몬테카를로 유전 알고리즘을 사용한 자산 선택

김정현*, 이주홍*
*인하대학교 컴퓨터공학과
rusid275@naver.com
juhong@inha.ac.kr

Portfolio Asset Selection using Monte-Carlo Genetic Algorithm

Jung-Hyun Kim*, Ju-Hong Lee*
**Dept. of Computer Engineering, Inha University

요 약

자산 선택을 통해 최적 포트폴리오를 구성하는 데 있어 모든 경우의 수를 탐색하는 것은 불가능하다. 본 논문에서는 몬테카를로 시뮬레이션을 통해 탐색 공간을 축소하고, 축소된 탐색 공간에 유전 알고리즘을 적용함으로써 최적 포트폴리오에 근사하는 부최적 포트폴리오를 생성하는 방법을 제안한다.

1. 서론

자산 선택(Asset Selection)이란 개별 자산들을 선택하여 포트폴리오를 구성하는 투자 전략을 말한다. 자산 선택을 통한 최적 포트폴리오 구성은 n개의 자산 중 k개의 자산을 선택하는 조합 문제이며, 가능한 조합 중 샤프 지수(Sharpe ratio)가 높은 포트폴리오를 최적 포트폴리오라고 할 수 있다. 이 때 n과 k의 값이 증가하면 가능한 조합의 수가 기하급수적으로 증가하기 때문에 가능한 모든 경우를 탐색하여 최적해를 찾는 것은 시간적으로 불가능하다.

유전 알고리즘을 사용한 자산 선택에 관련된 기존 연구[2-6]들은 소규모 자산으로부터 최적 포트폴리오를 구성하는 데 유전 알고리즘이 효과적임을 보였으나, 실제 자산 시장에서 수천 개의 종목으로부터 최적 포트폴리오를 탐색할 때 발생하는 시간적 문제에 대해서는 다루지 않았다. 본 논문에서는 동일 가중치(Equally-weighted) 포트폴리오를 구성하기 위해 몬테카를로 시뮬레이션(Monte-Carlo simulation)을 통해 탐색 공간을 축소하고, 축소된 탐색 공간으로부터 유전 알고리즘을 사용해 최적 포트폴리오에 근사한 부최적 포트폴리오를 생성함으로써 자산 선택에 걸리는 시간을 효과적으로 단축하는 방법에 대하여 제안한다. 또한 최적 포트폴리오와 부최적 포트폴리오의 성능을 비교함으로써 본 논문에서 주장하는 방법론이 효과적임을 증명한다.

본 논문의 구성은 다음과 같다. 2장은 본 논문과 관련된 연구들을 소개하며, 3장은 본 논문이 제안하는 방법론이 서술되어 있다. 4장은 본 논문에서 진행한 실험이 서술되어 있다. 그리고 5장은 본 논문에 대한 결론이 서술되어 있다.

2. 관련 연구

Markowitz[1]는 포트폴리오의 수익을 극대화하고 위험을 최소화하기 위한 평균-분산 모델을 제시하여 자산 배분을 통한 최적 포트폴리오를 찾는 방법을 정립하였다. Arnone[2]은 자산 선택에 유전 알고리즘을 적용함으로써 포트폴리오를 최적화하는 모델을 제시하였다. Prisadarng[3]은 Multi-Objective 유전 알고리즘을 이용함으로써 카디널리티, Floor, Round-lot과 같은 현실적인 제약 조건 하에서 최적 포트폴리오를 찾고자 하였다. Lin[4]은 최소 거래 단위 제약 조건을 적용한 포트폴리오 최적화 문제를 유전 알고리즘을 적용하여 해결하고자 하였다. 그는 제안한 모델을 사용해 만든 포트폴리오가 효율적인 프론티어에 매우 근접했다고 주장했다. Soleimani[5]는 유전 알고리즘을 사용하여 최소 거래 단위, 카디널리티, Sector capitalization 제약을 적용한 포트폴리오 최적화 문제를 해결할 수 있음을 보였다. Bermúdez[6]는 카디널리티 제약 조건 하에서 효율적인 포트폴리오를 선택하기 위해 유전 알고리즘과 퍼지 순위 전략을 사용하였다.

3. 제안 방법

n개의 자산으로 이루어진 자산 집합 $S=\{a_1, a_2, \dots, a_{n-1}, a_n\}$의 부분집합인 포트폴리오 P의 샤프 지수(Sharpe ratio) s_P는 다음과 같이 계산된다.

$$E(R_P) = \sum_i w_i E(R_i)$$
$$\sigma_P^2 = \sum_i \sum_j w_i w_j \sigma_{ij}$$
$$s_P = E(R_P)/\sigma_P^2$$

R_P는 포트폴리오 P의 수익률이며, R_i는 포트폴리오 P를 구성하는 자산 i 의 수익률이다. w_i는 자산 i에 대한 가중치를 의미한다. σ_{ij}는 자산 i와 j의 공분산이다. 이 때 포트폴리오 P의 부분집합인 포트폴리오 Q의 샤프 지수 s_Q에 대하여 $s_P < s_Q$인 포트폴리오 Q가 충분히 존재한다고 가정한다(가설).

