

디리클레 분포 기반 모델 기여도 예측을 이용한 앙상블 트레이딩 알고리즘

(Ensemble trading algorithm Using Dirichlet distribution-based model contribution prediction)

정재용*, 이주홍**, 최범기**, 송재원***

(Jae Yong Jeong, Ju Hong Lee, Bum Ghi Choi, Jae Won Song)

요약

알고리즘을 이용하여 금융 상품을 거래하는 알고리즘 트레이딩은 시장의 많은 요인들로 인해 그 결과가 안정적이지 못한 문제가 있다. 이 문제를 완화시키기 위해 트레이딩 알고리즘들을 조합한 앙상블 기법들이 제안되었다. 하지만 이 앙상블 방법에도 여러 문제가 존재한다. 첫째, 앙상블의 필요 조건인 앙상블에 포함된 알고리즘의 최소 성능 요건(랜덤 이상)을 만족시키도록, 트레이딩 알고리즘을 선택하지 못할 수 있다는 점이다. 둘째, 과거에 우수한 성능을 보인 앙상블 모델이 미래에도 우수한 성능을 보일 것이라는 보장이 없다는 점이다. 이 문제점들을 해결하기 위해 앙상블 모델에 포함되는 트레이딩 알고리즘들을 선택하는 방법을 다음과 같이 제안한다. 과거의 데이터를 기반으로 상위 성능의 앙상블 모델들에 포함된 트레이딩 알고리즘들의 기여도를 측정한다. 그러나 이 과거 데이터에만 기반된 기여도들은 과거의 데이터가 충분히 많지 않고 과거 데이터의 불확실성이 반영되어 있지 않기 때문에 디리클레 분포를 사용하여 기여도 분포를 근사시키고, 기여도 분포에서 기여도 값들을 샘플하여 불확실성을 반영한다. 과거 데이터로부터 구한 트레이딩 알고리즘의 기여도 분포를 기반으로 Transformer을 훈련하여 미래의 기여도를 예측한다. 예측된 미래 기여도가 높은 트레이딩 알고리즘들을 앙상블 모델에 선택하여 포함시킨다. 실험을 통하여 제안된 앙상블 방법이 기존 앙상블 방법들과 비교하여 우수한 성능을 보임을 입증하였다.

■ 중심어 : 앙상블 ; 트레이딩 알고리즘 ; 디리클레 분포 ; Transformer

Abstract

Algorithmic trading, which uses algorithms to trade financial products, has a problem in that the results are not stable due to many factors in the market. To alleviate this problem, ensemble techniques that combine trading algorithms have been proposed. However, there are several problems with this ensemble method. First, the trading algorithm may not be selected so as to satisfy the minimum performance requirement (more than random) of the algorithm included in the ensemble, which is a necessary requirement of the ensemble. Second, there is no guarantee that an ensemble model that performed well in the past will perform well in the future. In order to solve these problems, a method for selecting trading algorithms included in the ensemble model is proposed as follows. Based on past data, we measure the contribution of the trading algorithms included in the ensemble models with high performance. However, for contributions based only on this historical data, since there are not enough past data and the uncertainty of the past data is not reflected, the contribution distribution is approximated using the Dirichlet distribution, and the contribution values are sampled from the contribution distribution to reflect the uncertainty. Based on the contribution distribution of the trading algorithm obtained from the past data, the Transformer is trained to predict the future contribution. Trading algorithms with high predicted future contribution are selected and included in the ensemble model. Through experiments, it was proved that the proposed ensemble method showed superior performance compared to the existing ensemble methods.

■ keywords : ensemble ; trading algorithm ; Dirichlet distribution ; Transformer

* 학생회원, 인하대학교 전기컴퓨터공학과

** 정회원, 인하대학교 컴퓨터공학과 교수

*** 정회원, 밸류파인더스

이 논문은 2021년도 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2019R1F1A1062094, 2021R1F1A1050120, NRF-2020R1F1A1069361)

접수일자 : 2022년 01월 13일

게재확정일 : 2022년 03월 28일

수정일자 : 2022년 03월 15일

교신저자 : 이주홍 e-mail : juhong@inha.ac.kr

I. 서 론

알고리즘 트레이딩은 트레이딩 알고리즘을 통하여 금융상품을 거래하는 것을 말한다. 트레이딩 알고리즘으로 시장 분석 방법 중 하나인 기술적 지표를 이용하거나 머신러닝을 이용할 수 있다[1, 2]. 트레이딩 알고리즘을 거래에 적용하는 것은 인간의 감정적인 판단으로 인한 오판을 감소시킬 수 있는 장점이 있다. 알고리즘 트레이딩의 시장 점유율은 시간이 지남에 따라 계속해서 증가하고, 투자회사와 증권회사 등 많은 투자기관에서 사용하고 있다. 하지만, 트레이딩 알고리즘은 일반화 성능이 높지 않다. 즉, 모든 구간에서 우수한 성능을 보장하지 못한다. 따라서, 일반화 성능을 높이기 위해 트레이딩 알고리즘들을 앙상블로 조합하는 방법이 다양하게 시도되었다[2,5-8].

