생성 AI 기반 뉴스 기사 심리지수를 활용한 부동산 가격 예측 모델

김수아, 권미주, 조수빈, 김은수, 김현희
동덕여자대학교 정보통계학과

20200007@dongduk.ac.kr, 20201039@dongduk.ac.kr, 20201079@dongduk.ac.kr,
20201047@dongduk.ac.kr, heekim@dongduk.ac.kr

Predictive Model for Real Estate Prices Using Sentiment Index of news articles based on Generative AI

Kim Sua, Kwon Miju, Cho Soobin, Kim Eunsoo, Hyon Hee Kim
Dept. of Information Statistics, Dongduk Women's University

요 약

부동산 시장은 다양한 요인에 의해 가격이 결정되며 거시경제 변수뿐 만 아니라 뉴스 기사, SNS 등 다양한 비정형 데이터의 영향을 받는다. 특히 뉴스 기사는 국민들이 느끼는 경제 심리를 반영하고 있어 부동산 가격에 영향을 크게 미치는 변수라고 판단된다. 본 연구에서는 뉴스 기사의 세분화된 감정 분석을 통해 전통적인 분석 방법보다 더 의미 있는 결과를 얻을 수 있는 부동산 가격 예측 모델을 생성하였으며 뉴스 기사로부터 심리 지수를 산출하기 위해 생성 AI 를 활용하였다.
제안하는 매매가격지수 예측 모델을 통해 부동산 시장과 뉴스 기사와의 관계성에 대해 파악할 수 있으며, 사회/경제적 동향을 반영한 부동산 가격 변동을 예측할 수 있을 것으로 보인다.

1. 서론

부동산 매매가격지수를 예측하는데 거시경제 변수인 금리, GDP, 대출금리, 소비자물가지수 등의 변수가 주로 활용되어 왔다 [1]. 그러나 2006 년 경제 위기 이후 소비심리는 중요한 경제정보로 간주되어 왔으며, 사람들의 소비 의사결정에 굉장히 중요한 요소임이 밝혀졌다 [2]. 하지만 이러한 변수들은 개인의 소비 패턴과 정책적 요소를 반영하지 못하는 한계가 있다.

본 연구에서는 앞선 변수와 더불어 선행 지표가 될 수 있는 뉴스기사를 감정 분석하여 심리 지수로 추출하고 이를 새로운 변수로 사용하여 부동산 시장 예측 모델에 새로운 차원을 부여하고자 하였다. 특히 심리 지수 추출을 위해 생성 AI 를 활용하여 뉴스 기사의 긍정 및 부정 점수를 산출하였으며, 이를 입력 변수로 활용하였다. 예측 모델로는 시계열 데이터에 적합한 Multihead attention LSTM 모델 [3]을 적용하였다.

* 본 연구는 2023 학년도 중소기업벤처부의 SW ·콘텐츠 인재 양성사업(벤처스타트업아카데미) 지원에 의한 연구임

실험 결과 뉴스기사 심리지수를 사용한 부동산 매매가격지수 예측 모델이 이를 사용하지 않은 경우보다 성능이 우수함을 알 수 있었다. 따라서 부동산 매매가격지수와 같이 경제적 변수에 의해 가격이 결정되는 다양한 분야에서 뉴스기사 심리지

수를 활용하면 보다 성능이 뛰어난 예측 모델을 만들 수 있을 것으로 기대된다.

2. 데이터 수집 및 전처리

모델의 특징 변수는 기준금리, 명목 GDP, 대출금리, 전국 아파트 매매 거래량 총 4 개를 사용하였다.

본 연구의 기간은 2019 년 1 월부터 2023 년 3 월이며, 학습 세트는 2019 년 1 월부터 2022 년 12 월까지의 총 48 개월로 설정하였다. 테스트 기간은 2023 년 1 월부터 2023 년 3 월까지이다.

LSTM 모델의 경우 경향성과 계절성은 분석에 영향을 미치지 않기 때문에 차분은 하지 않으며 금리를 제외한 모든 변수에 최대최소 정규화를 실시하였다.

3. 뉴스 심리지수 생성

부동산 가격은 정부 정책과 외부환경에 대해 영향을 많이 받으며 뉴스 기사는 이러한 정부의 정책과 심리를 반영하고 있으므로 아파트 매매 가격지수 예측에 뉴스 정보가 반영될 필요성이 있다. 뉴스 데이터는 국내에서 정보검색을 위해 가장 통상적으로 활용되고 있는 포털 사이트인 네이버를 이용하였다. 기간은 테스트 기간을 포함하는 총 51 개월간의 부동산 뉴스를 수집하였으며 월별로 600 개씩 총 30,600 개의 뉴스 정보 를 python 을 이용하여 크롤링하였다.

뉴스 전문을 읽고 감정 분석하는 것은 비효율적이기 때문에 정보 처리의 효율성과 정확성을 높이기 위해 Text Summarization 를 먼저 수행하였다. 본 연구에서 사용한 KOBART ("ainize/kobart-news")는 SKT 에서 공개한 한국어 BART 모델이다.

