몬테카를로 유전 알고리즘을 활용한 부분복제 지수 추종

윤동진*, 이주홍*, 송재원**
*인하대학교 전기컴퓨터공학과
**(주)밸류파인더스
djyoon1125@naver.com, juhong@inha.ac.kr, jwsong03@gmail.com

Sparse Index Tracking Using Monte-Carlo Genetic Algorithm

Dong-Jin Yoon*, Ju-Hong Lee*, Jae-Won Song**
*Dept. of Computer Engineering, Inha University
**ValueFinders Co. Ltd.

요 약

본 논문은 지수를 추종하기 위해 유전 알고리즘에 몬테카를로 샘플링을 추가한 방법을 제안한다. 몬테카를로 샘플링을 통해 효율적으로 축소된 탐색공간을 탐험하는 유전 알고리즘은 최적의 종목들을 선택한다. 제안된 방법을 KOSPI200 지수 추종에 대하여 실험하였다. 제안된 방법이 몬테카를로 샘플링을 사용하지 않는 유전 알고리즘에 비해 지수 추종 오차가 더 낮고 더 빠르게 수렴하는 것을 보여주었다.

1. 서론

금융시장은 여러 개의 자산들로 복합적으로 구성되어 있다. 표준화된 방법을 통해 금융시장의 성과를 하나의 집약된 수치로 나타낸 것을 지수(financial index)라 한다. 지수의 움직임을 따라가도록(index tracking) 구성한 펀드를 인덱스 펀드(index fund)라 한다. 펀드 매니저가 펀드 운용에 적극적으로 개입하는 액티브 펀드(active fund)에 비해 인덱스 펀드는 장기적으로 낮은 리스크를 가지면서도 안정적인 수익률을 내는 것으로 알려져있다.[1, 2] 특히 S&P500 지수의 경우, 최근 5년간 약 80%의 펀드들이 지수의 수익률을 따라오지 못하는 성과를 내어 액티브 펀드들의 평균 수익률은 인덱스 펀드의 수익률에 미치지 못하였음을 보여주었다.[3] 이에 따라 지수를 추종하는 인덱스 펀드의 인기가 치솟는 현상이 나타나게 되었다.

지수를 추종하는 방법은 목표 지수의 편입종목을 모두 사용하는 완전복제(full replication) 방식과 일부 종목만을 사용하는 부분복제(sparse replication) 방식으로 나뉜다. 완전복제 방식은 지수의 정보만 알 수 있다면 구현하기 쉽다는 장점이 있지만 수백, 수천 개의 종목을 관리하는데 발생하는 거래비용, 슬리피지[1], 비유동성 등의 문제로 인해 부분복제 방식이 주요한 연구주제가 되었다[4].

기존의 지수 추종을 위한 방법들로 LassoGP[5], CyCoDe[6], SLAIT[4] 등 수학적 최적화 기법을 활용한 방법들과 유전 알고리즘을 사용한 인공지능 기법들이 연구되었다.[7] 목표 지수의 편입종목들 중 최적의 부분집합을 탐색하는데 유전 알고리즘이 특히 좋은 성능을 보여주었다.[8] 기존의 방법들은 전체 종목으로 구성된 탐색공간을 초기 모집단으로 설정하였다. 무한대 크기의 탐색 공간으로 인해 효율적인 탐색 성능을 보장해 줄 수 없다는 본질적 한계를 가진다. 따라서 본 논문은 무한대 크기의 탐색공간을 보다 더 효율적으로 탐험하여 최적의 부분 집합을 찾아내기 위해 몬테카를로 샘플링[9, 10]을 사용하여 탐색 공간을 획기적으로 축소하고 유전 알고리즘으로 최적의 해를 찾는 방법을 제안하고자 한다. 효율적으로 축소된 탐색공간에서 수행되는 유전 알고리즘은 더 낮은 추종 오차로 더 빠르게 수렴한다. 이에 대한 실험으로 KOSPI200 지수에 대하여 적용시킨 결과를 보여주고, 유전 알고리즘을 사용하지 않는 방법과 함께 지수 추종 오차를 확인한다. 본 논문은 편입종목을 효율적으로 선택하는 방법에 집중하여 추종 오차를 비교할 때의 편입비율은 모든 종목이 동일하도록 설정하였다.