앙상블 방법이란 여러 알고리즘의 결정을 종합하여 결정을 내리는 것을 말한다. 앙상블을 사용하면 모델의 과적합(overfitting)을 방지하고 모델이 국지적 최적화(local optima)에 빠지는 것을 방지하여 모델의 일반화 성능을 개선할 수 있다[3,4]. 따라서, 트레이딩 알고리즘들을 조합하는 앙상블 방법들이 다양하게 시도되었다. [2]는 AND 연산자를 사용하고, [5]는 모델의 평균을 이용하며, [6]은 모델들의 중앙값을 이용하고, [7]은 적절한 한 개의 알고리즘을 선택하며, [8]은 일정 성능 이상의 모델들을 선택하여 투표(Vote)하는 방식으로 앙상블을 수행한다. 이러한 다양한 앙상블 방법들이 존재하지만, 기존 앙상블 방법들은 우수한 트레이딩 성능을 보장하지 못하고 있다.

앙상블 모델의 성능을 높이기 위해서는 다음의 두 가지 조건을 만족하여야 한다. 첫째 조건은 모든 알고리즘들은 최소한 무작위 성능 이상의 성능을 가지고 있어야 한다[3,4]. 둘째 조건은 과거 시점에서 좋은 성능의 모델이 미래 시점에도

좋은 성능일 확률이 높아야 한다는 것이다. 둘째 조건은 금융시장의 많은 요인과 금융 데이터의 비정상성(non-stationarity) 때문에 매우 다루기 어려운 문제이다[9]. 기존 앙상블 방법 중에서 주어진 두 가지 조건을 모두 만족하는 방법은 존재하지 않는다.

본 논문에서는 트레이딩 알고리즘의 앙상블 방법의 성능을 개선시키기 위해 다음과 같은 방법을 제안한다. 과거의 데이터를 기반으로 상위 성능의 앙상블 모델들에 포함된 트레이딩 알고리즘의 성능 기여도를 측정한다. 그러나 이 성능 기여도들을 미래 예측에 사용하기에는 과거의 데이터가 충분히 많지 않고 데이터의 불확실성을 반영할 수 없기 때문에 디리클레 분포를 사용하여 기여도를 근사시킨다. 과거 데이터 기반으로 구한 트레이딩 알고리즘의 기여도로 Transformer를 훈련하여 미래의 기여도를 예측한다. 현재 시점에서 예측된 기여도가 높은 트레이딩 알고리즘들을 앙상블 모델에 선택하여 포함시킨다.

II. 관련 연구

And 방법은 And 연산자를 이용하여 모델의 불확실성을 축소하였다. 두 개의 신경망으로 시계열 데이터의 추세를 예측하는 모델과 다음 시점의 시계열 데이터 값을 예측하는 모델을 만들고, 두 모델의 결정이 동일한 경우만을 매수/매도 결정을 수행하였다[2].

Averaging 방법은 Decision Tree, Logistic Regression, Gaussian Naive Bayes, Stochastic Gradient Descent 모델들을 모든 모델에 동등한 가중치를 부여하여 매수/매도 결정을 수행한다[5].

Median 방법은 Bagging 기법을 적용하였다. 훈련 데이터에 대한 복원 추출을 수행하여 각 LSTM 모델들을 훈련시키고, 여러 모델의 중앙값을 사용하여 매수/매도 결정을 수행한다[6].

Selection 방법은 PPO(Proximal Policy Optimization), A2C(Advantage Actor Critic),

DDPG(Deep Deterministic Policy Gradient) 기반으로 학습한 각 에이전트 중에서 시점마다 성능이 우수한 에이전트를 선택하여 매수/매도 결정을 수행한다[7].

Selection+Voting 방법은 뉴스를 입력으로 BERT 모델이 매수/매도를 결정하는 형태의 트레이딩 알고리즘을 구현하였다. 여러개의 BERT 모델 중에서 우수한 모델들을 선택하고, 선택된 모델들의 투표를 통하여 매수/매도를 결정한다[9].

Transformer는 Attention 방법을 적용한 신경망으로 기존 시퀀스 데이터를 다루는 신경망이 가지는 속도의 문제와 입력 데이터 간 거리에 영향을 받는 문제를 극복하였다[10].

Transformer는 NLP 분야에서 뛰어난 성능을 가지고 있을 뿐 아니라, 이를 시장 예측에서도 이용한 사례들이 있다[8,11-13]. [9,11,12]은 Transformer 모델의 인코더를 이용하여 뉴스 데이터와 소셜 미디어 데이터로부터 특징들을 추출하여 시장의 방향성을 예측한다. [13]은 금융 데이터를 Transformer로 훈련하여 시장을 예측하였다.

