

데이터마이닝과 학습기법을 이용한 부동산가격지수 예측

이지영, 유재필*
키스채권평가 차장

Prediction of Housing Price Index using Data Mining and Learning Techniques

Jiyoung, Lee, Jae Pil Ryu*
Deputy Department Head, KIS Pricing

요약 4차 산업에 대한 관심이 증폭되면서 데이터를 활용한 과학적 방법론이 발전하고 있지만 부동산 분야에 대한 연구는 데이터 수집의 한계점을 내포하고 있다. 더불어 일반 시장 참여자들의 지식이 확장되면서 정성적인 심리가 부동산 시장에 미치는 영향이 커지고 있다. 때문에 본 연구에서는 기존의 원천 데이터가 아닌 심리적 부분을 반영한 정량 데이터를 텍스트마이닝과 k-means 알고리즘을 통해 수집하는 방안을 제안하고 수집된 데이터를 바탕으로 인공신경망 학습을 통해 주택 지수의 방향성을 예측하고자 한다. 2012년부터 2019년까지의 데이터를 학습 기간으로 하고 2020년도를 예측 기간으로 설정하여 실험을 진행한 결과, 두 가지 CASE에서 예측 능력이 약 80% 이상으로 우수하였고 주택 지수의 상승 구간에서의 예측 강도 또한 우수한 결과를 보였다. 본 연구를 통해서 의사결정에 있어서 부동산 시장 참여자들에게 인공신경망과 같은 과학적 방식의 활용도 증가 및 고전적 방식에서 벗어난 원천 데이터의 대체 데이터 확보 등에 대한 노력이 증진되기를 기대한다.

주제어 : 빅데이터, 인공신경망, 텍스트마이닝, 주택지수, 부동산

Abstract With increasing interest in the 4th industrial revolution, data-driven scientific methodologies have developed. However, there are limitations of data collection in the real estate field of research. In addition, as the public becomes more knowledgeable about the real estate market, the qualitative sentiment comes to play a bigger role in the real estate market. Therefore, we propose a method to collect quantitative data that reflects sentiment using text mining and k-means algorithms, rather than the existing source data, and to predict the direction of housing index through artificial neural network learning based on the collected data. Data from 2012 to 2019 is set as the training period and 2020 as the prediction period. It is expected that this study will contribute to the utilization of scientific methods such as artificial neural networks rather than the use of the classical methodology for real estate market participants in their decision making process.

Key Words : Big Data, Artificial Neural Network, Text Mining, Housing Index, Property

1. 연구배경

부동산에 대한 사람들의 관심이 높아지면서 부동산 시장에 대한 연구의 필요성이 커지고 있다. 부동산 시장과 관련된 다양한 통계 자료가 있는데 이런 통계 자료는 과거와 현재의 상황을 분석할 수 있는 기초 자료가 될 뿐만 아니라 미래의 부동산 정책을 결정하고 부동산 가격을 예측할 수 있는 자료로도 사용될 수 있기 때문에 그 중요성과 기대는 매우 크다. 부동산과 관련된 기초 통계 자료들은 동적으로 계속 변화하는데 주기적으로 과거와 유사한 패턴이 보이는 경우도 있는 반면에 특정한 분포 없이 무작위하게 변하는 성향을 갖고 있는 데이터도 있다. 결국 시장 참여자들이 관심을 갖는 것은 미래를 예측하는 것인데 최근 딥러닝(deep learning)과 같은 과학적 기법들이 발전하면서 시계열 데이터를 이용한 예측 기법들에 대한 활용도가 커지고 있다. 즉 분석 기법과 툴(tool)이 발전하면서 수치적 데이터의 적합성 및 실효성 등이 함께 중요해지고 있다. 그러나 현재 부동산 데이터의 공시 주기는 주로 월간이며 최초 공시 시점도 최근인 경우가 많아서 기술적 분석을 위한 실험 데이터가 매우 부족하다. 때문에 많은 부동산과 관련된 연구에서 시계열 자료의 부족 현상이 연구의 한계점으로 대두되고 있다[1]. 예컨대 부동산 시장의 유동성을 나타내는 거래량과 같은 데이터는 2006년부터 현재까지 월간으로 약 180개의 자료만 존재하며 이는 일별로 환산하면 약 3개월 정도로 데이터의 양이 매우 적은 것을 알 수 있다. 더불어 최근 시장 참여자의 주체가 정부와 투자 기관을 넘어서 일반 사람들에게도 과열 현상이 발생하면서 다양한 시장 참여자들의 정성적 심리 요인이 부동산 시장에 큰 영향을 끼치고 있으며 많은 기관에서는 부동산 시장의 매물에 대한 매수 심리 분석을 통한 부동산 심리 지수를 발표하고 있다.