1) 매수호가 및 매도호가 차이에 의한 결제가격차이

2. 관련 연구

지수를 추종하기 위해 수학적 최적화 기법[4, 5, 6]과 유전 알고리즘[7, 9, 10]을 사용한 방법들이 연구되었다. 최적화하기 어려운 비볼록(non-convex) 목적함수를 쌍대성(duality)를 이용하여 해결 가능한 볼록(convex) 목적함수로 나누고 Gradient Projection 기법을 사용하여 최적화시키는 LassoGP [5]를 활용한 방법이 사용되었다. [6]은 각 차원 별로 Gradient Descent 알고리즘을 적용시킨 Coordinate Descent 알고리즘에 기반한 방법을 제안하였다. 목적함수를 해결 가능한 특정 볼록 함수로 한정시킨 다음 MM (Majorization-Minimization) 알고리즘을 사용하는 방법이 연구되었다.[4] 그리고 유전 알고리즘의 유전자를 이진 비트로 표현하거나[7, 9] 특정 지수가 가진 조건을 만족시키도록 표현하고 연산을 변형하여 지수를 추종하는 방법이 연구되었다.[10]

3. 제안 방법

지수를 구성하는 전체 종목의 개수를 N, 학습에 사용되는 기간을 T라 하고, 종목들의 종가를 V, 지수의 종가를 B로 나타낸다. 선택된 종목의 개수를 n이라 하고 종목의 편입비율을 W라 한다. 포트폴리오의 성과를 측정하기 위한 추종 오차로 목표 지수의 종가와 포트폴리오의 종가의 평균 제곱 오차인 ETE(Empirical Tracking Error)[4, 7]를 사용하였다.

<식 1> 추정 오차 계산식

$$V=\begin{bmatrix} v_1^{(1)} & \cdots & v_1^{(N)} \\ \vdots & \ddots & \vdots \\ v_T^{(1)} & \cdots & v_T^{(N)} \end{bmatrix},\ B=\begin{bmatrix} b_1 \\ \vdots \\ b_T \end{bmatrix}$$

$$W=\begin{bmatrix} w^{(1)} \\ \vdots \\ w^{(N)} \end{bmatrix}\quad s.t.\ w^{(i)}=\begin{cases} \frac{1}{n} & \text{if } i \text{ is selected} \\ 0 & o.w. \end{cases}$$

$$ETE(W)=\frac{1}{T}\|B-VW\|^2$$

3.1 Monte-Carlo sampling

기존의 방법들은 N개의 전체 종목을 유전 알고리즘의 초기 모집단으로 넘겨주었다. 본 논문은 유의미하게 탐색공간을 축소시키는 것으로 연구된 몬테카를로 샘플링[11, 12]을 사용하여 N개의 전체 종목 대신 효율적으로 선택된 K개의 자산으로 구성된 포트폴리오들을 유전 알고리즘의 초기 모집단으로 넘겨준다. 상세한 과정은 다음과 같다.

N개의 종목에서 M개의 종목으로 이루어진 포트폴리오 p_M개를 무작위 추출한 후, 낮은 추정 오차를 낸 상위 p_{MTOP}개 포트폴리오만을 남긴다. 남은 포트폴리오로부터 K개의 종목으로 이루어진 포트폴리오 p_K개를 무작위 추출한 후 상위 p_{KTOP}개 포트폴리오만 남긴다. 선택된 포트폴리오들을 유전 알고리즘의 초기 모집단으로 넘겨준다. 이 과정을 통해 유전 알고리즘을 가동하기 전 낮은 추정 오차를 가진 초깃값을 선별적으로 추출한다.

