

GPT-4o 기반 언어 모델을 활용한 주식 등락 예측: 뉴스 감성 분석의 영향력 평가

정호연¹, 홍성초², 임선영^{3*}

배재대학교 드론로봇공학과¹, 배재대학교 공학연구소², 배재대학교 컴퓨터공학과^{3*}
2580jhy@naver.com, scv.hong@pcu.ac.kr, sunnyihm@pcu.ac.kr

*교신저자

Stock Movement Prediction Using GPT-4o Based Language Models: Evaluating the Impact of News Sentiment Analysis

Ho-Yeon Jeong¹, Seong Cho Hong², Sun-Young Ihm³

Department of Drone · Robot Engineering, Pai Chai University¹

Engineering Research Institute, Pai Chai University²

Department of Computer Engineering, Pai Chai University³

요 약

본 연구에서는 GPT-4 기반의 자연어 처리를 활용하여 뉴스 기사 제목의 감성 분석을 통해 주식 가격 예측 성능을 개선하는 방법을 제안한다. Python을 이용한 웹 크롤링으로 주식 관련 뉴스 기사 제목을 수집하고, GPT API로 긍정도를 분석하여 이를 주식 데이터와 결합한 후 머신러닝 모델에 적용하였다. 여러 회귀 모델을 통해 예측 성능을 비교한 결과, 감성 분석을 포함한 모델에서 예측 성능이 올라감을 확인하였다. 이 연구를 통해 뉴스 감성이 주식 시장 예측에 미치는 영향을 확인할 수 있다.

1. 서 론

현대 금융 시장에서 주식의 가격 변동을 예측하는 것은 투자자와 기업 모두에게 매우 중요한 과제이다. 주식 시장은 다양한 요인에 의해 영향을 받으며, 그중에서도 경제 지표, 기업 실적, 시장 심리 등이 주된 변수로 작용한다. 특히, 최근에는 빅데이터와 인공지능을 활용한 주식 예측 모델이 많은 관심을 받고 있다. 이러한 모델들은 주로 과거 주가 데이터, 거래량, 기술적 지표 등을 바탕으로 미래 주가를 예측하는 데 중점을 두고 있다.

그러나, 주식 시장의 가격 변동에는 경제적 지표 외에도 심리적, 사회적 요인들이 중요한 영향을 미친다. 언론 기사와 같은 뉴스는 투자자들의 감정과 결정을 크게 좌우하며, 시장에 즉각적인 영향을 미칠 수 있다. 최근 연구에서는 뉴스와 같은 외부 정보가 시장의 변동성을 설명하는 데 중요한 역할을 할 수 있음을 보여주고 있다[1]. 특히, 뉴스 제목이나 내용에서 추출된 감성 정보는 시장의 불확실성을 높이거나, 반대로 긍정적인 기대감을 불러일으킬 수 있다.

이 연구에서는 주식 예측 모델에 뉴스 기사의 감성 분석 결과를 추가함으로써, 기존의 데이터만을 사용하는 모델과 비교하여 예측 정확도를 개선할 수 있는지 여부를 탐구하고자 한다. 이를 위해, 최신 언어 모델을 활용하여 뉴스 기사 제목을 긍정적 또는 부정적으로 분류하고, 해당 감성 점수를 주식 데이터에 통합하여 예측 모델을 구축한다. 연구의 궁극적인 목적은 언론 매체가 주식 가격 변동에 미치는 영향을 평가하고, 이를 통해 기존의 주식 예측 모델의 한계를 극복하는 데 기여하는 것이다.

2. 제안방법

본 연구에서는 주식 관련 뉴스 기사의 감성 분석을 활용하여 주가 예측 모델을 구축하였다. 이를 위해 다음과 같은 과정을 거쳤다.

