

주가 예측을 위한 Web Site 개발

조규철*, 이성희^o

*인하공업전문대학 컴퓨터정보과,

^o인하공업전문대학 컴퓨터정보과

e-mail: kccho@inhac.ac.kr*, lee98417716@gmail.com^o

System Development of the Stock Price Prediction

Cho Kyu Cheol*, Lee Sung Hee^o

*Dept. of Computer Science, Inha Technical College,

^oDept. of Computer Science, Inha Technical College

● 요약 ●

주식을 매매할 때, 주식의 차트와 가치를 분석한 다음 언제 주식이 상승가 또는 하락가가 될지 예측한 후 매매하게 된다. 하지만 일반적으로 주식을 예측하기 어려워 주식의 수익을 내기 힘들다. 따라서 본 논문은 지난날의 주식 가격 데이터를 분석해 주식의 가격을 예측하는 주식 차트 분석을 할 수 있게 ‘주가 예측을 위한 웹 사이트’를 개발하였다. 이 사이트는 주식의 차트 분석을 지원하고 주식을 언제 매매할지에 대한 의사결정을 도와줄 수 있을 것으로 기대된다.

키워드: 주가 예측(prediction of Stock Price), LSTM, 평균제곱오차(Mean Squared Error, MSE)

I. 서론

주식거래를 위한 인공지능 투자 플랫폼들이 많이 늘어났다. 이런 플랫폼들은 보통 인공지능을 이용하여 사용자들에게 주식 추천을 해주거나 자동으로 주식을 매매해준다. 주식을 추천해주는 플랫폼들은 사용자 관리 카테고리가 대부분 없어 주식을 관리하기 불편하고 사용자 관리 카테고리가 있는 플랫폼들은 자동매매 플랫폼이라서 주식을 매매할지 자신이 결정하기 힘들다. 따라서 이러한 불편한 점을 개선하여 사용자가 매매한 주식을 관리해주는 카테고리가 포함된 주식 가격 예측 웹사이트를 개발하고자 한다.

II. 주식 증가 예측 절차

1. 주식의 증가

정규시간에 여는 주식시장을 정규장이라고 하고 증가는 정규장에서 마감될 때 가격을 말한다. 보통 전날의 증가와 당일 증가를 비교해 투자해서 증가를 예측 데이터로 선택하였다.

2. 데이터 분석과 전처리

	Date	High	Low	Open	Close	Volume	Adj Close
0	2015-11-09	26880.0	26420.0	26760.0	26880.0	9283050.0	271.287384
1	2015-11-10	26820.0	26280.0	26720.0	26420.0	9877550.0	266.644897

Fig. 1. samsung electronics 주가 데이터[2]

데이터로는 ‘Yahoo Finance[1]’의 데이터를 사용한다. 데이터는 주식별 날짜에 따른 주식의 고가(High), 저가(Low), 시가(Open), 종가(Close), 거래량(Volume), 수정 종가(Adj Close)를 보여준다. 사용한 데이터는 ‘삼성전자’의 최근 5년간 데이터[2]이다.

데이터를 분석에 적합하도록 종가(Close)값만 선택하였고 10일간의 데이터를 기반으로 그다음 날의 종가를 예측하는 모델을 만들기 위하여 슬라이딩 윈도우(sliding window) 방법을 사용한다. 고정된 크기를 정해서 데이터를 묶어준 것을 윈도우라고 한다. 윈도우를 만들 때 그림2와 같이 한 칸씩 밀리게 여러 윈도우들을 만들어준다. 이렇게 윈도우들을 학습시키는 방법을 슬라이딩 윈도우라고 한다. 본 논문에서는 그림2처럼 11개로 고정 크기를 정하고 하나의 윈도우에서 10개의 데이터는 독립변수로, 마지막 데이터는 종속변수로 설정해준다.

	Open	High	Low	Close
2018-10-01				
2018-10-02				
2018-10-20				
2018-10-21				

Fig. 2. windows 설정

손실률이 낮은 모델을 만들기 위해서 각각의 윈도우 안에서 표준화를 해주어야 한다. 표준화 방법은 식의 표준화 공식을 이용해 주었다.

$$X' = \frac{X - \mu}{\sigma} \quad \text{식(1)}$$

X' : 표준화 값, X : 데이터 값
 σ : 표준편차, μ : 평균

3. 데이터 학습

학습에는 LSTM(Long Short-Term Memory, 장단기 메모리)을 이용한다. LSTM은 RNN(Recurrent Neural Network, 순환신경망)의 정보량 소실을 해결하기 위해 나온 신경망으로 앞에서 기억된 값을 다음 층으로 넘길지 결정하여 넘기게 된다. 본 논문에서는 모델 성능을 측정하기 위해 MSE(Mean Squared Error, 평균제곱오차)를 사용하였다. MSE는 실제 값과 예측값의 차이를 의미하고 0에 가까울 수록 좋다.

전체 데이터의 총 1236개 중 8:2 비율로 나누어 학습과 테스트를 함께 진행해주었다.

4. 데이터 학습 결과

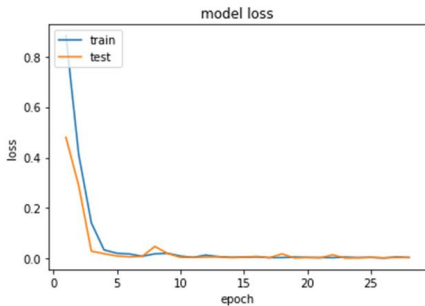


Fig. 3. 학습과 테스트의 손실률 그래프

그림 3은 matplotlib[3]으로 표현한 학습과 테스트의 학습 반복 횟수에 따른 손실률을 그래프로 나타낸 결과고 이를 활용한 실제 가격과 예측한 가격의 차이를 보여주는 그래프로 실제 가격에 수렴하였다.

III. 주식 가격 예측 웹사이트

웹 사이트는 Python 언어 기반인 Flask 웹 어플리케이션 프레임워크를 사용하여 웹 개발을 진행한다. 웹 페이지는 로그인 페이지, 로그인되었을 때만 보여주는 매입한 주식 페이지와 관심 있는 주식

페이지가 있다. 매입한 주식 페이지와 관심 있는 주식 페이지에서 주식을 선택 후 제출하면 두 페이지 다 선택한 주식의 예측한 가격과 Candle Chart를 보여준다.

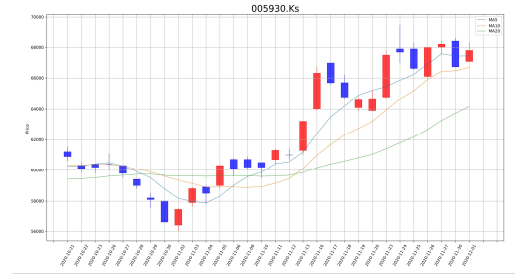


Fig. 4. matplotlib[3]으로 표현한 Candle Chart

IV. 결론

본 연구를 통해 사용자가 매입한 주식과 사용자가 관심 있는 주식으로 등록해놓은 주식 종가의 예측된 값을 알 수 있어 사용자가 좀 더 주식투자에서 수익을 낼 수 있을 것이다.

REFERENCES

- [1] Stock Data, <https://finance.yahoo.com/>
- [2] Stock Data of Samsung Electronics(005930.KS), <https://finance.yahoo.com/quote/005930.KS/history?p=005930.KS>
- [3] Matplotlib plot, <https://matplotlib.org/>