3.2 유전 알고리즘

유전 알고리즘은 Darwin의 적자생존 이론에 기초한 최적화 알고리즘이다.[13] 모집단을 초기화(initialization)한 후 부모 선택(parent selection), 재조합(recombination), 변이(mutation), 생존자 선택(survivor selection) 연산들을 하나의 세대로 하고 이를 반복시켜 최적의 해를 찾는다.[14] 몬테카를로 샘플링을 통해 넘어온 초깃값을 사용하는 유전 알고리즘은 최적값에 빠르게 수렴할 뿐만 아니라 더 낮은 추정 오차에서 수렴할 수 있다.

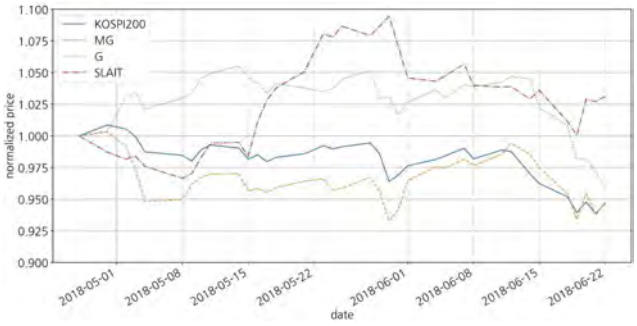
4. 실험

학습을 위해 2017년 12월 27일부터 2018년 4월 25일까지 약 4개월간 KOSPI200 지수에 편입된 종목들의 종가 데이터를 사용하였다. 검증을 위해 2018년 4월 26일부터 2018년 6월 25일까지 약 2개월간 동일 종목들의 종가 데이터를 사용하였다. 실험에 사용된 파라미터는 M, p_M, p_{MTOP}, K, p_K, p_{KTOP}은 각각 100, 1000, 10, 5, 100, 10 이며 3가지 지수 추종 알고리즘을 사용하여 성능을 비교하였다. MG는 몬테카를로 샘플링(MC)+유전 알고리즘, G는 유전 알고리즘, SLAIT[7]는 MM 알고리즘을 사용한 방법을 각각 의미한다.

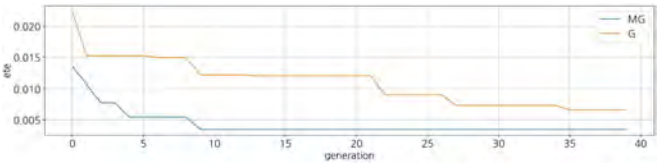
<표 1> 실험 결과 비교

	MG	G	SLAIT
ETE	**0.00043**	0.00250	0.00402

<그림 1> 알고리즘 별 지수 추종 비교



<그림 2> 유전 알고리즘 세대에 따른 ETE 변화 비교



<표 1>과 <그림 1>은 KOSPI200 지수와 몬테카를로 샘플링을 적용한 유전 알고리즘과 적용하지 않은 유전 알고리즘, MM 알고리즘을 사용한 방법의 지수 추종 결과를 비교한 것으로, 몬테카를로 샘플링을 적용했을 때 가장 좋은 결과를 보여주었다. 또한 <그림 2>는 몬테카를로 샘플링을 적용했을 때 유전 알고리즘이 좋은 초깃값에서 시작할 뿐만 아니라 더 작은 값으로 빠르게 수렴한다는 것을 보여주었다.

5. 결론

본 논문은 부분복제 지수 추종에 필요한 종목 선택을 위해 사용되었던 기존의 유전 알고리즘이 탐색 공간을 더 효율적으로 탐험할 수 있도록 몬테카를로 샘플링 과정을 추가하였다. 목표 지수의 전체 편입 종목에서 몬테카를로 샘플링을 통해 추정 오차가 낮은 여러 개의 종목들을 효과적으로 선택할 수 있었으며, 이를 초기 모집단으로 하여 수행시킨 유전 알고리즘은 그렇지 않은 경우보다 더 낮은 추정 오차로 더 빠르게 수렴할 수 있음을 KOSPI200 데이터를 사용한 실험을 통해 보여주었다.

