

# 모회사 뉴스 감성을 이용한 딥러닝 기반

## 자회사 주가 등락 예측

이호정\*, 손명균\*, 황병현\*, 윤고은\*, 전수현\*

### Deep learning based subsidiary stock price fluctuation forecast using parent company news sensitivity

Lee Ho Jung\*, Son Myeong Gyun\*, Hwang Byeonghyeon\*, Youn Go Eun\*, and Jeon Su Hyeon\*

#### 요 약

본 연구는 모회사의 뉴스 감성이 자회사의 주가 등락에 영향을 주는지 파악하기 위해 네이버 뉴스를 수집하여 분석하였다. 뉴스의 감성을 처리하기 위해 KoELECTRA를 사용하였으며, 주가 등락 예측 모델로는 LSTM을 사용하였다. sequence size별 예측을 진행했으며 분석 결과, 모회사의 뉴스가 주가 등락에 영향을 끼쳤음을 확인하였다.

#### Abstract

This study used to KoELECTRA to collect and analyze Naver news to understand whether the sensitivity of parent company news affects the stock price of subsidiaries, and used LSTM as a predictive model of stock price fluctuation. We made predictions by sequence size, and as a result of the analysis, we confirmed that the news of the parent company affected the stock price.

#### Key words

stock price fluctuation, KoELECTRA, LSTM, sensibility analysis

---

\*계명대학교, ( [ghwjd5684@kmu.kr](mailto:ghwjd5684@kmu.kr), [dmz38s1217@gmail.com](mailto:dmz38s1217@gmail.com), [grant0808@naver.com](mailto:grant0808@naver.com),  
[hhaneul0731@naver.com](mailto:hhaneul0731@naver.com), [djqtekrh4@gmail.com](mailto:djqtekrh4@gmail.com) )

※ 이 논문은 2022년도 계명대학교의 교육혁신팀의 빅데이터 사업부의 지원을 받아 수행된 연구임.

## 1. 서 론

주가는 기업의 가치를 반영하고 있는 대표적인 지표이다. 주가의 등락은 근본적으로 기업의 영업실적에 달려있다는 사실은 농후하다. 투자자가 투자할 기업을 선정할 기준으로 그 기업에 대한 정보를 알아야 선정할 수 있으며, 투자자들이 가장 쉽게 기업의 정보를 접할 수 있는 매체가 뉴스이다. 뉴스를 통해 기업에 대한 정보들을 파악할 수 있고, 이러한 정보들은 투자자들의 투자로 이어지게 된다.

이러한 배경을 기반으로, 비정형데이터가 주식시장에 영향을 미치는 선행연구가 입증됨에 따라 관련된 연구가 활발하게 진행되고 있다. 그러나 선행연구에 따르면, 모회사의 영향력이 자회사에 미치는 연구가 이루어지지 않고 있다.

따라서, 본 연구는 모회사의 뉴스가 자회사의 주가에 영향을 주는가를 파악하고자 한다.

## II. 분석을 위한 데이터 수집

본 논문에서는 감성분석을 활용하여 주가 등락의 영향을 파악하고자 한다. 주가 예측을 위해서는 기업의 주가 정보와 더불어 기업의 감성 지표를 활용하기 위해 뉴스 데이터를 수집한다.

데이터의 수집 기간은 2020년 1월 1일부터 2021년 12월 31일까지 총 730일이다. 그림 1과 같이 코로나 팬데믹 이후 코스피 지수는 약 3개월간의 폭락 이후 꾸준한 상승세를 보인다.[1] 이 기간 동안 코로나 및 주가와 관련된 뉴스가 많이 나왔을 것이라는 예상하에, 해당 기간을 선정하게 되었다.

그림 1. 코스피 차트

Fig. 1. KOSPI chart



KOSPI 상위 20개의 기업 중, 2020년 1월 1일 이전 상장된 기업과 자회사가 2개 이상 있는 모회사를 선정한 결과 셀트리온, POSCO홀딩스, SK로 선정되었다. 이때, 각 회사의 자회사 선정으로는 시가총액 상위 2개의 기업을 채택했다. 이에 해당하는 자회사는 셀트리온 헬스케어(이하 heal), 셀트리온 제약(이하 medi), POSCO케미칼(이하 chem), POSCO 인터내셔널(이하 inter), SK하이닉스(이하 hynix), SK이노베이션(이하 inno)이 선정되었으며, 전날 대비 주가가 증가했을 경우 True, 그 외의 경우를 False로 나타낸 결과, 각 회사별 등락률의 비율은 표 1. 와 같다.

