

군집분석과 유전자 알고리즘을 활용한 투자자 거래정보 기반 포트폴리오 투자전략

정동현¹ · 오경주²

¹²연세대학교 정보산업공학과

접수 2013년 12월 2일, 수정 2013년 12월 26일, 게재확정 2014년 1월 6일

요약

본 연구에서는 투자자 거래 정보를 활용한 포트폴리오 투자전략을 제안했다. 포트폴리오를 구성하는 과정에서 군집분석을 활용하여 기대수익이 높은 종목을 선정하고, 유전자 알고리즘으로 포트폴리오를 최적화하여 투자성고를 높이고자 했다. 2007년 4월부터 2013년 6월까지의 국내 주식시장을 대상으로 한 실증분석을 통하여, 본 연구에서 제안한 포트폴리오 투자전략의 유용성과 우수성을 확인했다. 본 연구의 결과는 특정 투자 주체의 매매행태를 분석하여 투자 의사결정에 이용할 수 있으며, 이를 통하여 높은 투자성고를 얻을 수 있음을 보여준다. 또한 인공지능 기법이 투자 의사결정에 유용하게 사용될 수 있음을 시사한다.

주요용어: 군집분석, 유전자 알고리즘, 투자자 거래 정보, 포트폴리오 투자전략.

1. 서론

전 세계적으로 금융시장에 대한 규제가 완화되고 통합화가 촉진되면서, 내국인 기관투자자와 개인투자자 위주의 주식시장에서 외국인 투자자가 중요한 투자 주체로 등장하기 시작했다. 그리고 이러한 외국인투자자는 세계 각지의 주식시장에서 막대한 영향력을 미치게 되었다. 1992년 외국인투자자에게 개방된 국내 주식시장의 경우에도, 외국인투자자의 주식보유량과 거래 비중의 확대에 따라서 기관투자자, 개인투자자와 더불어 외국인투자자가 중요 투자 주체로 분류되어왔다.

일부 내국인투자자들은 이러한 외국인투자자의 투자분석기법이 더 선진적이기 때문에 높은 성과를 낼 것이라고 예상하며 이들의 투자행태와 투자성고에 관심을 가져왔다. 반면에, 내국인투자자가 외국인투자자보다 국내 정보의 획득에 유리하기 때문에 내국인투자자의 투자성고가 더 높을 것이라고 예상하는 투자자도 있었다. 이처럼 어떠한 투자 주체가 정보의 우위 혹은 정교한 투자분석기법을 바탕으로 주식시장에서 우수한 성과를 얻고 있는지는 주요한 관심사로 떠올랐고, 이를 밝혀내기 위한 연구가 지속적으로 진행되고 있다.

Grinblatt 등 (2000)은 핀란드 주식시장을 대상으로 한 연구에서 외국인투자자가 모멘텀 투자를 통하여 내국인투자자보다 더 높은 투자성고를 얻는다고 주장했다. Kamesaka 등 (2003) 역시 일본 주식시장에서 외국인투자자가 뛰어난 투자성고를 얻고 있음을 보여주었다. 반면에, 기관투자자가 우수한 투자성고를 얻고 있음을 보여주는 연구도 있다. Bae 등 (2006)은 일본주식시장을 대상으로 한 연구에서

¹ (120-749) 서울특별시 서대문구 신촌동 134번지, 연세대학교 정보산업공학과, 석사과정.

² 교신저자 : (120-749) 서울특별시 서대문구 신촌동 134번지, 연세대학교 정보산업공학과, 부교수.
E-mail: johanoh@yonsei.ac.kr

기관투자자가 종목선택에서 뛰어난 능력을 보이며 우수한 투자성적을 얻는다고 분석했다. 이에 반하여, 개인투자자는 주식시장에서 우수한 투자성적을 얻지 못한다. Kamesaka 등 (2003), Bae 등 (2006)은 개인투자자가 주식시장에서 좋지 못한 투자성적을 얻는다고 분석했다.