Step 1. 뉴스 기사 제목 크롤링

타겟한 주식과 관련된 뉴스 기사 제목을 수집하기 위해 Python의 웹 크롤링 라이브러리인 BeautifulSoup을 사용하였다. 매일 주식 시장이 마감된 후, 해당 주식과 관련된 기사 제목 10개를 수집하였다. 각 기사의 제목은 해당 기업의 활동에 대한 평가 혹은 미래에 대한 예측을 포함하고 있다.

Step 2. 감성 분석을 통한 수치화

수집된 뉴스 기사 제목은 OpenAI의 GPT(Generative Pre-trained Transformer 4) API(Application Programming Interface)를 사용하여 감성 분석을 진행하였다. GPT 모델을 통해 각 제목의 긍정도 및 부정도를 분석하였으며, 이를 0에서 1 사이의 값으로 정량화하였다[2]. 1에 가까울수록 긍정적 의미를 나타낸다.

Step 3. 주식 데이터 수집

Python의 FinanceDataReader 라이브러리를 사용하여 주식 가격, 거래량, 고가, 저가, 종가 등의 주식 데이터와 USD/KRW 환율 및 코스피 지수를 포함한 경제 지표를 수집하였다.

Step 4. 머신러닝 모델 학습

다양한 관점에서 데이터를 분석하고, 보다 정교한 결과를 도출하기 위하여 다양한 회귀 모델을 활용하였다. 주식 데이터와 뉴스 감성 분석을 통해 얻은 긍정도 데이터를 결합한 후, 이를 예측 변수로 사용하여 다중 출력 회귀(MultiOutputRegressor)를 적용하였다.

Step 5. 모델 평가

모델 평가는 각 머신러닝 알고리즘이 예측한 주가 변동과 실제 주가 변동 간의 부호 일치 여부를 기준으로 이루어졌다. 구체적으로, 모델이 예측한 상승, 하락, 변동 없음과 실제 주가의 변동 방향을 비교하여 30일 동안의 예측 성공률을 계산하였다. 상승과 하락에 대한 예측 정확도를 각각 측정하였으며, 각 모델의 성능은 예측 성공 횟수를 기준으로 평가되었다. 마지막으로, 이러한 예측 정확도는 모델별로 비교 분석되었다.

3. 실험 및 결과

본 연구에서는 한국 주식 데이터를 바탕으로 주식 예측 모델을 구축하였으며, 주식 종목으로는 삼성전자를 대상으로 하였다. 감성 분석을 포함한 경우에는 뉴스 기사 제목에서 추출한 긍정도 점수를 추가하여 학습하였다. 긍정도 점수는 매일 기사 제목에서 GPT API를 통해 추출된 감성 정보를 바탕으로 주가와 관련된 투자자 심리를 정량화한 수치로, 0에서 1까지의 범위를 가진다.

수집된 데이터는 최근 31일을 기준으로 훈련 데이터와 테스트 데이터로 분할하였다. 학습 데이터는 가장 최근의 데이터를 제외한 모든 데이터를 사용하였으며, 가장 최근의 1일 데이터를 테스트 데이터로 사용하였다. 이를 통해 매일 새롭게 수집된 데이터를 테스트하며 모델의 성능을 평가하였다.

모델의 성능은 주가의 상승, 하락 예측 정확도를 기준으로 평가하였다. 예측된 주가가 실제 값과 동일한 방향으로 변화(상승 또는 하락)했는지 여부를 판단하여 정확도를 계산하였다. 상승 예측 성공률, 하락 예측 성공률을 각각 평가하였다.

또한, 예측 정확도는 최근 30일 동안의 결과를 바탕으로 측정하였으며, 각 모델의 예측 성능을 비교하였다. 감성 분석을 포함한 경우와 포함하지 않은 경우의 결과를 각각 분석하여 감성 분석이 주가 예측에 미치는 영향을 확인하였다.

감성 분석을 포함한 경우 대부분의 모델에서 상승 예측 정확도가 향상되었으며, 특히 GradientBoostingRegressor 모델이 상승 및 하락 예측 모두에서 가장 높은 성능을 보였다. 하락 예측 정확도는 감성 분석의 유무에 크게 영향을 받지 않았으나, 감성 분석을 포함한 모델이 전반적으로 더 높은 전체 정확도를 기록하였다.