표 1. 회사별 등락비율 (단위: %)

table 1. Rate of Increase Per Company (%)

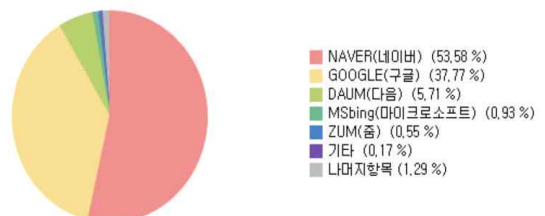
비율	heal	medi	hynix	inno	chem	inter
True	54.1	51.6	52.6	53.0	53.8	56.4
False	45.9	48.4	47.4	47.0	46.2	43.6

본 논문에서는 총 6개 기업을 연구 대상으로 선정한다. 2020년 1월 1일부터 2021년 12월 31일 포털 사이트 사용률 확인 결과[2] 그림 2와 같이 NAVER가 50% 이상을 차지하는 것을 알 수 있다. 따라서, NAVER 뉴스의 총 12개 일간지별 뉴스를 크롤링한다. 모회사와 자회사 간의 중복된 데이터를 삭제 결과, 셀트리온, POSCO홀딩스, SK 각각 2,269개, 7,964개, 39,658개로 나타났다.

그림 2. 포털 사이트 사용률

Fig. 2. portal site usage rate

Total : 177,048,093



## III. 데이터 분석 과정

학습데이터로 AI Hub에 있는 감성 대화 문치 데이터를 사용하여 KoELECTRA를 학습시켰다. KoELECTRA 모델은, 2020년 google에서 발표된 transformer 기반 모델인 ELECTRA 모델을 한국어로 학습한 모델이다.[3] 2020년에 학습데이터의 감성에서 불안, 슬픔, 당황, 상처, 분노를 부정으로 기쁨을 긍정으로 변환시킨다. 기사 본문에서 불필요한 문단을 제거하고자 회사 언급이 있는 문단을 시작으로 256개를 추출한다. 학습된 KoELECTRA를 사용하여 뉴스데이터 감성분석을 시행한다. 추출된 감성은 긍정과 부정에 대한 확률값을 가지게 된다. 도출된 결과를 딥러닝에 적용하여 유의미한 정보를 얻기 어려운 문제가 있어 변화시킬 필요가 있다. 따라서, 범위 (0, 1)을 가진 확률값을  $(-\infty, \infty)$ 로 변환하기 위해  $\logit$  function(1)을 활용한다. 날짜별 뉴스의 수에 변화가 있어 중요도를 적용 시키기 위해 전체 기간의 뉴스 수에서 해당 일자에 작성된 뉴스 수를 나누어 가중치를 만들었다(2). 최종적으로 가중치에서  $(-\infty, \infty)$ 로 변환한 확률값을 곱하여 딥러닝 입력값으로 사용했다. 즉, 최종으로 나온 감성 값은 (3)과 같다.[1]

$$\logit = \ln\left(\frac{positive}{1-positive}\right) \quad (1)$$

$$intensity_t = \ln\left(1 + \frac{n_t}{avg\ of\ N}\right) \quad (2)$$

$$sentimental\ index_t = \logit \times intensity_y \quad (3)$$

모회사의 뉴스 감성을 이용한 딥러닝 기반 자회사 주가 등락 예측을 위해 LSMT(Long Short-Term Memory) 모델을 통해 분석한다. LSTM 모델은 ANN(Artificial Neural Network) 기반으로 한 RNN의 단점인 경사 소실(Vanishing Gradient Problem)을 보완한 모델이다.[3]

LSTM모델을 학습하기 위해 수집한 데이터 중 2020년 1월부터 2021년 10월까지 train set, 나머지 데이터를 test set으로 사용하였다. 종가의 가중치를 줄이기 위해 MinMaxScaler 하였다. LSTM 설정을 500 epoch, 손실 함수로는 Mean Square Error를 사용하여 trian 및 test를 시행하였다.