경제학에서는 합리적인 의사결정을 하는 사람이 의사결정의 주체임을 가정하고 있지만 사실 사람들의 소비는 이성적이고 합리적이지만은 않다. 실제로 많은 연구에서 부동산 심리 지수가 부동산 가격과 거래량에 미치는 영향이 유의미하다고 설명하고 있다[2].

따라서 본 연구에서는 부동산 시장의 부족한 기초 통계 자료를 보완하고 시장 참여자들의 심리적 요인을 반영한 시계열 자료를 수집하여 이를 인공신경망(artificial neural network) 학습 기법을 통해 미래 부동산 가격의 추이를 예측하고자 한다. 시계열 자료를 수집하는 방법은 전문가 집단과 일반 집단으로 나눠 전자의 경우에는

부동산과 관련된 보고서를 바탕으로 텍스트마이닝(text mining)을 통해 주요 키워드(keyword)를 선정하고 후자의 경우는 k-평균 알고리즘(k-means clustering algorithm, 이하; k-means)을 통해 SNS(social network service)에서 부동산과 연관된 주요 키워드를 추출한다. 주택 소비자는 관심이 있는 자료에 대해 시장에서 제공되는 정보의 획득 과정이 이루어진다. 즉 사람들은 관심이 있는 자료에 대해 시장에서 제공되는 정보의 획득을 위해 인터넷 포털 사이트(internet portal site)에서 관심 키워드를 검색하며 이러한 정보 획득의 과정에서 발생하는 검색과 자료의 성격은 주택 가격의 방향성을 예측하는데 중요한 요인이 될 것으로 사료된다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 선행 연구를 기술하고 3장에서는 데이터 추출 방법과 인공신경망 학습에 관한 내용을 설명한다. 4장에서는 실험 계획 및 실험 결과에 대해서 기술하고 끝으로 5장에서는 결론을 제시한다.

2. 선행연구 고찰 및 차별성

본 장에서는 부동산과 관련된 연구 현황과 함께 텍스트마이닝을 이용한 연구 및 심리적 요인이 부동산 가격에 미치는 요인 등에 대한 선행연구를 기술한다.

2.1 부동산 연구

선행 연구를 살펴보면 부동산 시장에서 거시적 경제 변수 요인은 부동산 가격 결정 요인에 큰 영향을 준다. 김경민은 서울 아파트 시장과 거시 경제 변수 요인들 간 동적 상관관계를 실증분석 하였는데 고용률, 실업률, 소비자 물가 지수 등을 독립변수로 설정하여 종속변수인 서울 아파트 가격에 미치는 영향을 분석하였다[3].

부동산 가격 예측에 있어서는 VAR, ARIMA 모형 등의 시계열 분석 기법들이 있으며 주택과 상업용 부동산에 대한 부동산 경기 모형이 활발하게 연구되고 있다. 그러나 우리나라 시장에서 부동산 연구에 필요한 자료의 양적 정보의 한계성이 있으며, 특히 일별 자료가 제공되는 주식 시장과 달리 주로 월별 자료가 공시되고 있다.

최근에는 부동산 경기를 예측하기 위해 인공지능과 같은 과학적 기법을 적용하는 사례가 증가하고 있다. 김은미는 기계학습을 활용한 주택 매도 결정 요인 분석 및 예측 모델 구축을 연구하였고 머신러닝(machine learning) 모형과 OLS(ordinary least squares) 모형을 적용하여

예측력이 높은 모형을 도출하고 기계학습 방식의 활용 가능성을 제시하였다[4].

