

Reverse Low-High 라벨링 방법을 이용한 XGBoost 트레이딩 시스템 개발

한예찬, 박태신, 김재윤*
순천향대학교

ychan.sch@gmail.com, ptaesin@gmail.com, kimym38@sch.ac.kr*

Developing XGBoost Trading System using Reverse Low-High Labeling

Yechan Han, Taeshin Park, Jaeyun Kim*
Soonchunhyang Univ.

요 약

본 논문에서는 트레이딩 시스템 개발 시 데이터 라벨링의 중요성에 대해 논의한다. 기존의 Up-Down 라벨링 방법은 주가의 등락 정도를 파악하지 못한다는 한계점이 있다. 본 논문에서는 데이터 라벨링 시 Reverse Low-High 방법을 사용하며, KOSPI 200 과 S&P 500 주가 지수로 실증 분석을 진행한다. 실증 분석 결과 제안된 방법은 기존의 Up Down 방법보다 우수한 성과를 나타내는 것을 확인하였다.

1. 서 론

주식 투자에 대한 관심이 높아지고, 주가 데이터에 대한 접근이 쉬워지면서 머신러닝 기술을 활용한 다양한 투자 방법이 제안되고 있다. [1] 머신러닝 기술을 활용한 투자 방법은 주로 과거의 주가 데이터를 이용하여 매매 규칙을 발견하고 매매 규칙에 따라 백테스팅을 수행하여 검증하는 절차를 거친다. 주가의 상승과 하락을 예측하는 분류 문제의 경우, 이 과정에서 주가 데이터를 라벨링 하는 작업이 수행된다. 많은 연구에서는 주가 데이터 라벨링 시 Up-Down 라벨링 방법을 사용한다. [2][3] Up-Down 라벨링 방법은 $t+1$ 시점의 주가가 t 시점의 주가보다 오르면 Up, 내리면 Down 으로 라벨링을 수행한다. 이러한 라벨링 방법은 단순히 주가의 등락만을 구분하며, 등락 정도를 파악하지 못하기 때문에 이후 머신러닝 모델을 사용한 예측 과정에서 문제가 발생할 수 있다. 본 논문에서는 데이터 라벨링 작업 시 Sezer 가 제안한 라벨링 방법(본 논문에서는 해당 라벨링 방법을 Low High 라벨링 방법이라 정의)을 변환한 Reverse Low High 라벨링 방법을 사용하며, 기술적 지표와 XGBoost 모델을 활용하여 트레이딩 시스템을 구축한다. [4]

2. 본 론

2.1. 실험 데이터

본 논문에서는 KOSPI 200 과 S&P 500 일별 주가 지수 데이터를 사용하였으며, 2012 년 1 월 1 일 ~ 2017 년 12 월 31 일 기간 (6 년)의 데이터를 학습 기간으로, 2018 년 1 월 1 일 ~ 2020 년 12 월 31 일 기간 (3 년)의 데이터를 테스트 기간으로 설정하였다.

2.2. 실험 방법

- 데이터에서 기술적 지표를 추출하고 데이터 라벨링 작업을 수행하여 학습 데이터를 구축한다.

- XGBoost 모델을 사용하여 학습 데이터로 모델을 학습한다. (학습 데이터에서는 Up 과 Down 신호만 사용하며 Hold 신호는 사용하지 않는다)
- 학습된 XGBoost 모델을 사용하여 테스트 데이터를 예측하며, 예측 값으로 매매 신호를 생성하여 거래를 수행한다.

2.3. Reverse Low High 라벨링

본 논문에서는 기존의 Low High 라벨링 방법을 변환한 Reverse Low High 라벨링 방법을 제안한다. Reverse Low High 라벨링 방법은 아래와 같으며, 샘플 데이터로 라벨링을 수행한 결과는 Figure 1 과 같다.

- window size 를 설정한다. (본 연구에서는 window size 를 31 로 정하여 실험한다)
- 한 칸씩 이동하면서 window size 기간의 V 모양 최저점과 ^ 모양 최고점을 찾는다. 예를 들어 window size 의 크기가 5 라면 3 번째 날(가운데 위치한 기간)의 주가가 5 일의 기간 중 최저가인지 최고가인지를 확인한다.
- 최저점은 Down 으로 최고점은 Up 으로 설정하며, 나머지 부분은 모두 Hold 로 설정한다.

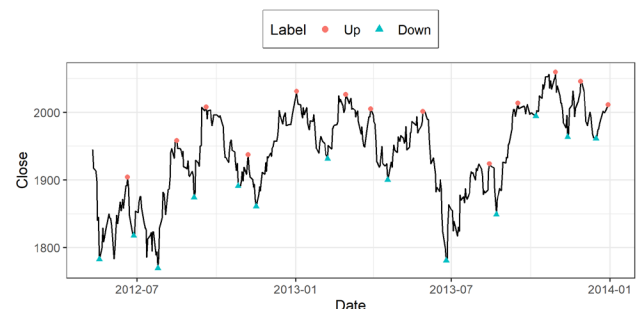


Figure 1 Reverse Low High 라벨링 결과

2.4. 매매 신호 생성

XGBoost 모형으로 예측한 값을 임계값 0.5 기준으로 기준 값 보다 작으면 Up, 기준 값 보다 크면 Down 으로 변환한다. 다음으로 변환한 값에서 처음 발생하는 Up 과 Down 을 찾아 각각 Buy 와 Sell 신호로 변경한 다음, 나머지 연속되는 신호는 모두 Hold 로 처리한다. 거래는 모두 1 주 단위로 진행되며, Buy 신호가 발생한 경우 다음날 시가로 매수하고 Sell 신호가 발생한 경우 다음날 시가로 매도한다. Figure 2, 3 은 샘플 데이터의 매매 신호 생성 전, 후를 보여준다.

