

부분복제 지수 상향 추종을 위한 진화 알고리즘 기반 3단계 포트폴리오 선택 양상을 학습 (3-stage Portfolio Selection Ensemble Learning based on Evolutionary Algorithm for Sparse Enhanced Index Tracking)

윤동진*, 이주홍**, 최범기**, 송재원***

(Dong Jin Yoon, Ju Hong Lee, Bum Ghi Choi, Jae Won Song)

요약

지수 상향 추종은 시장수익률을 쫓는 지수 추종에 기반을 두고 지수 이상의 수익을 내기 위한 목적함수를 최적화시키는 문제이다. 큰 거래비용과 비유동성 등의 문제를 피하기 위하여 지수 편입종목들 중 일부만을 선택하여 포트폴리오를 구성하는 방법을 사용하였다. 기존의 지수 상향 추종 방법들은 테스트하는 모든 구간에서 하나의 목적함수만을 가지고 최적의 포트폴리오를 찾고자 하였으나 변동성이 큰 금융시장에서 항상 잘 통하는 궁극의 전략을 찾고자 하는 것은 불가능에 가깝다. 또한, 시간에 따라 통계적 특성이 크게 변하는 금융시장의 특성 상 학습 데이터에 대한 목적함수를 최적화시키는 것 이상으로 일반화 성능을 향상시키는 것이 중요하지만 기존의 방법들은 이를 위한 직접적인 논의가 부족하다는 한계점을 가지고 있다. 이러한 문제점을 해결하기 위하여 본 논문은 여러 목적함수들을 조합하여 포트폴리오를 구성하는 양상을 학습과 학습 데이터에 대한 목적함수 이외의 다른 기준을 적용하여 포트폴리오를 선택할 수 있는 3단계 포트폴리오 선택 알고리즘을 제안한다. S&P500 지수를 사용한 실험에서 제안한 방법은 지수와 기존 방법들보다 27% 이상 높은 Sharpe ratio를 보이며 3단계 포트폴리오 선택 알고리즘과 양상을 학습이 지수 상향 포트폴리오를 선택하는데 효과가 있음을 보여준다.

■ 중심어 : 지수 상향 추종 ; 포트폴리오 선택 ; 진화 알고리즘

Abstract

Enhanced index tracking is a problem of optimizing the objective function to generate returns above the index based on the index tracking that follows the market return. In order to avoid problems such as large transaction costs and illiquidity, we used a method of constructing a portfolio by selecting only some of the stocks included in the index. Commonly used enhanced index tracking methods tried to find the optimal portfolio with only one objective function in all tested periods, but it is almost impossible to find the ultimate strategy that always works well in the volatile financial market. In addition, it is important to improve generalization performance beyond optimizing the objective function for training data due to the nature of the financial market, where statistical characteristics change significantly over time, but existing methods have a limitation in that there is no direct discussion for this. In order to solve these problems, this paper proposes ensemble learning that composes a portfolio by combining several objective functions and a 3-stage portfolio selection algorithm that can select a portfolio by applying criteria other than the objective function to the training data. The proposed method in an experiment using the S&P500 index shows Sharpe ratio that is 27% higher than the index and the existing methods, showing that the 3-stage portfolio selection algorithm and ensemble learning are effective in selecting an enhanced index portfolio.

■ keywords : Enhanced index tracking ; Portfolio selection ; Evolutionary algorithm

금융시장에서 사용될 수 있는 투자 전략은 크게 두 가지로 나누어진다. 적극적 투자 전략(active investment strategy)은 시장보다 큰 위험을 수반하지만 숙련된 지식과 정보를 통해 시

* 학생회원, 인하대학교 전기컴퓨터공학과 대학원생

** 정회원, 인하대학교 전기컴퓨터공학과 교수

*** 정회원, 벨류파인더스

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

접수일자 : 2021년 08월 .03일

수정일자 : 1차 2021년 09월 01일, 2차 2021년 09월 16일

제재확정일 : 2021년 09월 24일

교신저자 : 이주홍 e-mail : juhong@inha.ac.kr

장수익률을 넘는 이익을 얻고자 하는 전략이다. 반면, 효율적 시장 가설에 따라 시장수익률 이상의 이익을 얻기 힘들다는 것을 인정하고 안정적으로 시장수익률 만큼의 수익을 추구하는 전략을 소극적 투자 전략(passive investment strategy)이라 하며, 과거 여러 기록들에서 적극적 투자 전략보다 더 좋은 성과를 보여주었다[1-3].

대표적인 소극적 투자 방법인 지수 추종(index tracking)은 주식시장의 성과를 대표하는 값인 지수의 수익률을 추종하는 포트폴리오를 구성하여 투자하는 방법이다. 지수 추종 포트폴리오를 구성하는 방법은 완전 복제 방식(full replication)과 부분 복제 방식(sparse replication)으로 나누어진다. 완전 복제 방식은 지수에 편입된 모든 종목을 포함하여 지수의 수익률을 추종하는 방법으로, 수백 개의 크고 작은 종목들을 매매하는 과정에서 발생하는 거래비용과 슬리피지(호가 차이에 의한 결제가격 차이), 비유동성 등의 문제들로 인하여 실질적으로 사용하기에 어려움이 있다. 반면, 지수의 전체 편입 종목 대신 일부만을 사용하는 부분 복제 방식은 이러한 문제들로부터 다소 자유롭다. 부분 복제 방식은 실제 편입비율을 사용할 수 없기 때문에 각 종목의 편입비율을 정하기 위하여 목표 지수와 이를 추종하는 포트폴리오 간의 차이를 수치화한 추종 오차(tracking error)를 도입한다. 이를 최소화시키는 최적화 문제를 해결하는 것이 지수 추종 연구의 주요 과제이다[4-6].