2.2 데이터마이닝 및 심리 요인 연구

전해정은 언론 보도 내용을 반영하는 특정 키워드와 뉴스 내용을 독립변수로 설정하고 아파트 가격 추이를 분석하였다[5]. 노민지는 주택 시장을 분석하기 위해 기본적인 심리 지수들이 주택 가격에 미치는 영향을 분석하였다[6].

부동산과 관련된 다양한 심리 지수를 많은 기관에서 발표하고 있는데 특히 주택 가격 심리 지수가 실제로 주택 가격 변동에 밀접한 연관이 있다는 선행 연구가 존재한다[7]. 조태진은 심리 지수가 주택 가격에 어떠한 영향을 주는지 알아보기 위해 거시적 경제 변수와 심리 지수 7개를 바탕으로 주택 시장에 미치는 정도를 분석하였다[8]. 그 결과 경제 심리 지수와 소비자 심리 지수는 유의하지 않지만 부동산 전망 지수는 시차를 두고 주택 가격에 영향을 미치고 있으므로 선행적 지표로 심리 지수를 활용할 수 있음을 주장했다.

김대원은 주택가격에 대한 심리적 태도가 주택 매매 거래량에 미치는 영향을 분석한 결과 소비 심리 지수가 주택 시장을 예측할 수 있는 변수로 활용될 수 있으며 시장 참여자의 가격에 대한 심리적 태도는 기대치가 거래량의 결정으로 이어진다고 설명했다[9]. 허선영은 온라인 커뮤니티 자료를 활용하여 부동산 참여자에 대한 심리 지수를 산출하였는데 이는 부동산 시장에 시차를 두고 선행 또는 후행으로 가격 지수에 영향을 주는 것으로 나타났다[10].

3. 연구 모형

본 연구에서는 부동산과 관련된 전문가 보고서를 텍스트마이닝을 통해 6개의 주요 키워드를 추출하는 동시에 트위터(twitter)의 부동산 관련 글에서 k-means 알고리즘을 통해 연관 키워드 6개를 선정한다. 이렇게 선정된 키워드들은 네이버 트렌드(naver trend)를 통해서 검색 빈도 값을 추출하고 이는 주택 가격 예측을 위한 인공신경망 학습의 입력 값(input date)로 사용한다.

3.1 텍스트마이닝

텍스트마이닝은 언어와 같은 비정형 텍스트 자료를 자연어 처리 기술을 통해 필요한 정보를 추출하고 그것들

의 관련성 및 군집화 등의 분석을 통해서 의미 있는 정보를 파악하는 기술이다. 즉 텍스트마이닝은 빅데이터(big data)를 시각적으로 알기 쉽게 정보를 파악할 수 있다는 장점이 있다.

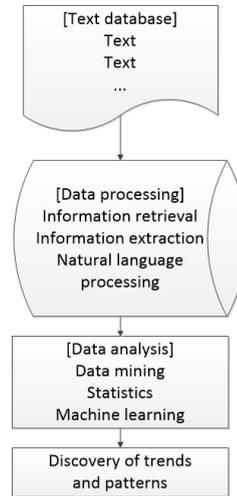


Fig. 1. Course of Text Mining

텍스트마이닝은 데이터 수집, 용어 추출, 정보 추출 그리고 정보 분석이라는 4 단계의 과정을 통해 마이닝을 수행한다. 데이터 수집을 완료하면 문장의 패턴과 단어들의 관련성 등을 고려하여 대표 키워드를 선정한다. 그리고 텍스트 데이터에서 의미가 있는 정보를 추출하는 과정을 거친 후 끝으로 빈도, 군집화(Clustering) 등의 분석을 통해 텍스트마이닝 수행을 완료한다[11]. Fig. 1은 앞서 설명한 텍스트마이닝의 과정을 도식화한 것이다.

3.1.1 전문가 집단 요인

전문가 집단에 해당하는 키워드를 추출하기 위해서 국내 금융기관에서 발간하는 최근 5년 동안의 부동산 관련 보고서를 주기적인 web scroll 방식을 통해 수집한다. 수집된 보고서를 텍스트마이닝을 통해서 선정된 주요 키워드는 양도소득세, 역세권, 대출, 매매, 전세 그리고 월세이다. 해당 키워드를 네이버 트렌트를 통해서 주별로 트렌드 값을 수집하였으며, Table 1은 주택가격지수와 해당 키워드 간의 상관계수를 보여주고 있다. 그 결과 매매와 역세권의 키워드가 상대적으로 상관계수가 높은 편인 반면에 월세는 -0.45로 다소 상관계수가 낮은 것으로 나타난다.