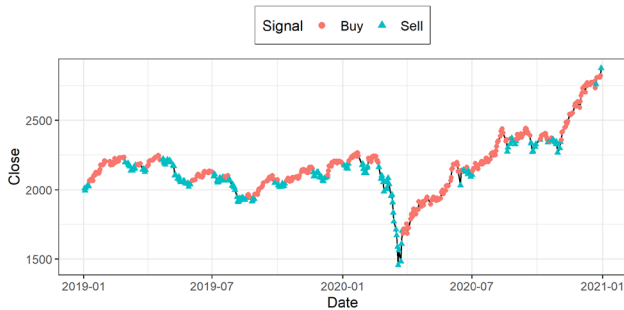


Figure 2 매매 신호 생성 전

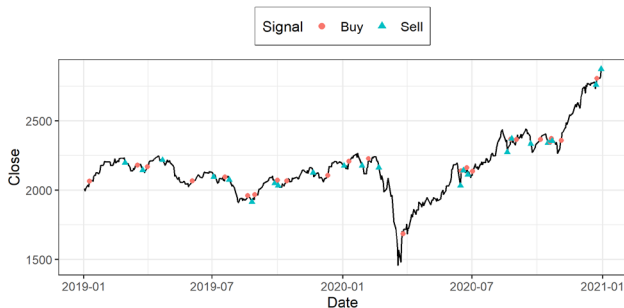


Figure 3 매매 신호 생성 후

2.5. 실험 결과

Table 1 은 KOSPI 200 주가 지수로 실험한 결과로 Up-Down 방법과 기존의 Low High 방법으로 라벨링을 수행하는 경우 누적 수익 값이 음수가 되어 포인트 손실이 발생하는 것을 확인 할 수 있다. 반면 Reverse Low-High 라벨링의 경우 778 포인트의 수익이 발생하는 것을 볼 수 있다. KOSPI 200 주가 지수의 경우 테스트 기간에 매수 후 보유(Buy&Hold) 전략을 취하는 경우 345 포인트의 수익이 발생하는데 이에 비해 2 배 이상의 포인트 수익이 발생한 것을 확인할 수 있다. Table 2 는 S&P 500 주가 지수로 실험한 결과로 KOSPI 경우와 마찬가지로 Up-Down 방법과 Low-High 방법으로 라벨링하는 경우 포인트 손실이 발생한다. S&P 500 주가 지수는 테스트 기간에 매수 후 보유 전략을 취하는 경우 1049 포인트의 수익이 발생하는데, Reverse Low-High 라벨링의 경우 KOSPI 경우와 마찬가지로 Buy & Hold 전략에 비해 2 배 이상의 포인트 수익을 내는 것을 확인할 수 있다. Up-Down 라벨링 방법은 KOSPI 200 과 S&P 500 두 경우 모두 트레이딩 성과가 매우 좋지 않은 것을 확인할 수 있다.

Table 1 KOSPI 200 트레이딩 결과

Method	N_trade	Win %	AVG gain	AVG loss	Cum profit
Up Down	141	0.46	24.60	33.92	-978
Low High	28	0.57	40.85	90.55	-432
Reverse Low High	28	0.46	105.26	39.33	778

Table 2 S&P 500 트레이딩 결과

Method	N_trade	Win %	AVG gain	AVG loss	Cum profit
Up Down	108	0.47	32.88	56.99	-1571
Low High	37	0.43	22.73	91.68	-1561
Reverse Low High	38	0.71	102.47	14.14	2611

3. 결 론

본 논문에서는 트레이딩 시스템 개발 시 데이터 라벨링의 중요성에 대해 논의하였다. 기존의 Low High 라벨링 방법을 변환한 Reverse Low-High 라벨링 방법의 우수성을 확인하기 위해 KOSPI 200 과 S&P 500 주가 지수 데이터를 바탕으로 실증 분석을 진행하였다. 분석 결과 Up-Down, Low-High 방법에 비해 Reverse Low-High 방법의 거래 성과가 더 우수한 것으로 나타났다. 그러나 본 연구는 KOSPI 200, S&P 500 주가 지수 만을 대상으로 실증 분석이 진행되었다는 한계가 있다. 따라서 채권, 원자재 등 다양한 상품에 제안 방법을 적용하여 유용성을 확인할 필요가 있다.

ACKNOWLEDGMENT

본 연구는 2021 년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음(2021-0-01399)

참 고 문 헌

- [1] Dash R. and Dash PK. "A hybrid stock trading framework integrating technical analysis with machine learning techniques" *The Journal of Finance and Data Science*, 2016.
- [2] Zhong X. and Enke D. "Predicting the daily return direction of the stock market using hybrid machine learning algorithms" *Financial Innovation*, 2019.
- [3] Nabipour M., Nayyeri P., Jabani H., Shahab S. and Mosavi A. "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data: A comparative analysis" *IEEE Access*, 2020.
- [4] Sezer OB. and Ozbayoglu AM. "Algorithmic financial trading with deep convolutional neural networks: time series to image conversion approach" *Applied Soft Computing*, 2018.