

변동성 라벨링을 이용한 앙상블 머신러닝 트레이딩 시스템 개발

박명석, 유성주, 김재윤*
순천향대학교

pmsk980122@sch.ac.kr, sjyoo1572@naver.com, kimym38@sch.ac.kr*

Developing an ensemble trading system using the volatility labeling

Myeong Seok Park, Sungju Yoo, Jaeyun Kim*
Soonchunhyang Univ.

요 약

본 논문은 변동성 라벨링을 통한 앙상블 머신러닝 트레이딩 시스템을 제안한다. 변동성 라벨링은 수익률과 표준편차를 이용하여 학습데이터를 라벨링하는 방식이다. 또한 예측 성능 향상과 안정적인 트레이딩 시스템을 구축하기 위해 앙상블 기법을 적용한 결과 전통적으로 활용되고 있는 방법보다 우수한 결과를 도출하였다.

I. 서 론

주식 시장 예측이 어려운 이유는 주식 데이터가 비선형성과 비정상성 특징을 가지고 있기 때문이다. 주가 데이터는 일정한 분산과 일정한 평균을 가지지 않으며, 자기상관관계에서 시차의 길이에 따라 예측이 가능하지 않아 통계적 모델만으로는 예측이 어렵고 복잡하다.[1] 이를 극복하기 위해 머신러닝을 이용한 주가 예측 연구가 활발해지고 있다. Rohit Choudhry [2]는 유전자 알고리즘과 SVM(Support Vector Machine)을 활용하여 주식 시장을 예측하기 위하여 다양한 기술적 지표를 입력 변수로 활용하였다. 다른 주가의 상관관계를 이용하여 기술적 지표의 유용성을 증명하고 주가 예측에 대한 성능을 확인했다. Dongdong Lv[3]는 미국 시장의 SPICS 424 개, 중국 시장의 CSICS 185 개를 실험 대상으로 적용하고 44 개의 기술적 지표를 머신러닝의 입력 변수로 활용했다. 머신러닝과 딥러닝 알고리즘의 성능을 비교했을 때, 일부 머신러닝 알고리즘이 주가 방향성 예측 성능이 더 좋은 것을 확인했다. Samuel Asante Gyamerah[4]은 두 개의 머신러닝 알고리즘(Adaptive Boosting 및 K-Nearest Neighbor)와 stacking 앙상블 분류기를 활용하여 주식의 움직임을 예측했다. 개별 머신러닝 알고리즘과 앙상블 분류기의 정확도를 비교했을 때 앙상블 분류기의 정확도가 모든 분류기 중에서 가장 높은 정확도를 보임을 확인했다. Shikha Metha[5] 또한 앙상블 머신러닝 접근 방식을 제안했다. 앙상블 학습 방식이 감소된 분산으로 최대 정확도를 달성할 수 있음을 확인했다.

그러나 위와 같은 연구들은 다음과 같은 한계점을 가진다. 첫째, 사용 데이터의 기간이 짧아 다양한 패턴 학습이 불가 했다는 점이다. 둘째, 주가의 상승과 하락의

기준을 단순히 가격의 차이를 바탕으로 정의했다는 점이다. 단순히 가격의 차이를 바탕으로 라벨을 정의하면 비선형성과 비정상성을 가지는 데이터를 사용하게 되는 문제가 발생한다. 셋째, 주가 예측을 활용한 트레이딩 시스템의 부재이다. 위와 같은 한계점을 바탕으로 본 논문에서는 주가 방향을 기존 방법과 다르게 정의하는 라벨링 방식(변동성 라벨링)과 가격의 차이를 바탕으로 상승과 하락을 정의하는 라벨링 방식(벤치마크 라벨링)을 비교하고 앙상블 머신러닝 트레이딩 시스템을 제안한다.