Table 1. Correlation coefficients for Key Keyword

	big deposit (= EF_1)	small deposit (= EF_2)	tax (= EF_3)
Price Index	0.16	-0.45	0.11
	station area (= EF_4)	trade (= EF_5)	loans (= EF_6)
Price Index	0.40	0.65	-0.04

3.2 k-means 알고리즘

k-means는 특정 기준을 인위적으로 정하지 않고 유사도가 높은 데이터끼리 하나로 묶어주는 방법을 의미한다. k-means는 군집의 중심 수를 선정하고 각 데이터의 대표성을 보이는 k개의 군집으로 선별하는데 k 값 외에 다른 사전정보가 없기 때문에 비교적 모형의 설계가 간단하다.

k-means는 각 그룹의 중심(centroid)과 그룹 내의 데이터 오브젝트와의 거리의 제곱 합을 비용 함수로 정하고, 이 함수 값을 최소화할 수 있는 방향으로 각 데이터 오브젝트의 소속 그룹을 업데이트 해 줌으로써 클러스터링을 수행하게 된다[12]. 본 연구에서는 최근 5년간의 트위터를 분석하기 위해 Neha Garg가 제안하는 R language를 이용해 부동산과 관련된 글에서 주요 키워드를 추출하고자 한다[13]. 이는 사용자들이 부동산과 관련된 글을 작성하면서 주요하게 거론된 키워드라고 볼 수 있다.

3.2.1 일반 집단 요인

일반 집단 요인에 해당하는 키워드를 추출하기 위해서 부동산과 관련된 트위터를 선정하고 선정된 트위터 내에서 작성되는 댓글을 k-means 알고리즘을 통해서 분석한다. 그 결과 일반 집단 요인으로는 전매, 아파트, 전세, 청약, 토지, 재건축이라는 주요 키워드가 선정되었다. 특히 최근에 새로운 정책 등으로 인해서 탄생한 새로운 단어들은 주요 키워드 선정 기준에서 제외했다. Table 2는 주택 가격 지수와 각 키워드 간의 상관계수를 보여주고 있는데 전문가 집단 요인에 비해서 상대적으로 상관계수

Table 2. Correlation coefficients for Key Keyword

	reconstruction(= GF_1)	Big deposit (= GF_2)	land (= GF_3)
Price Index	0.38	0.16	0.02
	subscription (= GF_4)	apartment (= GF_5)	resale (= GF_6)
Price Index	0.36	0.46	0.53

가 높게 나타난다. 특히 전문가 집단 요인 경우에는 월세와 대출이 음의 상관관계를 보인 반면에 일반 집단 요인 경우에는 모두 양의 상관관계를 보인다.

3.3 인공신경망 모형

인공신경망 학습 기법은 데이터들의 패턴을 인식하고 분류 문제(classification problem)를 해결하는 비선형 분석 모델(nonlinear analytical model)이다. 특히 인공신경망 모형은 인간의 우뇌 기능인 학습(learning), 병렬처리(parallel processing), 패턴인식(pattern recognition) 그리고 오차용인(fault tolerance)과 같은 생물학적 과정을 공학적으로 컴퓨터에 적용한 기법으로 인간 두뇌의 기본 단위인 뉴런(neuron)의 구조에 착안하여 개발되었다[14]. 인공신경망은 은닉마디(hidden units)로 이뤄진 은닉 층이 존재하기 때문에 고전적 방식인 회귀분석과 차별성이 있다.