지수 추종 문제가 연구됨에 따라 지수 추종이라는 패러다임에 기반을 두고 지수를 상향하는 포트폴리오를 찾고자 하는 시도들이 연구되었다[5,7-12]. 이를 조합한 문제를 지수 상향 추종(enhanced index tracking, enhanced indexation) 문제라고 하고, 지수를 상회하는 수익률을 내는 포트폴리오를 지수 상향 추종 포트폴리오라 한다.

기존의 연구들은 지수 대비 포트폴리오의 초과수익률에 대한 제곱평균제곱근과 평균 초과수익률의 선형 조합[5], 하방 위험[8,9] 등 지수 상향 추종 목적에 가장 잘 부합하는 특정 목적함수를 사용하여 문제를 해결하고자 하였다. 그러나 다양한 위험과 제약조건들이 존재하는 실제 환경에서 압도적으로 좋은 성능을 보이면서 실질적으로 사용될 수 있는 추종 오차 혹은 목적함수는 존재하지 않는다는[11]. 또한, 학습 구간에서의 성능 향상에 초점을 맞춘 연구들은 많이 이루어지고 있으나, 테스트 구간에서 학습 구간에서의 성능을 어떻게 지속시킬 것인가에 대한 논의는 부족하다.

본 논문의 목적은 상술한 기존 연구들의 한계점을 보완하여 지수 상향 포트폴리오를 선택하는 개선된 방법을 제안하는 것이다. 본 논문은 끊임없이 변화하는 금융시장에서 여러 목적함수들을 적응적으로 조합하여 포트폴리오를 구성할 수 있는 양상을 학습을 제안하고, 학습 데이터에 모델이 오버피팅되는 것을 막기 위해 학습 데이터에 대한 목적함수 이외의 조건을 사용하여 포트폴리오를 선택할 수 있는 3단계 포트폴리오 선택 방법을 제안한다. 제안된 방법의 성능을 평가하는 실험을 위해 약

5년간 S&P500 지수 데이터가 사용되었으며 지수와 기존 방법들보다 제안된 방법이 다양한 평가 기준에서 더 좋은 성과를 보여주어 제안 방법이 실질적으로 사용 가능하고 유의미하다는 것을 보여주었다.

논문의 나머지 구성은 다음과 같다. ‘II. 배경지식’에서 관련 연구와 제안 방법에 대한 배경지식에 대하여 설명한다. ‘III. 관련 연구’에서 지수 상향 추종을 위하여 사용되었던 기존의 방법들을 소개한다. ‘IV. 연구 동기 및 기여’에서 기존 연구들의 한계점과 본 논문의 기여를 소개한다. ‘V. 제안 방법’에서 IV장에서 언급한 문제점들을 해결하기 위해 제안한 방법에 대하여 설명한다. ‘VI. 실험’에서 제안된 방법의 성능을 평가하기 위한 실험을 소개하고 그 결과를 보여준다. 마지막으로, ‘VII. 결론’에서 본 논문의 결론과 추후 연구과제에 대하여 설명한다.

이후 사용되는 표기법은 다음과 같다.

- T 구간의 전체 길이 ($0 \leq t \leq T$)
- v_t 시점 t 에서의 포트폴리오 가치
- r_t 시점 t 에서의 포트폴리오 수익률 ($r_t = (v_t - v_{t-1})/v_{t-1}$)
- V_t 시점 t 에서의 지수의 값
- R_t 시점 t 에서의 지수 수익률 ($R_t = (V_t - V_{t-1})/V_{t-1}$)
- K 지수 상향 추종 포트폴리오의 편입종목 개수

II. 배경지식

1. Mean-variance framework

대부분의 지수 추종 문제와 지수 상향 추종 문제의 기반에는 Markowitz의 포트폴리오 선택 이론[13,14]이 자리 잡고 있다. [13,14]에서 제안된 Markowitz mean-variance framework는 포트폴리오의 이익(yield)과 위험(risk)을 각각 기대수익률과 분산으로 간주하고 이를 바탕으로 포트폴리오 선택 기준을 정립하였다. 이후 포트폴리오의 기대수익률과 분산은 다양한 형태로 지수 추종 및 지수 상향 추종 문제에 적용되었다.

2. 진화 알고리즘

유전 알고리즘(genetic algorithm)[15]은 문자열의 배열로 표현된 해를 최적화시키는 진화 알고리즘의 일종으로, 포트폴리오에 편입될 수 있는 수많은 종목들 중 목적함수를 최적화시키는 종목을 선택하는 자산 선택 문제에서 좋은 성능을 보이는 것으로 알려져 있다[16]. 차분 진화(differential evolution)[17]는 실수(real-valued)의 배열로 표현된 해를 최적화시키는데 효과적인 방법으로, 선택된 자산의 편입비율을 결정하는 자본 할당 문제에서 좋은 성능을 보여주었다[18].

포트폴리오 선택의 기준이 되는 목적함수를 최적화하기 위하여 MM(Majorization-Minimization) 알고리즘[9], Neural

Network[19], Projected Gradient Algorithm과 같은 nonlinear optimization[12] 등의 방법들이 사용되었다. 이들은 목적함수를 최적화시키는 단 하나의 해를 반환하는 반면, 진화 알고리즘[5,16,18]은 목적함수를 최적화시키는 여러 후보해들을 반환한다. 진화 알고리즘의 이러한 특성은 제안 방법에서 목적 함수 이외의 조건을 고려하여 포트폴리오를 선택하기 위해 이용되었다. 추가적으로, 진화 알고리즘은 목적함수와 제약 조건의 형태에 구애받지 않고 최적화를 수행할 수 있는 유연성과 병렬적으로 실행될 수 있는 연산들이 많아 병렬처리의 혜택을 많이 받을 수 있다는 점에서 활용도가 크다는 장점을 가진다.