II. 본론

본 논문의 제안 모델은 [Figure 1]과 같다. STEP 1에서는 주식 데이터를 가져와 기술적 지표를 추출한 뒤 변동성 라벨링 방식을 적용한다. STEP 2에서는 다양한 머신러닝 알고리즘을 이용해 학습 및 예측을 진행한다. STEP 3에서는 도출된 예측 값을 이용해 트레이딩 시뮬레이션을 진행하고 트레이딩 시스템을 구축한다.

2.1 STEP 1

주식 데이터는 시가총액 100 위 종목을 가져왔으며 학습 데이터를 동일한 기간 동안 하기 위해 2008 년 주가 정보가 없는 종목은 제외하였다. 기술적 지표는 선행연구에서 소개된 19 개를 사용했다. 변동성 라벨링 공식은 아래와 같으며 α 는 표준편차의 계수를 의미하며 σ 는 표준편차를 의미하고 n_preiod 는 수익률의 기간을 나타낸다.

$$DOWN = -\alpha \sigma \leq n_preiod \text{ profit\%}$$

$$UP = +\alpha \sigma \geq n_preiod \text{ profit\%}$$

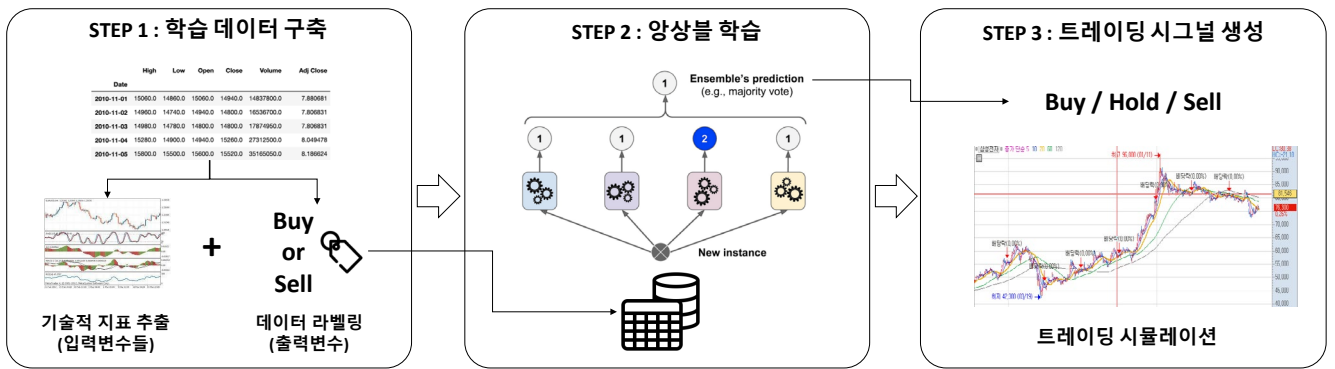


Figure 1 앙상블 트레이딩 시스템

2.2 STEP 2

앙상블 머신러닝 모델은 대표적인 방법인 보팅, 배깅, 부스팅 방법을 적용하였다. 보팅에 사용되는 알고리즘은 Decision tree, Logistic regression, Navie bayes, K-nn, MLP 사용하였으며 배깅은 Randomforest, 부스팅은 GBM 을 사용한다.

2.3 STEP 3

학습된 모델을 통해 예측 값을 생성하고 생성된 예측 값을 통해 트레이딩 시그널을 생성해 트레이딩 시뮬레이션을 진행한다. 매수 포지션이 발생하면 다음날 시가로 매수하며, 매도 포지션이 발생하면 다음날 시가로 매도한다. 매수 포지션이 발생하고 매도 포지션이 나오기 전까지는 보유하는 전략을 가지고 매도 포지션이 발생하고 매수 포지션이 나오기 전까지는 아무런 포지션을 갖지 않는다. 실거래와 유사한 결과를 도출하기 위해 거래 수수료는 0.015%를 부과하였다. 강한 매수 포지션과 매도 포지션을 잡기 위해 예측 임계 값을 down 은 0.25, up 은 0.75 로 조정한 뒤 진행했다. 임계 값을 변경한 이유는 강한 상승 신호와 하락 신호를 잡기 위함이다. 모델의 성능을 평가하기 위해 투자 성과 지표를 사용한다. 투자 성과 지표는 총 6 개로 거래 횟수(합계), 승률, 평균 수익, 평균 손실, Payoff ratio, Profit factor 가 있다.