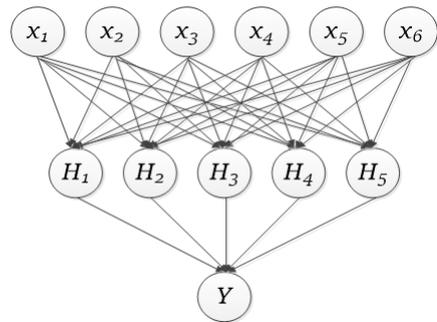


Fig. 2. Structure of the ANN

Fig. 2는 본 연구에서 구축하고자 하는 인공신경망 모형의 구조를 나타낸 그림인데 n 개의 입력 값인 x_i 가 중된 w_i 를 반영하여 신경망을 구축하고 w_i 가 반영된 x_i 의 편차인 b 를 합해 식(1)과 같이 A 를 구한다.

$$A = \sum x_i w_i + b \tag{1}$$

그리고 활성화 함수(activation function)인 f 를 통해서 출력 값을 계산하는데 이는 식(2)와 같이 계산한다.

$$f(A) = f(x_i w_i + b) \tag{2}$$

식(2)에서 w_i 와 x_i 는 실수 값이고 입력 변수가 많으면 산출 과정을 단순화하기 위해서 편차 값인 b 를 가상 입

력 값인 $x_0 = +1$ 로 가정하고 w_0 로 개정할 수 있다.

본 연구에서는 총 6개의 입력 층과 하나의 은닉 층에서 5개의 은닉마디를 갖는 인공신경망 모형을 모델링(modeling)하고 이에 대한 예측력을 분석하고자 한다.

4. 실험계획 및 분석

본 장에서는 앞 장에서 인공신경망 모형의 입력 값으로 선정된 총 12개의 키워드를 바탕으로 부동산 가격 방향성을 예측하고 결과를 분석하고자 한다.

4.1 실험계획

본 연구에서는 주택 가격의 모멘텀(momentum) 및 변동성에 따라 일반 대중들의 포털 사이트 검색 빈도가 연관관계가 있다는 것을 가정해 실험을 진행한다[15].

Table 3. Experimental plan

Factors	Details
Subjects to experiment	Korea House Price Index
Input Variables	Group of Experts Factors(CASE ₁) : $EF_1, EF_2, EF_3, EF_4, EF_5, EF_6$ General population factors(CASE ₂) : $GF_1, GF_2, GF_3, GF_4, GF_5, GF_6$
Keyword Extraction Method	Group of Experts Factors(CASE ₁) : Text Mining General population factors(CASE ₂) : K-Means
Experimental period	Learning Period : 2012.05~2019.12 Prediction period : 2020.01~2020.12
Source of Experimental Data Collection	Naver Trend Search Frequency (Weekly)
Learning Methods	Artificial Neural Network Model

이를 위해 전문가 집단과 일반 집단으로 나눠 해당 집단에서의 주요 키워드를 텍스트마이닝과 k-means 알고리즘으로 추출하고 해당 키워드의 정량적 검색 빈도를 수집하였다.

수집된 키워드들은 편리성을 위해 Table 1과 Table 2에서 정의한 함축 단어인 EF 와 GF 로 정의하며 이는 인공신경망 학습의 입력 값으로 사용된다. 목표 값은 주택 가격 지수로 하고 한 달 후 지수가 상승하면 상승 그리고 한 달 후 지수가 하락하면 하락이라고 정의한다. 실험의 모든 데이터는 주별 단위로 수집하고 인공신경망 학습

기간은 2012년도 5월부터 2019년까지이며, 예측 기간은 2020년도이다. Table 3은 본 연구의 실험계획을 정리한 표이며 Fig. 3은 실험 과정을 도식화한 그림이다.

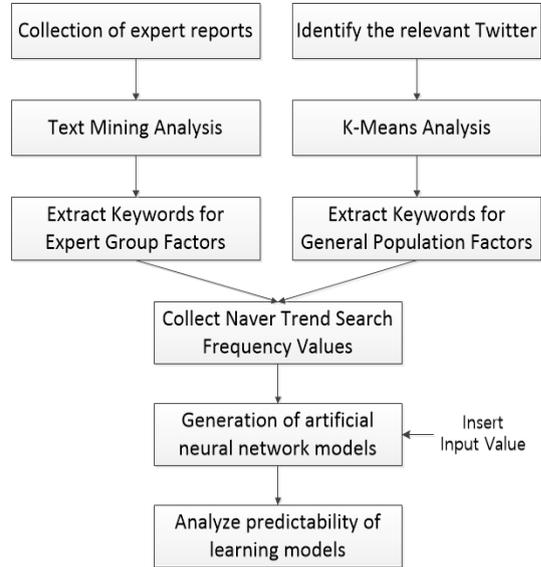


Fig. 3. Experimental Process

4.2 결과분석

본 절에서는 인공신경망을 이용해 주택 가격 지수의 방향성을 예측한 결과를 기술한다. Table 4는 각 CASE에 대해서 월별, 분기별로 나눠서 예측 결과를 정리하였고 예측 성공률, 방향성을 맞춘 횟수 그리고 예측 값에 대한 예측 강도 값을 나타내고 있다.