본 논문에서는 위 가설로부터 최적 포트폴리오에 근사한 부최적 포트폴리오를 구성하는 방법을 제안한다. 먼저, 무작위 추출을 통해 2k개의 자산으로 이루어진 포트폴리오(2k-자산 포트폴리오)를 m개 구성한다. 생성된 m개의 2k-자산 포트폴리오 중 샤프지수가 높은 상위 r%의 2k-자산 포트폴리오를 선택한다.

선택된 각각의 2k-자산 포트폴리오에 대하여, 유전 알고리즘을 사용해 2k개 자산 중 k 개를 선택하여 k -자산 포트폴리오들을 구성한다. 가설에 따라, 2k-자산 포트폴리오보다 높은 샤프 지수를 갖는 k-자산 포트폴리오가 충분히 존재할 것이다. 이러한 k-자산 포트폴리오를 찾기 위해 유전 알고리즘을 사용한다. 유전 알고리즘의 모델은 다음과 같다.

$$F = \max S(P_k) \tag{4}$$

$$g = (c_0, c_1, \ldots, c_m) \tag{5}$$

$$\sum_{i=1}^{m} c_i = k \begin{cases} 1 & \text{if } \ i_{th}\, asset \in P_k \\ 0 & else \end{cases} \tag{6}$$

목적함수 F는 포트폴리오 P_k의 샤프지수 $S(P_k)$를 최대화한다. g는 염색체로, m개의 c_i로 이루어진 벡터이다. c_i는 이진 변수로써 자산 i가 포트폴리오 P_k에 속하면 1, 속하지 않으면 0의 값을 가지며 합은 k이다. 즉, 염색체 g는 m개 자산 집합 중 k개의 자산을 선택한 k-자산 균등배분 포트폴리오이다.

4. 실험

본 논문에서 사용된 개별 자산 데이터는 KOSPI 100개 자산의 2011년 1월 1일부터 2011년 12월 31일까지의 1일 단위 데이터를 사용하였다. 실험은 Intel i5-2500 3.3Ghz, 16GB RAM의 환경에서 Python 3.7을 사용해 진행하였다.

실험은 다음과 같이 진행되었다. 먼저 가설 검증을 위해 100개의 자산을 대상으로 몬테카를로 시뮬레이션을 통해 10-자산 포트폴리오 2,000개를 생성하고, 생성된 10-자산 포트폴리오의 샤프 지수를 계산하였다. 이 중 임의의 포트폴리오들을 선택하여, 포트폴리오를 구성하는 10개의 자산 중 5개 자산을 선택하는 모든 경우의 5-자산 포트폴리오에 대하여 샤프 지수를 계산하였다. 계산된 10-자산 포트폴리오의 샤프 지수와 5-자산 포트폴리오의 샤프 지수의 비교를 통해 가설을 검증하였다.

<표 1 > 10-자산 포트폴리오와 그로부터 생성된 5-자산 포트폴리오들의 샤프 지수 비교

번호	10-자산 SR	5-자산 SR (MAX)	5-자산 SR (MIN)	10-자산 백분위 (%)
1	5.1844	8.1621	1.4409	23.02%
2	1.3805	6.0862	-1.1711	39.29%
3	0.9899	3.6598	-1.5367	43.25%
4	-0.0811	4.1829	-3.6556	50.79%
5	-0.8551	0.6728	-2.8643	56.35%

<표 1>은 10-자산 포트폴리오의 샤프 지수와 10-자산 포트폴리오로부터 만들어진 5-자산 포트폴리오 샤프 지수의 최대/최소값을 비교한 표이다. 10-자산 백분위는 10-자산 포트폴리오의 샤프 지수가 5-자산 포트폴리오들 중 상위 몇%에 위치하는 지를 나타낸 것이다. 위 실험 결과를 통해 2k-자산 포트폴리오로부터 k-자산 포트폴리오를 구성할 때 더 높은 샤프 지수를 갖는 포트폴리오가 충분히 존재함을 확인하였다.

<표 2> (A)와 (B)의 샤프 지수 통계량 비교

포트폴리오	평균 SR	표준편차	Max SR	Min SR
A	1.6933	1.4592	9.4887	-5.1919
B	-0.1806	1.5303	7.6961	-8.1299

다음으로, 제시한 방법의 검증을 위하여 몬테카를로 시뮬레이션을 통해 만들어진 10-자산 포트폴리오 2000개 중 상위 100개로부터 유전 알고리즘을 사용해 5-자산 포트폴리오(A)를 25,200개 생성하였다. 세대의 수는 10세대로 제한하였다. 비교를 위해 100개 자산으로부터 곧바로 무작위 추출하여 5-자산 포트폴리오(B)를 25,200개 생성하였다. (그림 2)는 (A) 포트폴리오와 (B) 포트폴리오의 샤프 지수를 히스토그램으로 나타낸 것이며, 노란색 영역은 100개 자산으로부터 생성 가능한 모든 5-자산 포트폴리오, 약 7,529만 개의 샤프 지수 중 상위 1%를 나타낸다. <표 2>는 (A)와 (B)의 샤프 지수의 통계량을 나타낸 표이다. 실험 결과 (A)의 샤프 지수가 (B)에 비해 더 높은 값을 보였다.