3. 몬테카를로-유전 알고리즘

유전 알고리즘이 자산 선택 문제에서 좋은 성능을 보여주었으나, 무한대에 가까운 해공간에서 최적의 해집합을 선택하는 것은 여전히 쉽지 않은 문제이다. 알고리즘의 효율성을 높이기 위하여 전체 종목에서 최적의 해집합을 선택하기 전, 중간 크기의 해집합을 몬테카를로 샘플링하는 과정을 추가한 몬테카를로-유전 알고리즘[20]이 제안되었으며, 이는 자산 선택 문제에 적용되어 좋은 성과를 보여주었다[21,22].

III. 관련 연구

[7]은 지수 상향 추종에 대한 초기 연구로서 Markowitz 모델을 지수 상향 추종 문제에 적용하고자 하였다. 이를 위해 평균 초과수익률과 포트폴리오 베타에 대한 제약조건을 추가하고 초과수익률의 분산을 추종 오차로 정의한 최적화 문제를 제안하였다.

[5]는 초과수익률의 분산으로 정의되던 기존의 추종 오차가 초과수익률의 편향(bias)을 무시하는 문제점이 있음을 지적하고 이를 보완하기 위해 초과수익률의 제곱평균제곱근(ete, empirical tracking error)을 추종 오차로 제안하였다. 그리고 이를 평균 초과수익률(er, excess return)과 선형 조합하여 일반화된 목적함수(ete_er(λ))를 제안하였다.

$$ete = \sqrt{\sum_{t=1}^T (r_t - R_t)^2 / T} \quad (1)$$

$$er = \sum_{t=1}^T (r_t - R_t) / T \quad (2)$$

$$ete_er(\lambda) \quad (\lambda \in [0, 1]) \quad (3)$$

$$= (1 - \lambda) \sqrt{\sum_{t=1}^T (r_t - R_t)^2 / T} - \lambda \sum_{t=1}^T (r_t - R_t) / T$$

[8,9]는 포트폴리오의 수익률이 지수의 수익률보다 낮은 경우만을 위험으로 간주하는 하방 위험(downside risk)을 목적 함수로 사용하였다. 지수 상향 추종 문제에 하방 위험을 적용하는 아이디어는 이후 Sharpe ratio에 하방 위험을 적용시킨 Sortino ratio[10] 등의 목적함수를 사용하는 연구들의 발판이 되었다.

$$downside\ risk = \sqrt{\sum_{t=1}^T \min(r_t - R_t, 0)^2 / T} \quad (4)$$

[12]은 포트폴리오의 이익과 위험을 각각 포트폴리오, 지수 수익률의 평균의 차와 분산의 차로 나타내고 이를 곱한 식을 최대화하여 지수 상향 추종 포트폴리오를 선택하였다. 이를 통해 초과수익률을 사용하는 기존 연구들과는 다른 방식을 통해 과레토 최적해를 찾고, 이를 시각적으로 해석하는 방법을 제안하였다.

IV. 연구 동기 및 기여

기존 연구들은 포트폴리오의 평균 초과수익률과 추종 오차 간의 상충관계(trade-off)를 적절히 나타내는 특정 목적함수를 선택하여 최적화하고자 하였다. [11]에서 지적한 바와 같이 급변하는 금융환경 속에서 항상 다른 목적함수보다 좋은 성과를 내는 궁극의 목적함수를 찾는 것은 불가능하나, 대부분의 연구들은 특정 목적함수를 선택하거나 여러 목적함수들에 대한 사후분석 결과를 보고하는 데에 그쳤다. 이는 다양한 목적함수들이 선택될 가능성이 있는 상황에서 실질적으로 최적의 결과라고 할 수 없을 뿐만 아니라 리밸런싱을 할 때마다 미래의 위험에 대비하여 최적화된 포트폴리오를 선택해야 하는 실전 투자에서 사용할 수 없다는 한계점을 가진다. 또한, 기존 연구들에서 학습 구간에서 성능을 높이는 데에 집중하여 일반화(generalization) 성능을 향상시키기 위한 직접적인 논의가 이루어지지 않고 있으나, 학습 기간의 성능이 테스트 기간에서 이어질 수 있도록 일반화 성능을 높이는 것은 포트폴리오의 위험 관리 관점에서 매우 중요하다.

본 논문의 기여는 상술한 문제점들을 보완하기 위한 새로운 포트폴리오 선택 방법을 제안함으로써 일반화 성능을 항상 시킨 데 있다. 특히, 제안된 방법은 모든 형태의 목적함수를 추가·적용할 수 있어 실용적으로 사용할 수 있을 뿐만 아니라, 투자 목적 이외의 문제에도 다양하게 활용될 수 있다는 점에서 큰 의미를 가진다.

V. 제안 방법

이번 장에서 최적의 지수 상향 추종 포트폴리오를 선택하는

제안 방법에 대하여 설명한다.

1. 3단계 포트폴리오 선택 알고리즘

지수를 구성하는 종목들 중 일부만을 사용하여 지수를 상향 추종하기 위해, 본 논문은 진화 알고리즘 기반 3단계 포트폴리오 선택 알고리즘을 제안한다(알고리즘 1).