III. 제안 방법

1. 연구 동기

앙상블 방법들을 비교하기 위한 요소(기능)들은 알고리즘 선택, 시간 정보 반영이다. 알고리즘 선택 기능은 여러 트레이딩 알고리즘 중에서 앙상블 방법을 수행하였을 때, 앙상블 방법의 성능을 향상 시킬 것이라고 기대되는 알고리즘들을 선택하는 과정이다. 시간 정보 반영 기능은 과거 시점에서 알고리즘의 성능을 기반으로 미래 시점의 성능이 좋은 알고리즘을 선택하는 기능이다.

표1에서 보는 바와 같이 기존의 앙상블 방법들은 두 가지 비교 요소들을 모두 충족시키지 못하고 있다. Averaging 방법과 And 방법은 별도의 알고리즘을 선택하는 과정이 없이 주어진 모델의 결정을 전부 반영하여 수행된다. 이때, 성능이 낮은 트레이딩 알고리즘이 존재한다면, 앙상블 모델

의 성능 저하에 영향을 미치게 된다. Selection 방법과 Median 방법은 적절한 모델을 선택하는 과정이 존재하지만, 하나의 모델만 사용하기 때문에 다른 우수한 트레이딩 알고리즘의 정보를 반영할 수 없다. Selection+Voting은 우수한 성능의 여러 트레이딩 알고리즘들을 선택하였지만, 다른 앙상블 방법과 마찬가지로 과거 데이터로 얻은 조합만 사용하기 때문에 미래 시점에서 의미가 없을 수 있다.

표 1. 앙상블 방법별 비교

	알고리즘 선택	시간 정보 반영
Averaging	X	X
Median	△	X
And	X	X
Selection	△	X
Selection+Voting	O	X
제안 방법	O	O

본 논문은 트레이딩 알고리즘의 기여도를 이용하여 앙상블의 성능을 향상 시킬 수 있는 트레이딩 알고리즘들을 선택하는 방법을 제시한다. 그림 1과 같이 이미 시간이 지나서 불확실성이 제거된 단순한 과거의 데이터로부터, 과거 데이터의 불확실성을 같이 고려하기 위하여 각 시간에 따른 분포를 가정하였다. 분포의 추정을 위해서 앙상블 방법을 사용하여 상위 성능의 앙상블들에 속하는 알고리즘들의 포함 회수를 그 알고리즘의 기여도로 산정하였다. 또한, 기여도의 분포를 디리클레 분포로 가정하여 샘플된 값들이 알고리즘의 기여도에 포함된 불확실성을 나타내게 하여 Transformer를 훈련시킨다. 이를 통해, 과거의 기여도 불확실성의 변화 추이로부터 학습된 Transformer가 미래에 대한 기여도의 변화 추이를 반영하여 예측할 수 있게 된다.

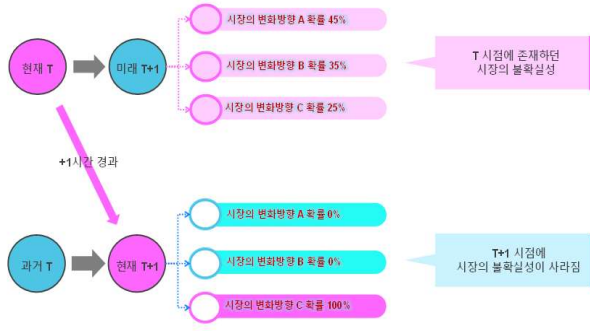


그림 1. 시간이 지남에 따라 과거 정보의 불확실성이 사라짐

2. 트레이딩 알고리즘의 평가

가. 앙상블 모델에 의한 간접 평가

기존 트레이딩 알고리즘은 개수가 제한적이거나, 유사하게 작동하는 알고리즘들이 존재하기 때문에, 모델의 다양성이 제한될 수 있다. 이를 해결하기 위하여 트레이딩 알고리즘을 개별로 평가하는 것이 아니라 앙상블 모델을 평가하여 그 안에 포함된 개별 트레이딩 알고리즘의 기여도를 부여함으로써 트레이딩 알고리즘의 성능을 간접적으로 평가한다(그림 2).

나. 성능별 상위 앙상블 선택

그림 2에서 보여준 바와 같이 성능이 우수한 V 개의 앙상블 모델을 찾는다. 이때 앙상블 모델을 평가하는 기준은 샤프지수(수식 1)를 사용한다. 이때, 샤프지수는 앙상블 모델을 이용하여 실제 투자 수행하였을 때, 변동성 대비 수익의 의미를 가진다.

$$\text{Sharpe ratio} = \frac{\text{mean}(r)}{\text{stddev}(r)} \quad (1)$$

이때, 각 트레이딩 알고리즘이 상위 성능의 앙상블 모델에 많이 포함되어 있으면 트레이딩 알고리즘의 앙상블 모델 성능에 대해 기여도가 높다고 평가한다.