표 2은 sequence size에 따른 LSTM 모델에서 종

가만을 사용한 모델의 정확도를 나타낸다. 각 자회사 기준, 가장 높은 정확도는 medi의 0.586, 가장 낮은 정확도는 inter의 0.476으로 두 회사 간의 정확도 차이는 0.11이다.

표 3는 sequence size에 따른 LSTM 모델에서 종가와 해당 회사의 모회사의 감성지수를 사용한 모델의 정확도이다. 가장 높은 정확도는 chem의 0.549, 가장 낮은 정확도는 inno의 0.476으로 두 회사간의 정확도 차이는 0.073으로, 종가만을 사용한 경우보다, 종가와 감성지수를 활용했을 때, 편차가 적다는 것을 확인할 수 있다.

표 4은 감성지수 활용 유무에 따른 정확도 차이를 나타낸 것이다. 전체적인 정확도는 0.002 증가했음을 알 수 있다.

표 2. 종가를 활용한 모델 정확도

table 2. Model Accuracy Using Closing Price

size	heal	medi	hynix	inno	chem	inter
5	0.545	0.472	0.527	0.472	0.527	0.509
10	0.519	0.66	0.46	0.439	0.56	0.419
20	0.4	0.625	0.524	0.55	0.55	0.5

표 3. 종가와 뉴스 감성을 활용한 모델 정확도

table 3. Model Accuracy Using Closing Price and news sensitivity

size	heal	medi	hynix	inno	chem	inter
5	0.545	0.545	0.509	0.472	0.581	0.509
10	0.5	0.419	0.439	0.38	0.54	0.579
20	0.55	0.675	0.5	0.575	0.525	0.449

표 4. 감성지수 활용 유무에 따른 정확도 차이

table 4. Accuracy Difference According to the Use of Emotional Index

size	heal	medi	hynix	inno	chem	inter
5	0	0.073	-0.018	0	0.054	0
10	-0.019	-0.241	-0.021	-0.059	-0.02	0.16
20	0.15	0.05	-0.024	0.025	-0.025	-0.051
avg	0.044	-0.039	-0.021	-0.011	0.003	0.036

## IV. 결 론

본 연구는 모회사의 뉴스 기사가 자회사의 주가 등락에 영향을 주는가를 알아보기 위해 LSTM과 KoELECTRA를 사용하여 분석하였다.

2년간의 주가 데이터를 사용하였으며, 정확도 확인 결과, 종가를 사용하여 주가등락을 예측했을 때 보다, 종가와 모회사 감성을 사용하여 주가등락을 예측했을 때의 평균 정확도가 증가한다는 결과가 도출되었다.

본 연구는 주가 관련 논문에서 시행한 연구들과는 달리, 모회사의 뉴스 기사가 자회사의 주가에 어느 정도의 영향을 줄 수 있는지에 대한 의문점에서 시작했다는 차별점을 가지고 있다. 그러나 모델학습에서 뉴스 수의 요소가 크게 작용하는 점과 주가 지표 중 종가만을 사용했다는 한계점이 존재한다. 향후 연구에서는 뉴스 수와 추가적인 주가 관련 변수를 고려할 필요가 있다.

## 참 고 문 헌

- [1] 강두원, 유소엽, 이하영, 정옥란. (2022). 뉴스 감성 분석을 이용한 딥러닝 기반 주가 예측에 대한 연구. 한국컴퓨터정보학회논문지, 27(8), 31-39.
- [2] "InternetTrend", InternetTrend. last modified Dec 31, 2021, accessed Nov 11, 2022, '<http://www.internettrend.co.kr/trendForward.tsp>'
- [3] Clark, Kevin, et al. "Electra: Pre-training text encoders as discriminators rather than generators." arXiv preprint arXiv:2003.10555 (2020).
- [4] 한태동. (2021). LSTM을 이용한 주가 예측: 기술 지표, 거시 경제 지표, 시장 심리의 조합을 중심으로. 융복합지식학회논문지, 9(4), 189-198.