

LSTM을 활용한 나스닥 차트 예측 모델 구현

NASDAQ Chart Prediction Model Implementation Using LSTM

변운성, 백호빈

Department of Physics and Astronomy, Seoul National University, Seoul 08826, Korea

E-mail: may670@snu.ac.kr, ghqlsd198@snu.ac.kr

Abstract.

주식의 차트를 완벽하게 분석하는 것은 모든 투자자들의 꿈일 것이다. 차트를 분석하는 방법에는 거시경제의 분석 방법도 있고 보조지표 등을 이용한 기술적인 분석도 있지만, 최근에는 인공지능의 발전으로 인하여 딥러닝을 이용한 투자 전략을 찾는 시도가 늘어나고 있다. 이러한 추세에 따라, 본 연구에서는 시계열 데이터를 다루는 인공지능망인 LSTM[1]을 이용해 나스닥의 차트를 학습시키는 연구를 수행한다. 본 연구에서 사용되는 모델은 비교적 기초적인 모델이기에 관측되는 여러 가지 한계점도 명확하였지만, 보조지표 등을 추가한 자신만의 합리적인 투자전략을 추가한다면 유의미한 예측 결과를 내놓을 수 있을 것으로 기대한다.

(Dated: 16 June 2022)

1. Introduction

최근에 인공지능이 급격하게 발전하면서, 그 활용도는 어마무시하게 많아지고 있다. 자율주행 자동차부터, 의료분야의 MRI와 심지어 투자에까지 인공지능을 사용해 능력을 개선하려는 노력이 행해지고 있다. 그 노력의 일환으로, 본 연구에서는 시계열 데이터를 다루는 인공지능망인 LSTM(Long-Short Term Memory model)과 나스닥 차트를 이용하여 과거 데이터로부터 미래의 주가를 예측하는 것이 과연 유의한지를 검증해보고자 하였다. 본 연구에서는 나스닥의 과거 약 20년간의 차트를 이용해 학습시키고, 최근 5년간의 데이터를 이용해 유효한지를 살펴보았다.

2. Experimental

2.1. Method & Result

인공지능 학습 코드는 기본적으로 TA3에서 제공된 Many-to-One 구조를 가진 LSTM 코드를 사용하였으며, LSTM을 이용해 학습시킨 결과는 Figure 1과 같다. :

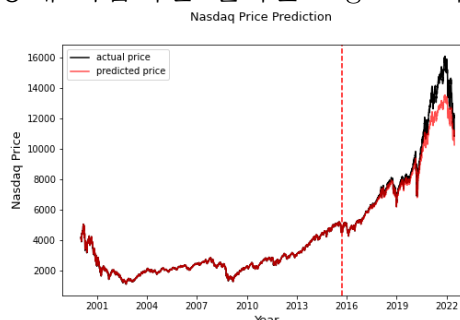


Figure 1: NASDAQ prediction/actual price

학습을 위한 데이터는 Yahoo Finance에서 불러온 NASDAQ의 차트 데이터 중 수정된 종가를 사용하였다. 즉, 전날 X일 동안의 수정된 종가의 가격변화를 이용해 다음날의 수정된 종가를 예측하였다. Figure 1을 살펴보면 2020~2022년 근방의 호황기에는 변동성이 너무 커서 인공지능의 가격변화량 예측이 실제 변화량을 따라가지 못했기 때문에, 가격의 변동폭을 대신하여 다음날 가격의 상승/하락 예측의 유효성에 대해서만 알아보았다. 때문에, 인공지능의 예측 결과를 실제 수익률과 연관시키기 위해서는 또다른 매수/매도 원칙을 적용시켜야 할 것이다.

Test data로 사용된 2015-09부터 현재까지의 나스닥 종가 데이터에서, 다음날 주가가 상승할 것이라고 예측했을 때 상승/하락했을 경우, 하락할 것이라고 예측했을 때 상승/하락했을 경우를 나타낸 것은 Table 1과 같다.