알고리즘 1 3StagePortSelect (학습변수 p , 구간 T_i)

학습변수 p : (목적함수 o , 필터링 조건 f)

구간 T_j 에 대해서 다음을 수행

1단계. 자산 선택

목표 지수를 구성하는 전체 종목들 중 몬테카를로-유전 알고리즘을 통해 목적함수 o 를 최적화시키는 종목들로 구성된 동일 가중치 포트폴리오를 다수 생성

2단계. 차분 할당

생성된 포트폴리오들에 대하여 차분 진화를 적용하여 목적함수 o 를 최적화시키는 자산 편입비율을 할당

3단계. 포트폴리오 필터링

생성된 후보 포트폴리오들 중 목적함수 기준으로 상위 포트폴리오들에 대해서 필터링 조건 f 가 최적인 포트폴리오를 선택

return 최종적으로 선택된 포트폴리오

1단계와 2단계는 몬테카를로-유전 알고리즘과 차분 진화를 사용하여 주어진 목적함수 o 를 최적화시키는 포트폴리오들을 생성하는 단계이다. 기존 연구들은 이 과정 이후 생성된 포트폴리오들 중 최적의 목적함수 값을 가지는 포트폴리오를 선택하였다. 그러나 본 논문은 새로운 포트폴리오 필터링 단계를 도입하여 이전 단계에서 생성된 포트폴리오들 중 목적함수 기준 상위 포트폴리오들을 선택하고, 이들이 충분히 학습 데이터에 적합된(fitted) 상황을 고려한다. 이에 따라, 목적함수 o 와 구별된 필터링 조건 f 를 사용하여 최적의 포트폴리오를 선택함으로써 학습 기간이 아닌 테스트 기간에서 성능이 높을 것으로 예상되는 포트폴리오를 선택한다.

2. 양상을 학습 알고리즘

기존의 연구들은 특정한 목적함수를 사용하여 포트폴리오를 선택하였다. 그러나 [11]에서 지적한 바와 같이 금융시장처럼 비정상적(nonstationary)이고 다양한 위험들이 존재하는 환경에서 한 가지 목적함수에 최적화된 포트폴리오가 항상 좋은 성과를 내는 것은 불가능한 일이다. 따라서 본 논문은 하나의 목적함수에 최적화된 포트폴리오를 선택하기보다, 여러 개의 목적함수 각각에 대하여 최적화된 포트폴리오들을 결합하는 양상을 학습 알고리즘 제안한다(알고리즘 2).

각각의 학습변수(P_i)를 사용하여 선택된 포트폴리오들을 효

과적으로 조합하기 위하여 총 m 개의 학습 구간(T_1, \dots, T_m) 및 검증 구간(V_1, \dots, V_m)으로 구성된 데이터를 사용한다. m 개의 (학습 구간(T_j), 검증 구간(V_j)) 데이터 쌍에 대하여, 먼저 3단계 포트폴리오 선택 알고리즘을 통해 포트폴리오($portfolio_{ij}$)를 선택한 후, 검증 구간(V_j)에서 선택된 포트폴리오의 평가지표($metric_j$)를 계산한다. 총 m 개의 검증 구간으로부터 m 개의 평

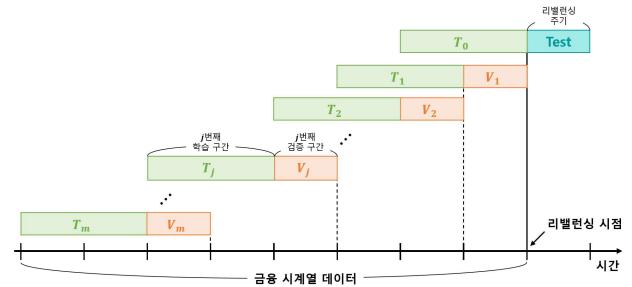


그림 1. 금융 시계열 데이터의 구성

알고리즘 2 Ensemble learning

(목적함수 집합 O , 필터링 조건 집합 F , 금융 시계열 데이터 D)

$$D_j = \begin{cases} T_j, & j=0 \\ (T_j, V_j), & j>0, j=0, \dots, m \end{cases}$$

$P = \{P_i = (o, f) | o \in O, f \in F, i=1, \dots, n\}$ // 학습변수 집합

for $i \leftarrow 1$ to n

for $j \leftarrow 1$ to m

$portfolio_{ij} \leftarrow 3StagePortSelect(P_i, T_j)$

$metric_j \leftarrow V_j$ 기간에서 $portfolio_{ij}$ 의 평가지표

end for

$w_i \leftarrow \frac{1}{m} \sum_{j=1}^m metric_j$ // P_i 에 대한 평균 평가지표

$portfolio_i \leftarrow 3StagePortSelect(P_i, T_0)$

end for

$w_port \leftarrow \sum_{i=1}^n w_i \cdot portfolio_i$

$opt_portfolio \leftarrow w_port$ 종목들 중 편입비율 상위 K 개 종목 선택

return $opt_portfolio$

가지표가 계산되는데 이들의 평균값이 해당 학습변수(P_i)가 갖는 중요도, 즉 가중치(w_i)가 된다. 상정된 모든 학습변수(P_1, \dots, P_n)에 대하여 이 과정을 반복하면 모든 학습변수에 대한 가중치(w_1, \dots, w_n)가 구해진다. 학습 구간 T_0 에서 모든 학습 변수(P_i)에 대하여 선택된 포트폴리오의 편입비율($portfolio_i$)과 계산된 가중치(w_i)에 대한 가중치합을 한 결과, 새로운 포트폴리오(w_port)가 생성된다. 마지막으로 편입 종목의 개수를 동일하게 유지하기 위하여 편입비율 상위 K 개 종목을 선택함으로써 최종 포트폴리오($opt_portfolio$)가 선택된다.

VI. 실험

이번 장에서 가장 대표적인 지수인 S&P500과 편입종목들의 종가를 이용하여 약 5년간 제안 방법의 성능을 평가한다. 일정 주기마다 리밸런싱을 발생시켜 학습과 테스트를 반복하는 rolling window 학습을 통해 실험을 수행하고 최종적으로 테스트 구간들을 이어 성능을 평가하였다.