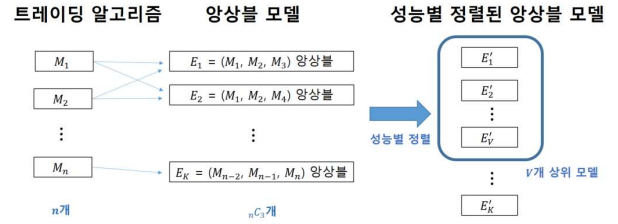


그림 2. 트레이딩 앙상블 모델과 평가

3. 트레이딩 알고리즘 기여도 분포

가. 초기분포 설정

트레이딩 알고리즘의 기여도는 형식적으로 알고리즘이 상위 성능의 앙상블 모델들에 포함될 확률로 정의한다(수식 2). 여러 알고리즘들에 대한 기여도들의 분포는 다변수 분포이며, 기여도 변수들의 합이 1로 제한되므로(수식 3), 이 조건을 만족하는 분포로써 디리클레 분포를 사용하였다. t 시점에서 i 번째 트레이딩 알고리즘(M_i)의 기여도(C_i^t)가 디리클레 분포를 따른다고 가정한다(수식 4). 이때 분포의 파라미터 α_i^t 는 과거 데이터에서 상위 V 개의 앙상블 모델에 M_i 가 포함된 횟수를 의미한다.

$$E[C_i^t] = \frac{\alpha_i^t}{\sum_{j=1}^n \alpha_j^t} \quad (2)$$

$$\sum_{i=1}^n C_i^t = 1 \quad (3)$$

$$C^t = (C_1^t, C_2^t, \dots, C_n^t) \sim \text{Dir}(\alpha_1^t, \alpha_2^t, \dots, \alpha_n^t) \quad (4)$$

나. 분포 업데이트

그림 3에서 본 바와 같이, 과거 시점의 분포의 파라미터에서 현재 시점의 상위 앙상블에 포함된 각 트레이딩 알고리즘의 수를 합하여 파라미터를 업데이트할 수 있다.

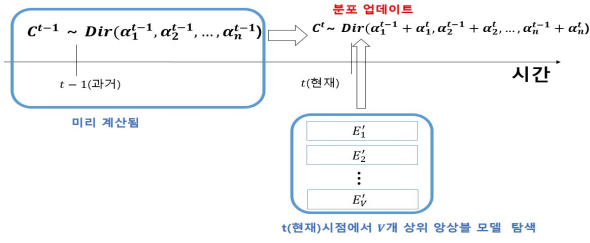


그림 3. 시간에 따른 분포 업데이트

과거 시점($t-1$)에서 모델의 기여도 분포가 정의되어 있다면(수식 5), 과거 시점부터 현재 시점(t)까지 구간에서 상위 성능의 양상불 모델을 이용하여 이전 시점에 구한 기여도 분포의 likelihood를 구할 수 있다(수식 6).

$$C^{t-1} \sim \text{Dir}(\alpha_1^{t-1}, \alpha_2^{t-1}, \dots, \alpha_n^{t-1}) \quad (5)$$

$$P(\alpha^t | C^{t-1}) = \frac{3V!}{\alpha_1^t! \alpha_2^t! \dots \alpha_n^t!} (C_1^{t-1})^{\alpha_1^t} \dots (C_n^{t-1})^{\alpha_n^t} \quad (6)$$

디리클레 분포는 다항분포의 Conjugate Prior이기 때문에 베이지 정리에 따라 사후분포를 구할 수 있다(수식 7)[14].

$$C^t \sim \text{Dir}(\alpha_1^{t-1} + \alpha_1^t, \alpha_2^{t-1} + \alpha_2^t, \dots, \alpha_n^{t-1} + \alpha_n^t) \quad (7)$$

다. 감소 계수 적용

금융 시계열 데이터에서는 최근 정보일수록 더 큰 가중치를 가져야 한다. 따라서, 감소 계수(γ)를 사용하여 과거 정보의 가중치를 감소시킨다(수식 8).

$$C^t \sim \text{Dir}(\lfloor \alpha_1^{t-1}/\gamma \rfloor + \alpha_1^t, \dots, \lfloor \alpha_n^{t-1}/\gamma \rfloor + \alpha_n^t) \quad (8)$$

4. 기여도 예측

금융 데이터는 비정상성을 가지기 때문에, 이전 시점에서 얻은 데이터 분포가 다음 시점에도 그대로 유지되는 것을 기대하기 어렵다. 따라서, 이전 기여도 정보를 그대로 사용하는 방식이 아닌 다음 시점의 기여도를 예측하는 방법을 사용하여 이러한 문제를 극복하고자 하였다. 이를 위해, 시계열 데이터에 대하여 우수한 성능을 가진 Transformer 모델을 사용하여 다음 시점에 대한 정보를 예측하였다.