본 논문은 최적의 종목을 선택하는 것에 집중하여 각 종목의 편입비율을 모두 동일하게 설정하였다. 향후 최적의 종목을 선택했던 유전 알고리즘에 이어 진화 알고리즘, mixed-integer programming 등의 방법을 통해 최적의 편입비율을 구하는 연구가 필요할 것이다. 그리고 유전 알고리즘의 경우 독립적으로 실행 가능한 연산들이 많기 때문에 GPU를 활용한 병렬처리로 가속화시켜 더 넓은 탐색공간을 빠르게 탐험할 수 있도록 하는 연구도 필요할 것이다.

6. Acknowledgement

이 논문은 2019년도 정부(과학기술정보통신부) 의 재원으로 한국연구재단의 기초연구사업(과제번호: 2019R1F1A1062094)과 2019년도 중소벤처기업부의 기술개발사업 지원에 의한 연구임 [과제번호: S2796242]

참고문헌

[1] Malkiel, Burton G. "Returns from Investing in Equity Mutual Funds 1971 to 1991", The Journal of Finance 50, no. 2, 549-72 (1995)

[2] BUSSE, JEFFREY A., AMIT GOYAL, and SUNIL WAHAL. "Performance and Persistence in Institutional Investment Management", The Journal of Finance 65, no. 2, 765-90 (2010)

[3] https://www.spindices.com/spiva/#/reports

[4] K. Benidis, Y. Feng and D. P. Palomar, "Sparse Portfolios for High-Dimensional Financial Index Tracking," in IEEE Transactions on Signal Processing, vol. 66, no. 1, pp. 155-170, 1 Jan.1, 2018

[5] Gasso, G., Rakotomamonjy, A., Canu, S. “Recovering sparse signals with a certain family of nonconvex penalties and DC programming“ IEEE Trans. Signal Process. 57(12), 4686 - 4698 (2009)

[6] Yen, Y., Yen, T. "Solving norm constrained portfolio optimization via coordinate-wise descent algorithms" Comput. Stat. Data Anal. 76, 737 - 759 (2014)

[7] Beasley, J., Meade, N., Chang, T.J. “An evolutionary heuristic for the index tracking problem” European Journal of Operational Research 148, 621 - 643 (2003)

[8] Giuzio, M. “Genetic algorithm versus classical methods in sparse index tracking”, Decisions in Economics and Finance 40, 243 - 256 (2017)

[9] Juan Díaz, María Cortés, Juan Hernández, Óscar Clavijo, Carlos Ardila & Sergio Cabrales. “Index fund optimization using a hybrid model: genetic algorithm and mixed-integer nonlinear

programming", The Engineering Economist, 64:3, 298-309 (2019)
[10] O. Strub, N. Trautmann, "A two-stage approach to the UCITS-constrained index-tracking problem", Computers & Operations Research 103, 167-183 (2019)
[11] Berkemeier, Thomas & Ammann, Markus & Krieger, Ulrich & Peter, Thomas & Spichtinger, Peter & Pöschl, Ulrich & Shiraiwa, Manabu & Huisman, Andrew. "Technical note: 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. 17. 8021-8029 (2017)
[12] 김정현, 이주홍, "몬테카를로 유전 알고리즘을 사용한 자산 선택", 한국정보처리학회 춘계 학술대회 VOL 26 NO. 01 PP. 0387 ~ 0389 (2019. 05)
[13] Holland, J. H. "Adaptation in natural and artificial systems: an introductory analysis with pplications to biology, control, and artificial ntelligence" University of Michigan Press (1975)
[14] A. E. Eiben and James E. Smith. "Introduction to Evolutionary Computing (2nd. ed.)" Springer Publishing Company, Incorporated (2015)