본 연구에서는 최근 사회에서 주목받고 있는 Chat GPT 를 활용한 데이터 라벨링을 수행하였다. GPT 모델 중 2020 년에 발표된 모델인 'text-davinci-002'를 사용하였으며 175 억 개의 파라미터를 가지며 대화형 AI, 문장 생성, 질의응답에서 높은 성능을 나타낸다. 각 뉴스 기사 요약문에 대해 감정 분석 후 긍정 1, 중립 0, 부정 -1 로 수치화하였다. GPT 를 통한 자동 라벨링 결과 중 라벨링 되지 않은 부분과 중립 내용에 대해 사람이 직접 개입하는 반자동 라벨링을 통해 정확한 라벨링을 수행 하였다. 이후 월별로 감정분석 수치의 총합을 식 (1)과 같이 계산하여 예측모형의 새로운 입력변수로 사용하였다.

뉴스 심리지수 = ∑ (Pos(i)×1 + Neu(i)×0 + Neg×-1) <식 1>

뉴스 심리지수: 월간 뉴스 감정분석 값

Pos: 긍정, Neu: 중립, Neg: 부정

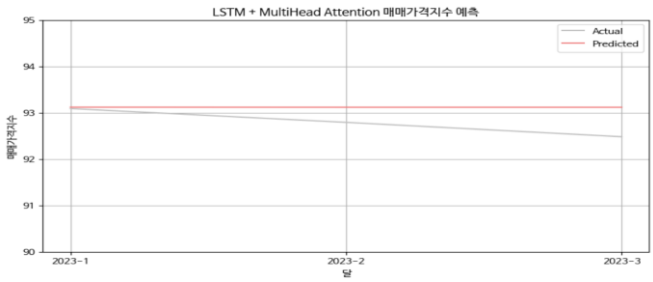
4. Multi-head attention based LSTM

부동산 매매가격지수 예측을 위해 LSTM-based Multi head Attention 알고리즘을 사용하였다. 입력 데이터에 대해 attention 을 적용하는 부분에 LSTM 을 사용함으로써 key, query, value 를 생성하는 과정에서 입력 데이터의 중요성을 예측 모델에 강조하였다. n_heads=16, hidden_layer 는 각각 64, 32 를 사용하였다. 손실 함수, 평가 지표는 mse, 최적화 알고리즘은 Adam 과 learning_rate 는 1e-3 을 사용하였다. 1, 2, 3 개월 후 가격 예측도는 RMSE (Root Mean Squared Error)로 나타내었으며 <표 1>과 같다.

구분	1 month	2 month	3 month
심리지수 미반영모델	0.0934	0.390	0.70
심리지수 반영모델	0.03	0.326	0.636

<표 1> 모델별 예측 성능 평가

(그림 1)은 예측 모델의 결과로 나타난 예측값과 실제값의 비교를 나타낸다. 미래 예측 성능은 뉴스 심리지수를 반영한 모델의 Scaling 이전의 1, 2, 3 개월의 총 RMSE 가 0.1841 로 나타났다. 이를 통해 본 연구에서 가장 주요하게 보는 뉴스 심리지수가 포함된 모델이 심리지수가 포함되지 않은 모형보다 아파트 매매가격지수의 예측 성능이 높은 것으로 나타났다.



(그림 1) 아파트 매매가격지수 예측값 실제값 비교

5. 결론

본 연구 결과로 인해, Multi-head attention LSTM 모형에서 뉴스 기사의 감정 분석을 심리지수로 결합할 경우 아파트 매매가격지수의 예측력이 소폭 향상된 것을 확인할 수 있다. 이러한 개선된 예측력은 부동산 뉴스와 같은 비정형 데이터가 부동산 시장과 가격에 유의미한 영향을 미친다는 것을 보여주며 금융데이터 예측 향상에 큰 공헌을 할 수 있을 것으로 기대된다. 또한, 부동산에 특화된 감성사전이 존재하지 않는 상황에서 Chat GPT 를 활용한 감정 분석 라벨링을 수행한 것에 의의가 있다. 본 연구에서는 뉴스 기사를 통한 심리지수만을 포함하였다면 추후 연구에서는 부동산 카페와 같은 대중의 실질적 여론을 포함한 심리지수를 이용한다면 보다 더 정교한 예측 결과가 나올 것으로 예상한다.

참고문헌

[1] 박주영, 2004,「소비심리와 소비, KDB 산업은행.

[2] 함종영 · 손재영(2016). 사전확률분포를 이용한 주택시장 예측모형 비교 연구-Bayesian VAR 모형을 중심으로. 부동산・도시연구, 8(2), 25-38.

[3] 배성완 . 유정석(2018). 머신 러닝 방법과 시계열 분석 모형을 이용한 부동산 가격지수 예측. 주택주택연구, 26(1), 107-13