가. 훈련 데이터 샘플링

과거의 각 시점에서 구한 기여도 분포로부터 p 개의 데이터를 샘플링한다(그림4). $c_{i,(j)}^t$ 는 t 시점의 디리클레 분포에서 M_i 의 j 번째 샘플 기여도 값이다.

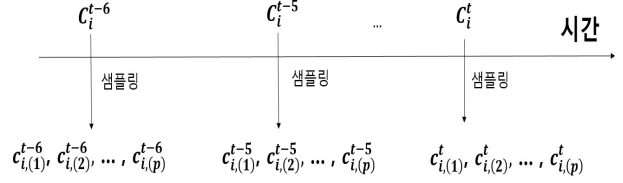


그림 4 훈련 데이터 생성

나. Transformer 훈련

그림 5는 기여도 예측에 대한 Transformer를 훈련 시키는 과정이다. 특정 시점 이전의 샘플된 기여도 값들을 입력 데이터로, 특정 시점 이후 샘플된 기여도 값을 출력 데이터로 구성된 데이터 셋을 사용하여 Transformer를 훈련시킨다.

$$X : [[c_{i,(1)}^{t-16}, c_{i,(1)}^{t-15}, \dots, c_{i,(1)}^{t-11}], \dots, [c_{i,(p)}^{t-16}, c_{i,(p)}^{t-15}, \dots, c_{i,(p)}^{t-11}]]$$

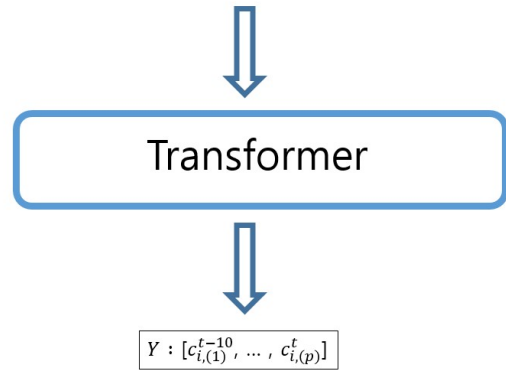


그림 5. Transformer 훈련

다. 기여도 예측

그림6는 (나.)에서 훈련된 Transformer를 사용하여, 미래 기여도를 예측하는 과정이다. 현재 시점을 포함한 가장 가까운 과거 시점들의 샘플들을 입력으로 사용 미래 기여도를 예측한다. 훈련된 Transformer에 p 개의 샘플 데이터를 입력으로 하여 p 개의 미래 기여도를 예측한다. 이때,

$\widehat{c}_{i,(j)}^{t+1}$ 는 미래 시점($t+1$)에 M_i 에 대한 j 번째 기여도 예측 값이다.

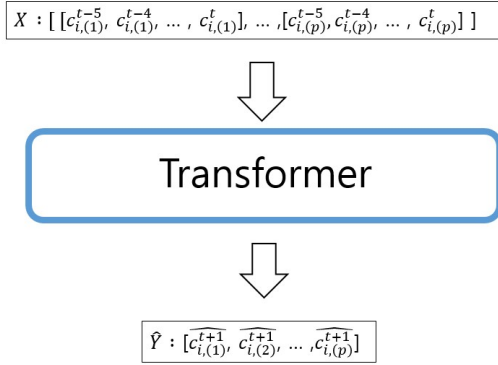


그림 6. 가중치 예측

5. 모델 선택

각 트레이딩 알고리즘별 기여도 예측평균을 구한다(수식 9). 그림7과 같이, 기여도 예측평균이 높은 트레이딩 알고리즘을 앙상블에 포함되도록 선택한다(수식 10).

$$\widehat{c}_i^{t+1} = \frac{1}{p} \sum_{j=1}^p \widehat{c}_{i,(j)}^{t+1} \quad (9)$$

$$\widehat{c}_i^{t+1} = \begin{cases} 0 & \text{if } \widehat{c}_i^{t+1} \leq \frac{1}{n} \\ \widehat{c}_i^{t+1} & \text{if } \widehat{c}_i^{t+1} > \frac{1}{n} \end{cases} \quad (10)$$

6. 앙상블 수행

각 트레이딩 알고리즘이 내린 매수/매도 결정에서 기여도를 가중치로 곱하여 매수/매도의 결정을 선택한다.