	실제		
		상승	하락
		예측	예측
예측	상승	544	442
	하락	409	295

Table 1: 나스닥의 가격 상승/하락 예측과 실제 변화 비교

Test data 기간 내에서 나스닥의 종가가 전날의 종가에 비해 상승한 날의 비율은 56.4%로, 전체적으로 우상향하는 장이었음을 볼 수 있다. 다음날 주가가 상승할 것으로 예측한 경우와 하락할 것으로 예측한 경우, 예측이 실제 가격의 움직임과 일치할 확률은 각각 55.2%, 41.9%였다. 상승할 경우에는 50%가 넘는 승률을 기록했지만, 전체 기간에서의 나스닥이 오를 확률인 56.4%를 유의미하게 넘지는 못하는 것으로 밝혀졌다. 또한, 상승/하락으로 예측하는 각각의 경우에 대해서 유의미한 확률의 차이가 있는지를 알아보기 위해 카이제곱검정을 수행한 결과, P-value는 0.25208로, 상승/하락 예측

여부에 따른 실제 변화 방향의 차이는 유의미하게 나타나지 않는 것으로 판단되었다.

2.2. Error Analysis

본 연구에서 유의미한 결과가 나오지 않은 이유에 대해서 생각해 본 결과, 몇 가지 문제점을 생각해볼 수 있었다.

첫번째로, 주식 차트는 랜덤성이 짙어서 예측할 수 없는 random walk를 한다는 가정이 사실이라면, 시계열 데이터를 sequence로 LSTM에 대입하는 것은 의미가 없게 된다. 이전 X일의 가격의 변화가 미래 가격변화의 방향을 담보해줄 수 없기 때문이다. 하지만 이러한 가정은 차트 무용론자와 차트 분석가들 간의 의견차이가 분분한 주제이기 때문에, 여기서는 고려하지 않도록 한다.

두번째로, 주식 가격은 과거의 가격보다는 다른 외생변수들과 더욱 큰 연관성이 있다는 가정이다. 실제로 매크로 시장경제가 돌아가는 상황은 아주 복잡하다. 때문에, 단순히 가격만을 보고 예측하기를 시도하기 보다는, 뉴스, 물가, 거래량, SNS 언급량 등의 지표들 또한 주가 변화 예측 모델에 대입해 준다면, 더욱 정확하게 예측할 수 있을 것으로 생각한다.

마지막으로, LSTM에 넣은 train data가 stationary하지 않다는 점이다.[2] 나스닥 주가 차트는 지난 20년이 넘게 꾸준히 우상향했다는 점에서 stationary하지 않은데, 이는 기존 데이터 분석에서 사용하는 통계량을 나스닥 차트에서는 적용하지 못한다는 것을 뜻한다. 이를 보정하기 위해서는, 미래 주가의 가격을 예측하는 것이 아니라, 상승/하락을 1,0으로 라벨링을 하고, loss function을 cross-entropy 함수를 사용하는 등의 방법을 사용할 수 있을 것이다.

3. Conclusion

본 실험에서는 LSTM을 사용해 시계열 데이터인 나스닥 차트의 가격을 예측하

는 시도를 해 보았다. 그 결과, 나스닥의 전체 기간 동안에서 상승할 확률은 56.4%인데 반해, 예측의 승률은 55.2%, 41.9%로, 나스닥의 상승확률을 넘기지 못한 것으로 밝혀졌다. 비록 유의미한 나스닥의 상승확률을 넘기는 유의미한 승률을 기록하진 못 했지만, 인공지능을 투자에 적용시켜본다는 점에서 의미있는 시도였을 것으로 기대한다. 또한, 보조지표와 알파-전략으로 불리는 자신만의 추가적인 투자전략을 덧붙인다면, 충분히 매력적인 모델을 만들어낼 수 있을 것이라고 기대한다.

References

- [1] Sepp Hochreiter; Jürgen Schmidhuber (1997). "Long short-term memory". Neural Computation.
- [2] Boashash, B. (ed.), (2003) Time-Frequency Signal Analysis and Processing: A Comprehensive Reference, Elsevier Science, Oxford, 2003