결과적으로, 감성 분석을 포함한 학습은 주가 예측 성능을 개선하는 데 긍정적인 영향을 미치며, 특히 상승 예측에서 그 효과가 두드러짐을 확인할 수 있었다.

3.1 상승 예측 정확도 (Rise_pred Accuracy)

상승 예측을 위해 긍정도를 포함하여 학습한 결과, GradientBoosting이 72.7%로 가장 높은 정확도를 기록했으며, 나머지 모델들은 대체로 40-50%대의 상승 예측 정확도를 보였다. 반면, 긍정도를 포함하지 않은 경우 상승 예측 정확도가 전반적으로 낮아졌으며, RandomForest는 44.4%로 상대적으로 높은 성능을 보였지만, Linear 모델은 28.57%로 크게 감소하였다. 따라서 긍정도를 포함한 학습이 상승 예측에서 전반적으로 더 나은 성능을 보이는 것으로 나타났다.

3.2 하락 예측 정확도 (Fall_pred Accuracy)

하락 예측에서도 긍정도를 포함한 학습 결과, GradientBoosting이 73.68%로 가장 우수한 정확도를 기록했으며, 다른 모델들도 대체로 50-60%대의 하락 예측 정확도를 보였다. 긍정도를 포함하지 않은 경우, 하락 예측 정확도는 전반적으로 약간 감소했으나 비교적 유지되었으며, GradientBoosting과 RandomForest가 58.33%로 높은 성능을 유지했다. 이를 통해 하락 예측에서도 긍정도를 포함한 경우 상대적으로 우수한 성능을 보인다는 결론을 도출할 수 있다.

Model Name	긍정도 포함		긍정도 미포함	
	상승 예측	하락 예측	상승 예측	하락 예측
Linear	0.0	53.33	28.57	43.75
Random Forest	53.33	66.67	42.86	56.25
LGBM	43.75	57.14	44.44	58.33
KNeighbors	53.85	64.71	42.11	54.55
DecisionTree	41.18	53.85	46.15	62.5
GradientBoosting	72.73	73.68	44.44	58.33
XGBoost	50	64.29	46.67	60.0

4. 결론 및 향후 연구

본 연구에서는 최신 자연어(Natural Language Processing, NLP) 처리 모델인 GPT-4o를 활용하여 뉴스 기사 제목의 감성 분석을 기반으로 주가 등락 예측 모델을 구축하였다. 뉴스 기사 제목의 감성을 정량화하여 주식 시장 데이터에 결합함으로써, 단순 시계열 데이터를 사용한 기존 모델에 비해 예측 성능이 향상됨을 확인할 수 있었다. 특히 상승 예측에서 감성 분석이 주가 예측의 성능을 크게 향상시키는 결과를 보였으며, 이는 뉴스가 투자자 심리와 시장 방향성을 잘 나타냄을 알 수 있다. 다만 본 연구는 특정 기업(삼성전자)의 주식 데이터만을 분석하여 다른 기업이나 산업에 일반화되기에는 한계가 있을 수 있다. 또한 기사 제목만으로 분석을 수행하는 것은 정보가 제한적일 수 있다. 향후 딥러닝 모델로 확장하여 더욱 정교한 주가 예측 모델을 구축할 예정이다.

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1C1C2011105). 본과제(결과물)는 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다. (2021RIS-004)

참고 문헌

[1] Trang-Thi Ho, Yennun Huang, "Stock Price Movement Prediction Using Sentiment Analysis and CandleStick Chart Representation," Sensors, vol. 21, no. 23, 7957, 2021.
 [2] Khedr, A. E., Yaseen, N., "Predicting stock market behavior using data mining technique and news sentiment analysis," International Journal of Intelligent Systems and Applications, vol. 9, no. 7, 22, 2017.