III. 결론

트레이딩 시뮬레이션을 실행한 결과는 Table 1, 2 과 같다. 벤치마크 라벨링과 변동성 라벨링의 결과를 비교해보면 평균 수익, 평균 손실, Payoff ratio, Profit factor 에서 벤치마크 라벨링 결과보다 변동성 라벨링의 성과가 높은 것을 확인할 수 있다. 변동성 라벨링에서 Payoff ratio 와 Profit factor 의 성능이 벤치마크 라벨링의 비해 개선이 된 것을 볼 수 있다. 변동성 라벨링에서 제안한 앙상블 모형인 보팅 모형이 다른 앙상블 기법인 배깅과 부스팅의 성과보다 높은 것을 확인할 수 있다. 이는 제안한 보팅 모형이 안정적인 수익을 가져간다는 것을 의미하며 다른 머신러닝 알고리즘보다 성능이 좋은 것을 확인할 수 있다.

Table 1 벤치마크 라벨링을 이용한 트레이딩 결과

모델	거래 횟수	승률	평균 수익	평균 손실	Payoff Ratio	Profit factor
Decision tree	266	0.48	3104.65	2878.16	1.06	0.99
K-nn	275	0.48	2948.85	2737.96	1.05	0.98
Naive Bayes	96	0.63	3338.36	4875.03	0.89	1.42
MLP	165	0.53	2962.44	2996.49	0.98	1.13

Logistic regression	130	0.57	3316.58	3187.07	1.03	1.28
Voting	164	0.53	2925.27	2992.88	1.00	1.15
GBM	206	0.51	2952.76	2918.26	1.03	1.06
Random forest	216	0.51	2932.76	2913.82	1.04	1.10

Table 2 변동성 라벨링을 이용한 트레이딩 결과

모델	거래 횟수	승률	평균 수익	평균 손실	Payoff Ratio	Profit factor
Decision tree	286	0.47	5444.10	4572.17	1.17	1.04
K-nn	195	0.45	6647.84	5197.36	1.30	1.09
Naive Bayes	212	0.45	6006.77	4852.66	1.26	1.07
MLP	160	0.44	7528.58	5616.73	1.38	1.14
Logistic regression	136	0.44	7882.94	5862.92	1.38	1.13
Voting	185	0.44	8266.17	5740.48	1.42	1.29
GBM	283	0.47	5517.42	4612.71	1.17	1.03
Random forest	153	0.45	7392.22	5625.25	1.36	1.13

ACKNOWLEDGMENT

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

참 고 문 헌

- [1] 이형용 "한국 주가지수 등락 예측을 위한 유전자 알고리즘 기반 인공지능 예측기법 결합모형" Entru Journal of Information Technology 7.2 pp.33-43 (2008) : 33.
- [2] Choudhry, Rohit, and Kumkum Garg. "A hybrid machine learning system for stock market forecasting." World Academy of Science, Engineering and Technology 39.3 (2008): 315-318.
- [3] Lv, Dongdong, et al. "An empirical study of machine learning algorithms for stock daily trading strategy." Mathematical Problems in Engineering 2019 (2019).
- [4] Gyamerah, Samuel Asante, Philip Ngare, and Dennis Ikpe. "On stock market movement prediction via stacking ensemble learning method." 2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFER). IEEE, 2019.
- [5] Mehta, Shikha, et al. "Ensemble learning approach for enhanced stock prediction." 2019 Twelfth International Conference on Contemporary Computing (IC3). IEEE, 2019.