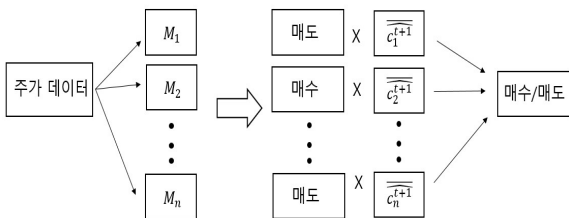


그림 7. 앙상블 수행

제안 방법에 대한 전체적인 구성은 그림8로 표현된다.

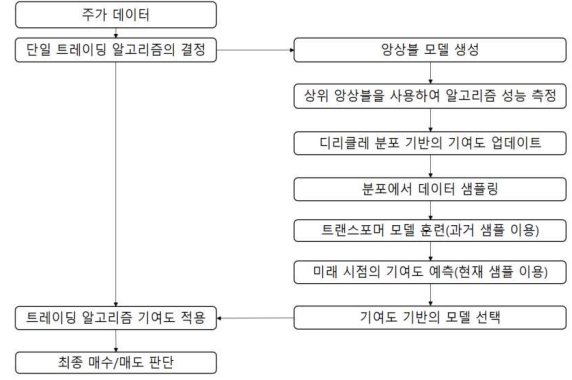


그림 8. 제안 방법에 대한 전체적인 구성

IV. 실험

1. 실험 설정

제안한 방법의 성능을 검증하기 위하여 시장의 횡보, 하락과 상승의 정보를 모두 반영하고 있는 구간인 2018년 9월 27일부터 2021년 7월 29일까지의 10종목의 주식 데이터로 실험을 수행하였다. 주식 종목은 현대차, 삼성전자, 삼성전기, 한국조선해양, 롯데케미칼, 현대모비스, 한국전력, KT&G, LG디스플레이, 신한지주를 사용하였다. 앙상블을 적용하기 위한 트레이딩 알고리즘은 전체 30개를 사용한다.

기술적 지표를 기반으로 한 트레이딩 알고리즘 다음과 같이 15개를 사용하였다. 이동 평균선 정배열(10,30,60), 급등주 포착 알고리즘(650), MACD(12,26), MACD oscillator(1.01,0.99), 볼린저 밴드(2,12), Fast stochastic K(0.7,0.3), Fast stochastic D(10,5), Slow stochastic K(5), Slow stochastic D(5), DMI(12), CCI(20), 이동 평균선 밀집 (5 , 20 , 60) , 20 일 선돌파, 60일선 돌파, Swing pattern(1).

머신러닝 알고리즘은 다음과 같이 9개를 사용하였다. Logistic regression, Random forest, Gradient Boosting 알고리즘이 각각 장기, 중기, 단기 예측을 위해 3개씩 있다.

딥러닝 알고리즘은 다음과 같이 6개를

사용하였다. ANN, LSTM이 장기, 중기, 단기 예측을 위해 3개씩 있다. 각 알고리즘은 2012년 1월 2일부터 2018년 6월 29일까지의 기간의 데이터를 훈련에 사용하였다.

기여도 예측에 사용되는 Transformer 모델은 배치 크기는 128, 멀티 헤드의 수는 100개, 손실 함수는 cross entropy, 최적화 기법으로는 Adam Optimizer, Drop out rate는 0.1, Learning rate는 0.001 사용하여 모델을 훈련하였다.

2. 실험 결과

그림 9은 10개의 주식 종목을 동일 가중치 포트폴리오로 구성했을 때의 투자 결과이다. 실험 결과에서 회색 선은 실제 주가의 가격 변동을 나타낸다. 초록색 선은 제안한 앙상블 방법을 사용하였을 때 투자 결과이고, 파란색 선은 Averaging 앙상블의 결과이고, 하늘색 선은 Median 앙상블의 투자 결과이고, 자홍색 선은 Selection 앙상블 투자 결과이고, 남색 선은 AND 앙상블 투자 결과이고, 남색 선은 Selection+Voting 앙상블 투자 결과이다. 그림 9에서 보여준 바와 같이, 제안한 앙상블 방법이 다른 앙상블 방법보다 성능이 좋은 것을 알 수 있다. 또한, 실제 주가는 변동성이 크게 나타나지만, 제안한 앙상블 방법을 사용한 투자 결과는 수익률이 좋으면서도 변동성이 적음을 알 수 있다. 이를 통해 과거 데이터 정보만을 이용하는 앙상블 방법은 시계열의 비정상성으로 인하여 성능이 떨어지지만, 미래 기여도 예측 기반의 앙상블 방법은 이러한 문제를 완화할 수 있음을 볼 수 있다.