국내 주식시장을 대상으로 투자자 유형에 따른 투자행태와 투자성적을 분석한 연구도 있다. Choe 등 (1999)은 국내 주식시장에서 외국인 투자자의 투자성적이 뛰어남을 보여주었고, Ko와 Kim (2004), Kim과 Cheon (2004), Oh와 Kim (2013) 역시 외국인투자자의 투자 우수성과 시장 영향력을 보여주었다. 반면에 Yi와 Lee (2004)는 모멘텀 투자를 수행하는 기관투자자의 성과가 가장 뛰어나다고 주장했다.

이러한 기존의 연구들은 주식시장에서 투자자 유형에 따른 투자성적을 사후적으로 고찰하는데 그 의의를 두었다. 하지만, 본 연구에서는 주식시장에서 특정 투자 주체가 우수한 투자성적을 얻는다는 선행 연구 결과에 주목하여, 투자자 유형에 따른 거래 정보를 활용하여 포트폴리오를 구성하는 전략을 제안하고, 이러한 투자전략의 유용성에 관하여 분석했다. 기대수익이 높은 주식을 선택하여 포트폴리오를 구성하는 투자전략에 대한 연구는 많은 이들의 관심을 끌어들였다 (Hong 등, 2011; Kam과 Shin, 2012). 하지만, 투자자 유형에 따른 거래 정보를 활용하여 포트폴리오를 구성하는 투자전략에 대한 연구는 전무하다.

또한 본 연구는 포트폴리오를 구성하는 과정에서 인공지능 (artificial intelligence) 기법인 군집분석 (cluster analysis)과 유전자 알고리즘 (genetic algorithm)을 활용했다. 군집분석을 통하여 특정 투자 주체의 거래 비중이 높은 군집을 따로 분류하고, 군집에 속한 주식으로 포트폴리오를 구성했다. 그리고 유전자 알고리즘을 사용하여 종목별 투자 비중을 최적화함으로써, 포트폴리오의 투자성적을 높이고자 했다.

투자자 유형에 따른 투자성적을 분석한 다수의 기존 연구들이 2000년대 초반까지의 국내 주식시장을 연구기간으로 설정한 반면, 본 연구는 2007년 4월부터 2013년 6월까지의 국내 주식시장을 연구기간으로 설정했다. 2008년 금융위기 이후 최근까지의 금융시장 변화를 고려할 때, 최근 6년간의 연구 결과는 주식시장에 대한 새로운 시사점을 줄 수 있다고 하겠다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 투자자 유형에 따른 거래 정보를 활용하여 포트폴리오를 구성하는 방법에 대하여 설명하고, 3절에서는 실증분석 결과를 토대로 제안된 포트폴리오 투자 전략의 유용성을 분석했다. 마지막으로 결론에서는 본 연구의 시사점 및 향후 연구에 대하여 서술했다.

2. 연구방법

본 연구에서 제안하는 포트폴리오의 구성 및 재조정 방식은 크게 3단계로 구분할 수 있다. 1단계는 투자자 유형별 거래 비중을 이용하여 포트폴리오의 구성 종목을 선정하는 단계로서, 인공지능 기법인 군집분석을 활용하여 특정 투자 주체의 거래 비중이 높은 군집에 포함된 종목을 선택한다. 2단계는 포트폴리오를 최적화하는 단계로서, 유전자 알고리즘으로 포트폴리오 구성 종목의 투자 비중을 결정한다. 마지막 3단계는 포트폴리오를 재조정하며 지속적으로 성과를 측정하는 단계로서, 포트폴리오 재조정시 때마다 1단계와 2단계를 반복하여 포트폴리오를 재구성하며 성과를 측정한다.