실험은 총 5개의 서버를 사용하여 수행되었다. 1대의 Intel Core i9-10900X@3.70GHz, 128GB RAM, 2대의 Intel Xeon Silver 4210R@2.40GHz, 256GB RAM, 1대의 Intel Xeon Silver 4110@2.10GHz, 256GB RAM, 1대의 Intel Xeon Silver 4210@2.20GHz, 512GB RAM 환경에서 수행되었다.

1. 실험 설정

제안 방법의 성능을 테스트하기 위한 기간으로 2016/1/4 ~ 2020/10/7, 약 5년의 기간을 설정하였다. 편입자산의 개수는 10개와 20개, 리밸런싱 주기는 3개월(60일), 6개월(120일)에 대하여 각각 실험하였다. 학습 구간(T_j)의 길이는 120일, 검증 구간(V_j)의 길이는 리밸런싱 주기와 동일하게 설정하였다. 검증 데이터의 개수(m)는 검증 구간 길이의 총합이 240일이 되도록 설정하였다. 각각의 목적함수에 대하여 최적의 포트폴리오를 선택할 때 각 자산의 최대 편입비율은 20%를 넘지 않도록 하였으며, 리밸런싱 시 발생하는 거래비용은 매수금액의 0.015%, 매도금액의 0.315%로 설정하였다.

진화 알고리즘은 난수를 사용하는 확률적 알고리즘(stochastic algorithm)으로 20개의 random seed로 실험하고 이를 중 가장 최적의 값을 가지는 포트폴리오를 선택하였다.

obj 로 표기된 목적함수로 기준 방법에서 가장 대표적으로 사용되는 $\text{ete_er}(\lambda)$ ($\lambda \in \{0.3, 0.5, 0.7\}$), downside risk 를 사용하였으며 양상을 학습 알고리즘을 사용한 제안 방법은 optimized- obj 로 표기하였다. filter로 표기된 필터링 조건에 대한 실험을 위하여 여러 가지 조건들을 제안하였다. 포트폴리오 자산들의 시가총액의 가중치합이 최대인 포트폴리오를 선택하는 방법을 cap, 각 종목에 대한 편입비율이 가장 분산되어 투자된 포트폴리오를 선택하는 방법을 diversified, 지수와 포트폴리오 간의 상관계수가 가장 큰 포트폴리오를 선택하는 방법을 corr이라 표기하였다.

포트폴리오의 성능을 평가하기 위하여 다양한 지표들이 사용되었다. Sharpe ratio, Information ratio, VWR(Variability-Weighted Return)[23], 평균 초과수익률, MDD를 통해 최종 성능을 평가하였다. Sharpe ratio는 수익률의 평균과 표준편차의 비율을, Information ratio는 지수 대비 포트폴리오의 초과수익률에 대한 평균과 표준편차의 비율을 나타낸다. VWR은 Sharpe ratio의 문제점을 보완하기 위해 제안

된 평가지표로서, 변동성(σ)을 고려하여 연평균 성장률(\bar{v})에 가중치를 부여한 값이다. 기준 대상과 동일한 연평균 성장률을 가지는 무변동 포트폴리오와 기준 대상 간의 가치 차이의 표준편차를 변동성으로 간주하고, 투자자의 성향에 따라 정해지는 매개변수(σ_m, τ)에 의해 가중치로 변환된 변동성을 연평균 성장률에 곱하여 계산한다.

$$\text{Sharpe ratio} = \frac{\text{mean}(r)}{\text{stddev}(r)} \quad (5)$$

$$\text{Information ratio} = \frac{\text{mean}(r-R)}{\text{stddev}(r-R)} \quad (6)$$

$$VWR = 100(e^{250\bar{v}} - 1) \times (1 - (\sigma/\sigma_m)^\tau) \quad (7-1)$$

$$\bar{v} = \ln(v_T/v_0)/T \quad (7-2)$$

$$v_t' = v_0 \times e^{\bar{v}t} \quad (7-3)$$

$$d_t = (v_t - v_t')/v_t' \quad (7-4)$$

$$\sigma = \text{stddev}(d) \quad (7-5)$$

평균 초과수익률은 지수에 대한 포트폴리오의 초과수익률의 평균을 의미한다. MDD(Maximum Drawdown)는 테스트 구간 내에서 최저가 이전의 최고가와 최저가의 차이를 최저가로 나눈 값으로 최대 낙폭을 의미하며, 값이 작을수록 최대 낙폭이 커 위험도가 높은 것으로 해석된다.

$$\text{평균 초과수익률} = \sum_{t=1}^T (r_t - R_t)/T \quad (8)$$

$$MDD = \frac{\text{maximum value(before minima)} - \text{minimum value}}{\text{minimum value}} \quad (9)$$

양상을 학습의 검증 기간에서 포트폴리오에 대한 평가지표로 VWR을 변형시킨 VWS(Variability-Weighted Spread)를 정의하여 사용하였다. VWS는 항상 양수의 값을 가지는 특성으로 인해 각 학습변수에 대한 성과를 편입비율로 변환하기 쉽다는 장점을 가진다. VWR과 VWS에서 사용되는 매개변수(σ_m, τ)의 값은 (0.2, 4)를 사용하였다.