<표 3> 5-자산 포트폴리오 샤프 지수 계산에 소요된 시간

	개수	소요 시간
전체	75,287,520	10시간 이상
A	25,200	20.74초
B	25,200	5.01초

<표 3>은 5-자산 포트폴리오의 샤프 지수를 계산하는데 소요된 시간을 나타낸다. 100개 자산으로부터 조합 가

능한 전체 5-자산 포트폴리오의 수는 75,287,520개이며, 실험에 사용된 컴퓨터의 메모리 크기의 제약으로 인해 샤프 지수 계산에 10시간 이상이 소요되었다. (A)와 (B)의 샤프 지수를 계산하는 데에는 각각 20.74초, 5.01초로 (A)가 (B)의 약 4배의 시간이 걸렸다.

<표 4> 전체 5-자산 포트폴리오 중 (A)와 (B)의 샤프 지수가 전체의 t% 이상인 포트폴리오의 수와 비율

t (%)	A		B	
	개수	비율	개수	비율
1%	3842	15.25%	279	1.11%
0.50%	2081	8.26%	133	0.53%
0.10%	483	1.92%	22	0.09%
0.05%	259	1.03%	13	0.05%

<표 4>는 100개 자산으로부터 조합 가능한 모든 5-자산 포트폴리오의 샤프 지수의 상위 t% 이상에 속하는 (A)와 (B)의 포트폴리오 수와 비율을 나타낸 것이다. 모든 t에 대하여 전체의 상위 0.05%이상에 속하는 (A)의 포트폴리오의 수가 (B)의 포트폴리오 수보다 약 20배 많았다. 또한 (A)의 1.03%에 해당하는 포트폴리오는 전체 포트폴리오의 상위 0.05% 이상에 속하여 제안한 방법을 통해 최적 포트폴리오에 근사하는 포트폴리오를 충분히 만들 수 있음을 보였다.

5. 결론

본 논문에서는 최적 포트폴리오에 충분히 근사하는 부최적 포트폴리오를 찾기 위하여 몬테카를로 시뮬레이션을 통해 탐색 공간을 축소하는 방법을 제안하였다. 실험 결과 본 논문에서 제안한 방법을 적용하는 것이 무작위 추출에 비해 소요되는 시간은 다소 늘었으나 최적에 근사하는 포트폴리오가 크게 많아지는 것을 확인하였다.

본 논문에서는 동일 가중치의 포트폴리오에 대하여 실험하였으나, 가중치가 동일하지 않은 최적 포트폴리오를 탐색하는 데에는 더 많은 시간이 걸릴 것이다. 따라서 향후 유전 알고리즘을 병렬화하고, 이를 사용해 자산 배분 가중치를 조절하여 빠르게 최적 포트폴리오로 근사할 수 있는 방법에 대하여 연구할 예정이다.

참고문헌

[1] Harry Markowitz, “Portfolio Selection:Efficient Diversification of Investment”, New York: John Wiley & Sons, 1959.

[2] Salvatore Arnone, Andrea Loraschi, Andrea Tettamanzi, "A Genetic Approach to Portfolio Selection", Neural Network World, 3, 6, (1993), 597-604

[3] Prisadarng Skolpadungket, Keshav Dahal, Napat Harnpornchai, "Portfolio Optimization using Multi-Objective Genetic Algorithms", 2007 IEEE Congress on Evolutionary Computation, (2007), 516-523

[4] Chang-Chun Lin, Yi-Ting Liu, "Genetic algorithms for portfolio selection problems with minimum transaction lots", European Journal of Operational Research, 185, (2008) 393 - 404

[5] Hamed Soleimani, Hamid R. Golmakani, Mohammad H. Salimi, "Markowitz-based portfolio selection with minimum transaction lots, cardinality constraints and regarding sector capitalization using genetic algorithm", Expert Systems with Applications, 36, (2009), 5058 - 5063

[6] José D. Bermúdez, José V. Segura, Enriqueta Vercher, "multi-objective genetic algorithm for cardinality constrained fuzzy portfolio selection", Fuzzy Sets and Systems, 188, 1, 16-26, (2012)

[7] Sharpe, William F, “The sharpe ratio”, Journal of portfolio management, 21, 1, 49-56, (1994)

[8] Berkemeier, Thomas, 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, 17, 12, (2017), 8021-8029,