2.1. 포트폴리오 구성 종목 선정

군집분석은 유사한 성격을 가진 데이터를 모아서 하나의 군집으로 분류하고, 다른 군집과 비교하는 분석기법으로, 본 연구에서는 가장 일반적으로 사용되는 k -평균 군집분석 (k -means clustering) 방법을 사용했다. k -평균 군집분석 방법은 비계층적 군집분석 방법론으로, 군집에 속한 개체와 군집의 중심과의 거리를 계산하고 거리의 합을 최소화하는 방식으로 개체를 분류한다.

k -평균 군집분석 방법을 활용하기 위해서는 먼저 군집의 분류 개수 k 를 결정해야 한다. 몇 개의 군집으로 데이터를 분류하느냐에 따라서 군집분석의 결과가 달라지기 때문에, k 의 선택은 매우 중요하다. 본 연구에서는 군집분석을 통하여 포트폴리오를 구성하고, 투자성과를 높이고자 하였다. 이를 위하여, 군집의 분류 개수 k 를 조절할 수 있는 k -평균 군집분석 방법을 활용하였다. 즉 군집의 분류 개수 k 를 조절하여, 기대수익을 높일 수 있는 최적의 k 값을 찾아 군집분석을 수행하고, 포트폴리오의 투자성과를 높이고자 하였다. 이러한 목적에 비추어, 학습구간을 설정하여 k 값에 변화를 주면서 군집분석을 수행했고, 최적의 k 값을 선택하여 실험구간에 적용했다.

본 연구에서는 투자자 유형별 거래 비중을 이용하여 군집을 나누고, 기대수익이 높은 군집을 선택하기 위하여 군집분석을 사용했다. 따라서 다음의 5가지 특성변수를 사용하여 군집분석을 수행했다.

$$\text{주가 등락률} = \frac{\text{현재주가} - \text{과거주가}}{\text{과거주가}}$$

$$\text{기관투자자 거래 비중} = \frac{\text{기관투자자 순 거래량}}{\text{총 거래량}}$$

$$\text{외국인투자자 거래 비중} = \frac{\text{외국인투자자 순 거래량}}{\text{총 거래량}}$$

$$\text{개인투자자 거래 비중} = \frac{\text{개인투자자 순 거래량}}{\text{총 거래량}}$$

$$\text{VR (Volume Ratio)} = \frac{\text{주가 상승일 거래량} + 0.5 * \text{주가 보합일 거래량}}{\text{주가 하락일 거래량} + 0.5 * \text{주가 보합일 거래량}}$$

주가 등락률은 일정기간 동안 움직인 주가의 등락폭을 의미한다. 그리고 투자자 유형별 거래 비중은 일정기간 동안의 전체 거래량 중에서 각각의 투자자 유형에 따른 순 거래량이 얼마나 되는지를 보여준다. 마지막 변수 VR은 기술적 분석에서 사용되는 거래량 지표로서 상승일의 거래량을 하락일의 거래량으로 나누어 계산하며, 주가 상승일의 거래량 비중을 보여준다.

위의 다섯 가지 변수를 Z 점수 (Z score)로 표준화하여 군집분석을 수행한다. 그리고 분류된 군집 중에서 기관투자자 또는 외국인투자자의 거래 비중이 높은 군집을 선택하여, 군집에 속한 종목들로 포트폴리오를 구성한다.

2.2. 포트폴리오 최적화

유전자 알고리즘은 John Holland에 의해 1975년에 소개된 인공지능 기법으로 유전적 계승과 다윈의 생존경쟁에 기초하여 개발되었다. 교배 (crossover), 돌연변이 (mutation)를 통해 생성된 개체군들 중 적합도 (fitness)가 가장 높은 개체들을 다음 세대의 개체군으로 선정하여 교배와 돌연변이를 지속적으로 반복해 나가며, 최적화 문제를 해결하는데 사용한다. Oh 등 (2005), Byun 등 (2009)이 유전자 알고리즘을 사용하여 포트폴리오를 최적화 할 수 있음을 보여주었다.