표의 맨 마지막 열인 예측 값에 대한 예측 강도(Result/Target)에서 Target 값이 1에 가까울수록 해당 시기의 주택 지수가 상승한 것을 의미하고 0에 가까울수록 하락한 것을 의미한다. Result 값 또한 1에 가까우면 가까울수록 강한 상승 방향으로 예측된 것을 의미하고 0에 가까울수록 강한 하락 방향으로 예측된 것을 의미한다. 즉 Result 값과 Target 값의 차가 작을수록 예측 강도가 실제 값인 목표 값과 거의 유사하기 때문에 우수한 결과라고 볼 수 있다.

예측 성공률을 보면 3~4월 달에 저조한 것으로 보이는데 같은 시기에 주택 지수가 하락하는 구간으로써 이는 학습 기간인 2012년 이후에 주택 지수가 상승 추이를 보인 것과 관련이 있다. 즉 주식 시장과는 다르게 부동산 가격은 과거부터 지속적인 상승을 보이면서 하락 구간에 대한 학습 능력이 다소 떨어진 결과라고 사료된다. 특히

주택 지수가 계속 상승을 보인 7월 이후에는 예측 능력은 물론 강도 또한 매우 우수하다는 것을 확인할 수 있는데 Result 값과 Target 값의 차도 약 0.03으로 다른 기간에 비해서 매우 낮게 나타난다.

CASE별로 살펴보면 방향성 예측 능력은 물론 Result 값과 Target 값의 차이도 CASE₂가 CASE₁에 비해서 일괄적으로 우수하다. 예컨대 CASE별로 예측 능력과 강도가 서로 반비례하게 나올 수도 있는 반면 본 실험에서는 비례적으로 나온 것을 알 수 있다. 전체 기간 동안의 강도는 CASE₁이 0.07이고 CASE₂가 0.14이며, 예측 능력은 각각 86.54%, 80.77%로 나타났다.

Fig. 4는 예측 강도와 Target 값을 비교하여 도식화한 그림인데 우선 Target은 상당 기간 이상 1로 나오면서 상승 모멘텀을 유지하고 있다. CASE₁와 CASE₂를 비교하면 상승 부분에서는 CASE₂의 변동성이 CASE₁에 비해서 낮았다. 앞에서 설명했듯이 3~4월 달에 Target 값이 하락 했을 때 CASE₁이 하락 방향성을 예측하면서 CASE₂에 비해서 예측 능력이 우수했다.

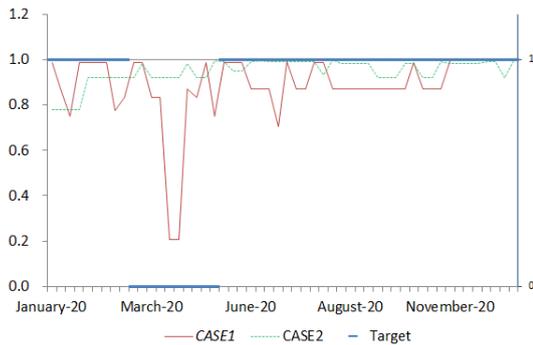


Fig. 4. Prediction Strength and Target Value

Table 4. Experimental results

monthly	Factor Groups	Success rate	Number of passes	Result/Target
1	CASE ₁	100.00	4/4	0.90/1.00
	CASE ₂	100.00	4/4	0.78/1.00
2	CASE ₁	100.00	4/4	0.93/1.00
	CASE ₂	100.00	4/4	0.92/1.00
3	CASE ₁	20.00	1/5	0.89/0.20
	CASE ₂	20.00	1/5	0.93/0.20
Total (1Q)	CASE ₁	69.23	9/13	0.91/0.69
	CASE ₂	69.23	9/13	0.88/0.69
4	CASE ₁	50.00	2/4	0.53/0.00
	CASE ₂	0.00	0/4	0.94/0.00
5	CASE ₁	50.00	2/4	0.93/0.50
	CASE ₂	50.00	2/4	0.97/0.50