$$VWS = e^{250\bar{s}} \times (1 - (\sigma/\sigma_m)^\tau) \quad (10-1)$$

$$s_t = v_t/V_t \quad (10-2)$$

표 1. S&P500 지수에 대한 상향 추종 포트폴리오 성과

리밸런싱 주기	종목 개수(K)	학습변수		Sharpe ratio	Information ratio	VWR	평균 초과수익률	MDD
		목적함수(obj)	필터링 조건(filter)					
Benchmark index		S&P500		0.670		11.625		-9.30
3개월(60일)	10	downside risk		0.687	0.060	11.870	1.68E-05	-10.80
		ete_er($\lambda=0.3$)		0.669	0.157	12.307	3.64E-05	-6.56
		ete_er($\lambda=0.5$)		0.449	-0.636	7.202	-1.50E-04	-9.15
		ete_er($\lambda=0.7$)		0.628	0.026	11.483	8.79E-06	-10.15
	20	optimized-obj		0.673	0.128	12.356	3.45E-05	-7.47
		optimized-obj	cap	0.974	1.153	19.839	2.93E-04	-6.08
6개월(120일)	10	optimized-obj	diversified	0.887	0.651	16.808	1.93E-04	-7.89
		optimized-obj	corr	0.657	0.052	11.803	1.30E-05	-7.37
		downside risk		0.668	0.055	11.818	1.10E-05	-10.56
		ete_er($\lambda=0.3$)		0.546	-0.379	9.401	-6.99E-05	-10.58
	20	ete_er($\lambda=0.5$)		0.673	0.122	12.066	2.14E-05	-7.10
		ete_er($\lambda=0.7$)		0.535	-0.320	9.324	-7.74E-05	-8.32
평균	10	optimized-obj		0.750	0.444	13.985	8.86E-05	-7.83
		optimized-obj	cap	0.752	0.562	14.515	1.07E-04	-7.36
		optimized-obj	diversified	0.723	0.337	13.187	7.26E-05	-8.26
		optimized-obj	corr	0.663	0.024	11.661	4.25E-06	-10.36
	20	downside risk		0.427	-0.669	6.867	-1.66E-04	-10.62
		ete_er($\lambda=0.3$)		0.728	0.315	13.293	7.14E-05	-9.71
	6개월(120일)	ete_er($\lambda=0.5$)		0.620	-0.112	10.545	-2.79E-05	-9.15
		ete_er($\lambda=0.7$)		0.739	0.341	14.634	1.24E-04	-8.61
평균	20	optimized-obj		0.601	-0.058	10.786	-1.66E-05	-6.24
		optimized-obj	cap	0.814	0.833	15.759	1.52E-04	-7.37
		optimized-obj	diversified	0.635	0.024	11.355	5.34E-06	-12.36
		optimized-obj	corr	0.810	0.572	15.214	1.34E-04	-5.37
	6개월(120일)	downside risk		0.756	0.471	14.164	9.65E-05	-10.10
		ete_er($\lambda=0.3$)		0.597	-0.191	10.558	-3.36E-05	-10.58
평균	10	ete_er($\lambda=0.5$)		0.690	0.161	12.337	2.92E-05	-7.10
		ete_er($\lambda=0.7$)		0.593	-0.119	10.252	-3.29E-05	-10.23
		optimized-obj		0.835	0.778	16.024	1.64E-04	-9.68
		optimized-obj	cap	0.890	0.829	17.770	2.23E-04	-6.22
	20	optimized-obj	diversified	0.456	-0.513	7.360	-1.44E-04	-38.96
		optimized-obj	corr	0.539	-0.451	9.221	-8.16E-05	-9.82
평균	10	downside risk		0.635	-0.021	11.180	-1.04E-05	-10.52
		ete_er($\lambda=0.3$)		0.635	-0.025	11.390	1.08E-06	-9.36
		ete_er($\lambda=0.5$)		0.608	-0.116	10.538	-3.18E-05	-8.13
		ete_er($\lambda=0.7$)		0.624	-0.018	11.423	5.62E-06	-9.33
	20	optimized-obj		0.715	0.323	13.288	6.76E-05	-7.81
		optimized-obj	cap	0.857	0.844	16.971	1.94E-04	-6.76
	6개월(120일)	optimized-obj	diversified	0.675	0.125	12.178	3.17E-05	-16.87
		optimized-obj	corr	0.667	0.049	11.975	1.73E-05	-8.23

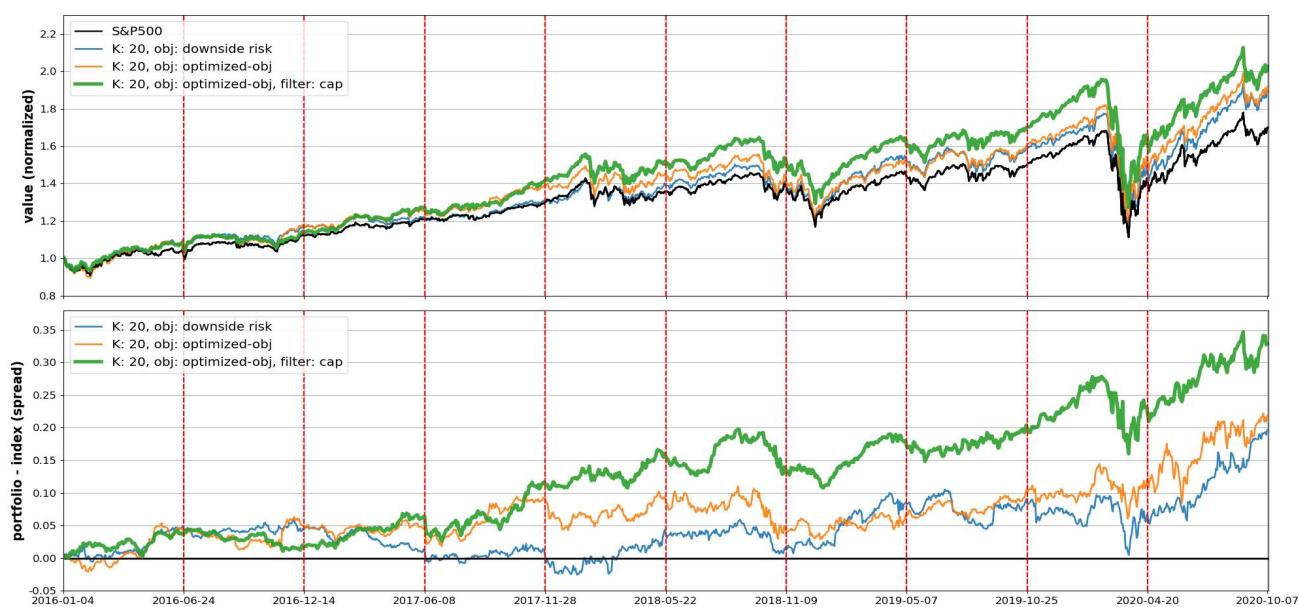


그림 2. S&P500 지수를 상향 추종하는 포트폴리오 (편입종목 개수: 20, 리밸런싱 주기: 6개월)

