*, 1-15
  30. Kuremotoa, T., S. Kimuraa, K. Kobayashib, M. Obayashia, 2014, "Time series forecasting using a deep belief network with restricted Boltzmann machines", *Neurocomputing* 137, Elsevier: 47-56
  31. McCulloch, W. S., W. Pitts, 1943, "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics* 5(4), Springer: 115-133
  32. Minsky, M., 1969, *Perceptrons*, M.I.T. Press
  33. Nair, V., G. E. Hinton, 2010, "Rectified Linear Units Improve Restricted Boltzmann Machines", In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 807-814
  34. Rumelhart, D. E., J. L. McClelland, PDP Research Group, 1986, *Parallel Distributed Processing Explorations in the Microstructure of Cognition Vol1*, The MIT Press
  35. Vincent, P., H. Larochelle, Y. Bengio, Pierre-Antoine Manzagol, 2008 "Extracting and composing robust features with denoising autoencoders", In *proceedings of the 25th international conference on Machine learning (ICML)*, 1096-1103

36. Zhang, G. P., 2003, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing* 50, Elsevier: 159-175.
37. 한국감정원 보도자료(2015.12.14), "한국감정원 · 한국거래소 「부동산 금융상품 투자지수 공동개발」을 위한 업무 협약식 개최"
38. <https://keras.io>, 파이썬 딥러닝 라이브러리
39. [www.r-one.co.kr](http://www.r-one.co.kr), 한국감정원 부동산통계정보

- 
- 논문 접수일 : 2017. 7. 24
  - 1차 수정일 : 2017. 8. 28
  - 2차 수정일 : 2017. 9. 11
  - 게재 확정일 : 2017. 9. 15