

<박사학위논문 소개>

주택시장 예측을 위한 부동산 감성지수 개발 연구*

A Study of the Development of Real Estate Sentiment Index for Housing Market Prediction

저 자: 박 재 수 (Park, Jae-Soo)**

지도교수: 이 재 수 (Lee, Jae-Su)***

스마트 혁명으로 촉발된 초연결 사회가 도래하면서 유통 정보량은 이전과는 비교할 수 없을 정도로 증가되었다. 소셜네트워크 서비스를 통한 왕성한 의사소통은 집단적 기대심리를 형성하고 전통적 시장에 영향을 끼치는 주요 변수가 되고 있다. 그런데 부동산 시장을 예측하는 전통적 사회과학적 연구방법론은 주택보급률, 경제성장률, 금리 등 정형화된 변수만 활용했을 뿐 부동산 심리 요인은 거의 다루지 않고 있어 개선의 목소리가 높아지고 있다.

본 연구의 목적은 첫째, 부동산 관련 신문기사와 방송뉴스를 활용하여 부동산시장의 움직임을 설명 또는 예측할 수 있는 감성지수를 기계학습 방법론을 활용하여 개발하는 것이다. 둘째, 새롭게 개발된 감성지수가 아파트 매매가격의 예측에 유용성이 있는지를 파악하는 것이다. 마지막으로, 신문과 방송뉴스 기사를 활용한 감성지수가 아파트 매매가격지수의 예측력 향상을 비교·분석하는 것이다.

분석자료는 KB 국민은행에서 제공하는 아파트 매매가격지수와 비정형 텍스트 데이터인 신문과 방송의 부동산 관련 뉴스기사이다. 부동산 관련 뉴스기사의 텍스트 데이터는 일간지 3개(조선, 동아, 중앙)와 경제지 3개(매경, 서경, 한경) 및 지상파 3사(KBS, MBC, SBS)의 인터넷 사이트에서 웹 크롤링을 통하여 수집하였다.

새로운 감성지수 개발을 위해 토픽모델, 텍스트랭크, TF-IDF, 나이브 베이즈 등 텍스트 분석 및 기계학습 방법을 적용하였다. 그리고 감성지수의 유용성 검증에는 시계열 분석모형인 ARIMA 및 ARIMAX 모형을 적용하였다. 비정형 데이터인 신문과 방송뉴스에서 긍정과 부정 문장들의 극성을 분류하는데 주관적 개입을 차단하고 객관적이고 체계적인 방법론을 도입하기 위

하여 기계학습을 이용하였다. 개발된 감성지수의 유용성을 검증하기 위하여 아파트 매매가격지수를 종속변수로 하는 ARIMA 모형을 구축하였다. 또한 감성지수를 외생변수로 투입한 ARIMAX 모형을 적용하고 모형의 예측 오차인 RMSE와 MAE 값을 비교하여 유용성을 평가하였다. 분석 결과, 신문과 방송 부동산 기사를 활용해 도출한 감성지수를 아파트 매매가격지수의 외생변수로 하여 ARIMA 모형을 추정할 경우 모형의 예측력이 유의미하게 향상됨을 확인하였다.

본 연구의 차별성은 첫째, 빅데이터와 기계학습 방법을 활용하여 감성지수를 도출하고 연구자의 주관적인 개입이 배제된 체계적이고 자동화된 방법론을 제시하였다는 점이다. 둘째, 신문과 방송뉴스 기사에서 신문 감성지수와 방송 감성지수를 도출하여 언론매체의 부동산시장에 대한 영향력의 비교·분석을 한 점이다. 부동산시장에 대한 소비자 심리를 파악하기 위해 기존의 설문조사 방법을 벗어나 빅데이터와 기계학습을 접목하여 부동산시장에 대한 소비자 심리를 파악할 수 있는 새로운 방법론을 제시한 점에서 학문적 의의가 있다.

주제어 : 빅데이터, 감성분석, 토픽 모델, 나이브 베이즈, 자기회귀이동평균모형

Keyword : Big data, Sentiment Analysis, Topic Model, Naive Bayes, ARIMA Model

* 본 논문은 필자의 강원대학교 박사학위 논문인 박재수(2020) '주택시장 예측을 위한 부동산 감성지수 개발 연구'를 요약한 것임.

** 강원대학교 부동산학 박사, goodmans@kakao.com, 주저자

*** 본 학회 정회원, 강원대학교 부동산학과 교수, jslee25@kangwon.ac.kr, 교신저자