$$\bar{s} = \ln(s_T/s_0)/T \quad (10-3)$$

$$s_t' = s_0 \times e^{\bar{s}t} \quad (10-4)$$

$$d_t = (s_t - s_t')/s_t' \quad (10-5)$$

$$\sigma = \text{stddev}(d) \quad (10-6)$$

(s_t : t 시점에서 지수 대비 포트폴리오의 가치)

사용된 평가지표들은 모두 큰 값일수록 좋은 특성을 가진 포트폴리오라는 것을 의미한다.

2. 실험 결과

상향 추종 포트폴리오의 성능을 판단하기 위해 리밸런싱 주기, 편입종목의 개수가 다른 4가지 실험에서 테스트 기간 동안의 평가지표를 측정하였다. 표 1은 S&P500 지수에 대한 포트폴리오의 성과를 보여준다. 거의 모든 지표에서 3단계 포트폴리오 선택 알고리즘과 양상을 학습을 사용한 제안 방법이 기존 방법에 비하여 평균적으로 더 좋은 성능을 보여주었음을 확인할 수 있다. 특히, 포트폴리오 필터링 조건들 중 시가총액을 고려한 방법(cap)이 가장 좋은 성과를 보였는데 [24]에서 언급된 바와 같이 자산의 시가총액은 변동성과 음의 상관관계를 가진다는 특성이 일반화 성능을 높이는데 효과적인 역할을 한 것으로 분석된다. 또한, 각 실험에서 Information ratio를 기준으로 가장 좋은 성능을 보여준 목적함수를 볼드체로 나타내었는데, 동일한 지수에서 같은 기간 동안 수행된 실험임에도 불구하고 조건이 약간 달라지는 것만으로도 각각의 목적함수로 최적화시킨 결과가 크게 변하는 것을 확인할 수 있다. 이러한 결과들은 특정 목적함수에 의존하지 않는 양상을 학습이 올바른 접근법이라는 것을 나타내며, 3단계 포트폴리오 선택 알고리즘이 일반화 성능을 높이는 효과가 있다는 것을 보여준다. 그림 2은 S&P500 지수에 대하여 20개의 편입종목과 리밸런싱 주기를 6개월로 하여 실험한 결과로 제안 방법을 사용하여 최적화시킨 포트폴리오가 지수를 가장 안정적으로 상향 추종하고 있음을 확인할 수 있다.

VII. 결 론

부분복제 지수 상향 추종 포트폴리오를 생성하기 위해 사용되던 대부분의 기존 방법들은 하나의 목적함수만을 선택하고, 일반화 성능을 올리기 위한 직접적인 방법을 가지고 있지 않다는 한계점이 있다. 이에 본 논문은 하나의 목적함수 대신 여러 목적함수를 적절히 조합하여 포트폴리오를 구성하는 양상을 학습과 일반화 성능을 높이기 위하여 목적함수 이외의 값을 고려할 수 있는 진화 알고리즘 기반 3단계 포트폴리오 선택 알고리

즘을 제안하였다. 제안 방법의 효과를 확인하기 위하여 약 5년 간의 S&P500 지수 데이터를 사용한 실험을 수행하였으며, 편입종목의 개수와 리밸런싱 주기가 다른 여러 환경에서 제안 방법이 기존 방법보다 더 좋은 성능을 내는 것을 여러 평가지표를 통해 보여주었다.

제안된 3단계 포트폴리오 선택 알고리즘과 양상을 학습은 새로운 목적함수와 필터링 조건을 추가함으로써 쉽게 변형 및 적용이 가능하다는 장점을 가지고 있다. 좋은 성과들이 보고된 목적함수들과 일반화 성능에 영향을 미칠 수 있는 포트폴리오 베타, 업종(섹터) 편입비율 등의 적절한 금융 변수들을 활용한다면 제안된 모델의 성능을 더 높일 수 있을 것이다.

REFERENCES

- [1] B. M. Barber and T. Odean, "Trading is hazardous to your wealth: The common stock investment performance of individual investors," *Journal of Finance*, vol. 55, no. 2, pp. 773 - 806, Apr., 2000.
- [2] Ian R. Appel, Todd A. Gormley and Donald B. Keim, "Passive investors, not passive owners," *Journal of Financial Economics*, vol. 121, no. 1, pp. 111-141, Jul., 2016.
- [3] Percentage of Large-Cap funds that underperformed the S&P 500(2020). <https://www.spglobal.com/spdji/en/research-insights/spiva/#/reports> (accessed Jul., 23, 2021).
- [4] Juan Diaz, Juan, et al., "Index fund optimization using a hybrid model: genetic algorithm and mixed-integer nonlinear programming," *The Engineering Economist*, vol. 64, no. 3, pp. 298–309, Jul., 2019.
- [5] J. E. Beasley, N. Meade, and T.-J. Chang, "An evolutionary heuristic for the index tracking problem," *European Journal of Operational Research*, vol. 148, no. 3, pp. 621-643, Aug., 2003.
- [6] O. Strub and N. Trautmann, "A two-stage approach to the UCITS-constrained index-tracking problem," *Computers & operations research*, vol. 103, pp. 167-183, Oct., 2019.
- [7] Richard Roll, "A mean/variance analysis of tracking error," *The Journal of Portfolio Management*, vol. 18, no. 4, pp. 13-22, 1992.
- [8] A. A. Gaivoronski, S. Krylov, and N. Van der Wijst, "Optimal portfolio selection and dynamic benchmark tracking," *European Journal of Operational Research*, vol. 163, no. 1, pp. 115 - 131, May, 2005.