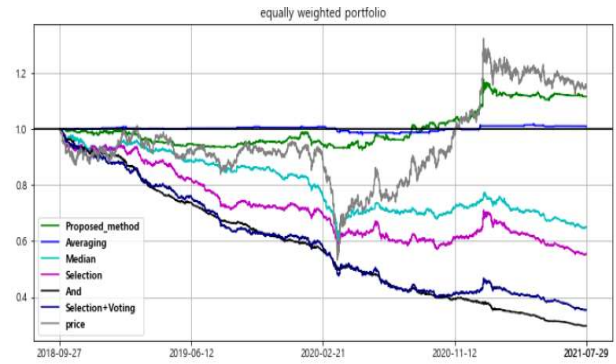


그림 9. 동일 가중치 포트폴리오

그림10는 제안 방법의 성능이 가장 우수한 경우이다. 주가의 변동성이 강한 구간에서 적은 변동성을 유지하는 것을 볼 수 있고, 주가가 상승하는 구간에서는 수익률이 좋은 것을 볼 수 있다.



그림 10. 제안 방법의 성능이 가장 우수한 경우

그림11은 제안 방법의 성능이 가장 낮은 경우이다. 주가의 변동성이 강한 구간에서 적은 변동성을 유지하였지만, 수익률이 떨어졌다. 이러한 문제를 극복하기 위하여서 투자를 수행할 때 포트폴리오를 구성한 투자를 수행하고, 그림 9는 포트폴리오에 대해서 제안한 방법의 성능이 우수함을 보였다.

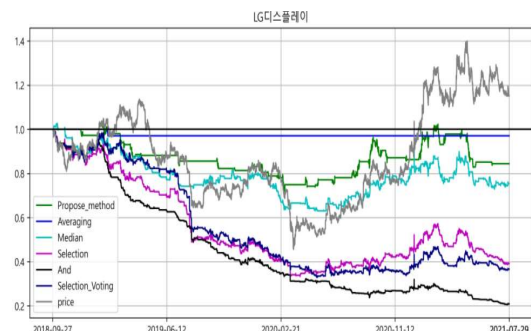


그림 11. 제안 방법의 성능이 가장 낮은 경우

표 2는 주가와 각 앙상블 방법을 사용한 투자 결과에 대한 투자 성능 지표들을 나타낸다. 이때, 성능평가 지표는 샤프지수를 사용하였다(수식 1). 10개의 종목에 대해서 제안한 방법의 샤프지수가 다른 앙상블 방법들과 비교하여 평균적으로 우수한 성능을 나타내었다. 이를 통해서 금융 시장의 비정상성으로 인해서 과거의 우수한 조합들이 현재 시점에도 우수하게 적용되지 않음을 볼 수 있고, 미래 기여도를 예측하는 방법이 유효함을 볼 수 있다. 각 종목의 투자 결과가 상이 하기 때문에, 이에 대한 일반적인 성능평가를 위해 포트폴리오를 구성한 투자 결과를 평가하였고 제안 방법의 성능이 우수함을 확인하였다.

표 2. 앙상블 방법별 샤프지수 비교

	제안 방법	주가	Averaging	median	selection	AND	Selection + Voting
현대차	0.9	0.6	0.8	-0.3	-0.1	-1.5	-0.8
삼성전자	0.2	0.7	0.6	-0.02	-0.5	-3.5	-1.1
삼성전기	0.4	0.4	0.4	-0.6	-0.5	-2.4	-1.0
한국조선해양	0.3	0.2	-0.3	-0.9	-0.6	-2.7	-0.9
롯데케미칼	0.08	0.15	0.08	-0.28	-0.8	-1.6	-1.6
현대모비스	0.4	0.3	-0.5	-0.8	-0.4	-2.7	-1.9
한국전력	-0.1	0.001	0.6	-1.1	-1.1	-3.3	-2.4
KT&G	0.1	-0.2	-1.4	-0.8	-1.1	-2.0	-1.8
LG디스플레이	-0.2	0.2	-0.5	-0.3	-1.0	-2.6	-1.0
신한저축	-0.2	0.02	-0.4	-0.9	-0.6	-2.6	-0.9
포트폴리오	0.53	0.29	0.23	-1.3	-1.5	-6.4	-2.8

V. 결 론

본 논문은 트레이딩 알고리즘이 가지는 비정상성과 앙상블 방법의 조건을 기반으로, 기존 앙상블 방법이 가지는 문제를 찾고 이를 극복하기 위한 2가지 방법을 제안하였다. 첫째, 트레이딩 알고리즘의 성능에 따른 기여도 분포를 디리클레 분포로 근사하여 학습에 사용될 데이터의 부족에 따른 불확실성을 반영하였다. 둘째, Transformer를 사용하여 미래 시점에 기여도가 높은 알고리즘으로 앙상블의 요소 알고리즘을 선택하였다. 실험을 통하여 제안 방법과 기존 앙상블 방법들을 비교하였고, 기존 앙상블 방법과 같이 과거 정보만을 사용한 알고리즘을 종합하

는 방법보다 미래의 기여도를 예측하여 알고리즘을 종합하는 방법의 성능이 평균적으로 우수함을 확인하였다. 또한, 여러 종목에 대한 일반적인 성능평가를 위해 포트폴리오를 구성한 성능평가 결과에서도 제안 방법의 성능이 우수함을 보였다. 이를 통해, 제안한 방법의 성능이 우수함을 확인하였다.