6	CASE ₁	100.00	5/5	0.86/1.00
	CASE ₂	100.00	5/5	0.98/1.00
Total (2Q)	CASE ₁	69.23	9/13	0.78/0.54
	CASE ₂	53.85	7/13	0.96/0.54
7	CASE ₁	100.00	4/4	0.93/1.00
	CASE ₂	100.00	4/4	0.99/1.00
8	CASE ₁	100.00	5/5	0.89/1.00
	CASE ₂	100.00	5/5	0.98/1.00
9	CASE ₁	100.00	4/4	0.87/1.00
	CASE ₂	100.00	4/4	0.94/1.00
Total (3Q)	CASE ₁	100.00	13/13	0.90/1.00
	CASE ₂	100.00	13/13	0.97/1.00
10	CASE ₁	100.00	4/4	0.90/1.00
	CASE ₂	100.00	4/4	0.95/1.00
11	CASE ₁	100.00	4/4	0.97/1.00
	CASE ₂	100.00	4/4	0.99/1.00
12	CASE ₁	100.00	4/4	1.00/1.00
	CASE ₂	100.00	4/4	0.98/1.00
Total (4Q)	CASE ₁	100.00	13/13	0.96/1.00
	CASE ₂	100.00	13/13	0.97/1.00
Total (1Y)	CASE ₁	86.54	45/52	0.89/0.81
	CASE ₂	80.77	42/52	0.95/0.81

5. 결론

본 연구는 심리적인 비정형 데이터(unstructured data)를 정형화(regularization)하여 부동산과 관련된 키워드를 추출하고 해당 키워드들의 검색 빈도 값을 수집하여 인공지능망 모형의 실험 데이터로 사용하였다. 키워드 추출에 있어서는 전문가 요인과 일반 요인으로 크게 나눠 각각 6개의 주요 키워드를 정량적인 기법을 통해 선정하고 각 키워드들의 주별 검색 빈도 값을 수집함으로써 부동산 원천 데이터의 부족 현상에 대한 한계점을 극복하기 위해 노력했다. 예측 목표 값인 주택 지수를 예측하기 위해서 2012년부터 2019년까지의 학습기간과 2020년도의 예측기간으로 나눠 실험을 진행하였고 그 결과 학습 데이터가 부족했던 하락 구간을 제외하고는 대부분의 구간에서 우수한 예측 능력을 확인했다. 특히 전문가 집단 요인이 일반 집단 요인에 비해서 예측력 및 예측 강도에 있어서 예측 결과가 우수하게 나타났으며 CASE 별로 52주 중에서 각각 45회, 42회의 방향성을 예측하였다. 다만 2012년도부터 학습 데이터에 있어서 주택 지수가 지속적인 상승 모멘텀을 보이고 있어서 다양한 시장 상황에서의 예측력을 분석할 수 없었다는 점에서 본 연구의 한계점을 보이고 있다.

본 연구는 전통적인 부동산 원천 데이터에 국한하지 않고 텍스트마이닝 및 k-meas과 같은 과학적 기법을 이용해서 부동산과 관련된 키워드를 추출하고 정성적 심리 현상을 정량적인 수치로 반영되는 포털 사이트 검색 빈도 값을 인공지능망 학습의 실험 데이터를 사용했다는 점에서 의의가 있다고 사료되며 다양한 데이터 확보에 대한 방법론적인 확장 연구가 진행되기를 기대한다.

REFERENCES

[1] J. Y. Lee & J. P. Ryu. (2021). Prediction of Housing Price Index Using Artificial Neural Network. *Journal of the Korea Academia-Industrial*, 22(4), 228-234. DOI : 10.5762/KAIS.2021.22.4.228

[2] D. W. Kim & J. S. Yu. (2013). Analysis on How Psychological Attitudes on the House Price Affect the Trading Volume. *Korean Association For Housing Policy Studies*, 21(2), 73-92. DOI : G704-000825.2013.21.2.005

[3] K. M. Kim. (2018). A Study on Dynamic Correlations between the Seoul Apartment Market and Factors of Macroeconomic Variables. *Korea Real Estate Academy*, 73(1), 115-129.