- [9] K. Benidis, Y. Feng and D. P. Palomar, "Sparse Portfolios for High-Dimensional Financial Index Tracking," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 155-170, Jan., 2018.
- [10] N. Meade and J. E. Beasley, "Detection of momentum effects using an index out-performance strategy," *Quantitative Finance*, vol. 11, no. 2, pp. 313-326, Apr., 2010.
- [11] H. Li, Qin Huang, and Baiyi Wu, "Improving the naive diversification: An enhanced indexation approach," *Finance Research Letters*, vol. 39, Mar., 2021.
- [12] Francesco Cesarone, Lorenzo Lamariello and Simone Sagratella, "A risk-gain dominance maximization approach to enhanced index tracking," *Finance Research Letters*, vol. 29, pp. 231-238, Jun., 2019.
- [13] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77 - 91, 1952.
- [14] H. Markowitz, "Portfolio Selection: Efficient Diversification of Investments," Cowles Foundation Monograph, vol. 16, 1959.
- [15] Holland, J. H., "Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence," MIT Press, pp. 89-120, 1992.
- [16] Margherita Giuzio, "Genetic algorithm versus classical methods in sparse index tracking," *Decisions in Economics and Finance*, vol. 40, no. 1, pp. 243-256, Nov., 2017.
- [17] Rainer Storn and Kenneth Price, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces," *Journal of global optimization*, vol. 11, no. 4, pp. 341-359, 1997.
- [18] Andriopoulos Kostas, et al., "Portfolio optimization and index tracking for the shipping stock and freight markets using evolutionary algorithms," *Transportation Research Part E: Logistics and Transportation Review*, vol. 52, pp. 16-34, 2013.
- [19] Hongbing Ouyang, Xiaowei Zhang and Hongju Yan, "Index tracking based on deep neural network," *Cognitive Systems Research*, vol. 57, pp. 107-114, Oct., 2019.
- [20] Thomas Berkemeier, et al., "Monte Carlo genetic algorithm (MCGA) for model analysis of multiphase chemical kinetics to determine transport and reaction rate coefficients using multiple experimental data sets," *Atmospheric Chemistry and Physics*, vol. 17, no. 12, pp. 8021-8029, 2017.
- [21] 김정현, 이주홍, "몬테카를로 유전 알고리즘을 사용하여 축소된 탐색 공간에서 자산을 효율적으로 선택하는 방법에 관한 연구," 한국지능시스템학회 논문지, 제30권, 제1호, 21-27쪽, 2020년 2월
- [22] 윤동진, 이주홍, 송재원, "몬테카를로 유전 알고리즘을 활용한 부분복제 지수 추종," 한국정보처리학회 학술대회논문집, 제27권, 제2호, 751-754쪽, 2020년 11월
- [23] Improved Sharpe Ratio, using log returns(2021). <https://www.crystalbull.com/sharpe-ratio-better-with-log-returns/> (accessed Jul., 23, 2021).
- [24] Richard W. Sias, "Volatility and the institutional investor," *Financial Analysts Journal*, vol. 52, no. 2, pp. 13-20, Apr., 1996.
- [25] 이윤선, 이주홍, 최범기, 송재원, "비정형, 정형 데이터의 이미지 학습을 활용한 시장예측," 스마트미디어저널, 제10권, 제2호, 16-21쪽, 2021년 06월
- [26] 천성길, 이주홍, 최범기, 송재원, "대규모 외생 변수 및 Deep Neural Network 기반 금융 시장 예측 및 성능 향상," 스마트미디어저널, 제9권, 제4호, 26-35쪽, 2020년 12월
- [27] 하기수, 정현태, 박아론, 백성준, "양자 유전알고리즘을 이용한 특징 선택 및 성능 분석," 스마트미디어저널, 제1권, 제1호, 36-41쪽, 2012년 3월

저자소개



윤동진(학생회원)
 2019년 인하대학교 컴퓨터공학과 학
 사 졸업.
 2019년 ~ 현재 인하대학교 전기컴퓨
 터공학과 석사과정.

<주관심분야 : 진화 알고리즘, 병렬처리>



이주홍(정회원)
 1983년 서울대학교 전자계산기공학과
 학사 졸업.
 1985년 서울대학교 전자계산기공학과
 석사 졸업.
 2001년 KAIST 컴퓨터공학전공 박사
 졸업.
 2001년 ~ 현재 인하대학교 컴퓨터공
 학과 교수.

<주관심분야 : 머신러닝, 금융투자공학>



최범기(정회원)
 1986 서울대학교 수학과 학사 졸업.
 1995 Florida State university
 computer science 석사 졸업.
 2017 ~ 현재 인하대학교 컴퓨터공학
 과 겸임교수.

<주관심분야: 머신러닝, 금융투자공학>



송재원(정회원)
 2005년 성공회대학교 컴퓨터공학 학
 사 졸업.
 2007년 인하대학교 컴퓨터공학 석사
 졸업.
 2013년 인하대학교 컴퓨터공학 박사
 졸업.
 2013~2014 대구경북과학기술원 개발
 팀장.
 2014~2017 인천국제공항공사 선임연구원.
 2017~2017 가천길대학병원 연구교수.
 2017~현재 (주)밸류파인더스 대표이사.

<주관심분야: 머신러닝, 데이터마이닝, 금융IT>