REFERENCES

- [1] Drakopoulou Veliot, "A review of fundamental and technical stock analysis techniques.", *Journal of Stock & Forex Trading*, vol. 5, no. 1, Nov. 2016.
- [2] Giacomel, Felipe, Renata Galante, and Adriano Pereira, "An algorithmic trading agent based on a neural network ensemble: a case of study in North American and Brazilian stock markets." *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, IEEE, vol. 2, Singapore, Dec. 2015.
- [3] Sagi, Omer, and Lior Rokach, "Ensemble learning: A survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, Feb. 2018
- [4] Huang, Faliang, Guoqing Xie, and Ruliang Xiao, "Research on ensemble learning." *2009 International Conference on Artificial Intelligence and Computational Intelligence*, IEEE, vol. 3, Shanghai, China, Nov. 2009.
- [5] Ullah, A. K. M., et al., "Combining machine learning classifiers for stock trading with effective feature extraction.", *preprint arXiv:2107.13148, arXiv*, 2021.
- [6] SUN Shaolong, WANG Shouyang and WEI Yunjie, "A new ensemble deep learning approach for exchange rates forecasting and trading." *Advanced Engineering Informatics*, vol. 46, Oct. 2020.
- [7] Hongyang Yang, Xiao-Yang Liu, Shan Zhong and Anwar Walid, "Deep reinforcement learning for automated stock trading: An ensemble strategy," *SSRN*, Oct. 2020.
- [8] Man, Xiliu, Jianwu Lin, and Yujiu Yang. "Stock-UniBERT: A News-based Cost-sensitive Ensemble BERT Model for Stock Trading." *2020 IEEE 18th International Conference on Industrial Informatics (INDIN)*, IEEE, vol. 1, Warwick, United Kingdom, Jul. 2020.

- [9] HUANG, Norden E., et al., "Applications of Hilbert-Huang transform to non-stationary financial time series analysis." *Applied stochastic models in business and industry*, Vol. 19, pp. 245-268, Oct. 2003.
- [10] VASWANI, Ashish, et al., "Attention is all you need." *Advances in neural information processing systems*, pp. 5998-6008, 2017.
- [11] Daiya Divyanshu, and Che Lin, "Stock Movement Prediction and Portfolio Management via Multimodal Learning with Transformer." *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 3305-3309, Toronto, ON, Canada, 2021.
- [12] JLIU, Jintao, et al., "Transformer-based capsule network for stock movement prediction." *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, pp. 66-73, 2019.
- [13] DING, Qianggang, et al., "Hierarchical Multi-Scale Gaussian Transformer for Stock Movement Prediction." *IJCAI*, no. 640, pp. 4640-4646, Jan. 2020.
- [14] TU Stephen, "The dirichlet-multinomial and dirichlet-categorical models for bayesian inference." *Computer Science Division*, UC Berkeley 2, 2014.

저 자 소 개



정재용(학생회원)

2020년 인천대학교 컴퓨터학과 학사 졸업.
2020년 ~ 현재 인하대 전기컴퓨터공학과 석사 과정.

<주관심분야 : 시계열 데이터, 머신러닝>



이주홍(정회원)

1983년 서울대학교 전자계산기공학과 (현 컴퓨터공학과) 학사 졸업.
1985년 서울대학교 전자계산기공학과 (현 컴퓨터공학과) 석사 졸업.
2001년 KAIST 컴퓨터공학전공 박사 졸업.
2001년 ~ 현재 인하대 컴퓨터공학과 교수.

2019년 ~ 현재 ㈜큐헛지 대표이사

<주관심분야 : 머신러닝, 금융투자공학>



최범기(정회원)

1986년 서울대학교 수학과 학사 졸업.
1995년 Florida State university computer science 석사 졸업.
2017년 ~ 2020년 인하대학교 컴퓨터공학과 겸임교수.
2019년 ~ 현재 ㈜큐헛지 이사

<주관심분야 : 머신러닝, 금융투자공학>



송재원(정회원)

2005년 성공회대학교 컴퓨터공학과 학사 졸업.
2007년 인하대학교 컴퓨터공학과 석사 졸업.
2013년 인하대학교 컴퓨터공학과 박사 졸업.
2013~2014 대구경북과학기술원 개발팀장.

2014~2017 인천국제공항공사 선임 연구원

2017~2017 가천길대학병원 연구교수

2017~현재 ㈜밸류파인더스 대표이사

<주관심분야 : 머신러닝, 데이터마이닝, 금융IT>