[4] E. M. Kim, S. B. Kim & E. S. Cho. (2020). Using Mechanical Learning Analysis of Determinants of Housing Sales and Establishment of Forecasting Model. *Journal of Cadastre & Land Informatix*, 50(1), 181-201. DOI : 10.22640/lxsiri.2020.50.1.181

[5] H. J. Chun. (2020). Big Data Time Series Analysis of the Relationship Between News Coverage Apartment Price and Trading Volume. *Korean Society of Real Estate Law Society*, 24(2), 53-69. DOI : 10.32989/rel.2020.24.2.53

[6] M. J. Noh & S. J. Yoo. (2016). A Relationship between Sales Prices of APT and Consumer Sentiment. *Korea Research Institute for Human Settlements*, 89(1), 3-13. DOI : 10.15793/kspr.2016.89..001

[7] H. S. Yoo & J. H. Chung. (2015). The Lead-Lag Relationship between Housing Purchase Price Index and Consumer Sentiment Index. *korea real estate research institute*, 25(4), 49-61. UCI : G704-001886.2015.25.4.002

[8] T. J. Cho. (2014). A Study on the effect of the sentiment index to the housing market. *Korean Association For Housing Policy Studies*, 22(3), 25-48. UCI : G704-000825.2014.22.3.002

[9] D. W. Kim & J. S. Yu. (2013). An Analysis on How Psychological Attitudes on the House Price Affect the Trading Volume. *Korean Association For Housing Policy Studies*, 21(2), 73-92.

UCI : G704-000825.2013.21.2.005

[10] S. Y. Heo, J. Y. Kim & T. H. Moon. (2019). Consumer Sentiment in On-line Community and the Variation of Real Estate Market. *Journal of the Korean housing association*, 17(4), 31-41. DOI : 10.22313/reik.2019.17.4.31

[11] J. P. Ryu, C. H. Han & H. J. Shin. (2016). Sector Investment strategies Using Big Data Trends. *Korea Institute of Enterprise Architecture*, 13(1), 111-121. DOI : G704-SER000010357.2016.13.1.004

[12] J. P. Ryu, H. J. Shin, M. H. Kim & J. K. Baek. (2017). Pattern Analysis of Stock Prices Using Machine Learning and Data Visualization. *Korea Institute of Enterprise Architecture*, 14(2), 189-197.

[13] G. Neha & R. Rinkle. (2017). Analysis and visualization of Twitter data using k-means clustering. *ICICCS*, 15(1), 15-16. DOI : G704-SER000010357.2016.13.1.004

[14] J. P. Ryu & H. J. Shin. (2018). Portfolio Selection Strategy Using Deep Learning. *Korea Institute of Enterprise Architecture*, 15(1), 43-50. DOI : 10.22865/jita.2018.15.1.43

[15] J. M. Lim & M. H. Lim. (2016). A Study of the Relationship Between Agents Sentiment and Housing Market. *Journal of the Korean regional development association*, 28(3), 147-164. UCI : G704-000688.2016.28.3.005

이 지 영(Jiyoung Lee)

[정회원]



- 2017년 2월 : 한성대학교 일반대학원 경제부동산학과(부동산학박사)
- 2006년 4월 ~ 현재 : KIS채권평가 평가본부 주식파생실
- 관심분야 : 부동산학, 기계학습, 딥러닝
- E-Mail : jiyoung.lee@kispricing.com

유 재 필(Jae Pil Ryu)

[정회원]



- 2017년 2월 : 상명대학교 일반대학원 공과대학 경영공학과(공학박사)
- 2013년 10월 ~ 2016년 10월 : KIS 채권평가 금융공학연구소
- 2016년 11월 ~ 현재 : KIS채권평가 평가본부 주식파생실
- 관심분야 : 금융공학, 기계학습, 딥러닝
- E-Mail : jaepilryu@kispricing.com