1단계에서 군집분석으로 기대수익이 높은 종목군을 선정하여 포트폴리오를 구성했다면, 2단계에서는 유전자 알고리즘을 활용하여 포트폴리오 구성 종목의 투자 비중을 최적화한다. 종목별 투자 비중은 Markowitz (1952)의 포트폴리오 이론에 따라서, 포트폴리오의 분산을 최소화 하는 비중으로 결정한다. 즉, 군집분석이 포트폴리오의 기대수익을 높이기 위한 목적으로 사용되었다면, 유전자 알고리즘은 포트폴리오의 위험을 낮추기 위하여 사용되었다.

w_i 를 i 종목의 투자 비중, w_j 를 j 종목의 투자 비중 그리고 σ_{ij} 를 i, j 종목의 공분산이라고 한다면,

유전자 알고리즘의 목적함수는 다음과 같다.

$$\text{Objective function} = \text{Min} \left(\sum_{i,j=1}^n w_i w_j \sigma_{ij} \right)$$

2.3. 포트폴리오 재조정 및 성과측정

마지막 단계에서는 슬라이딩 윈도우 방식으로 포트폴리오를 지속적으로 재조정 (rebalancing)하며 성과를 관측한다. Jang 등 (1993), Kohzadi 등 (1996) 그리고 Ahn 등 (2012) 역시 투자전략의 성과를 측정하기 위해, 슬라이딩 윈도우 방식을 사용했다. 본 연구에서 하나의 윈도우는 변수 수집 (data acquisition) 기간과 성과측정 (performance evaluation) 기간으로 나누어져 있다. 변수 수집이 끝나면 군집분석과 유전자 알고리즘을 활용하여 포트폴리오를 구성하고, 성과측정 기간 동안 포트폴리오의 성과를 관측한다. 성과를 관측하면서 동시에 포트폴리오 재조정에 사용할 변수를 수집하며, 성과측정이 끝나자마자 포트폴리오를 다시 구성하고 성과측정을 시작한다. 이러한 일련의 과정이 전체 실험기간 동안 지속되며, 본 연구에서 제안하는 포트폴리오 투자전략의 성과를 측정하게 된다.

3. 실증분석

실증분석에서 사용된 데이터는 코스콤에서 구한 2007년 4월부터 2013년 6월까지의 KOSPI200 데이터로서, 본 연구에서는 KOSPI200 상위 90종목을 선정하여 실증분석에 사용했다. 전체 기간 중에서 2007년 4월부터 2011년 6월까지의 4년 3개월을 학습구간, 2011년 7월부터 2013년 6월까지의 2년을 실험구간으로 분류했다. 그리고 학습구간에서의 실험으로 최적의 군집 분류 개수 k 를 선택하고, 이를 실험구간에 적용하여 본 연구에서 제안하는 포트폴리오 투자전략의 성과를 살펴보았다.

학습구간의 실험에서는 군집의 분류 개수 k 를 7에서 10까지 바꾸어 가며 실험했다. 그리고 분류된 k 개의 군집 중에서 기관투자자의 거래비중이 높은 상위 2개의 군집에 포함된 종목으로 포트폴리오를 구성하였고, 유전자 알고리즘을 활용하여 종목별 투자 비중을 최적화하였다. 또한 유전자 알고리즘을 활용하여 구성된 포트폴리오의 투자성과 비교를 위한 대조 집단으로, 종목별 투자 비중을 동일하게 할당한 포트폴리오와 시가총액 비중으로 할당된 포트폴리오를 각각 구성하였다. 즉 군집분석을 활용하여 포트폴리오에 편입할 종목들을 선정하고, 투자 비중을 달리하여 세 가지의 포트폴리오를 구성했다.

마지막으로 이렇게 구성된 각각의 포트폴리오를 변수 수집 기간과 성과측정 기간을 3개월, 1개월로 설정하여 슬라이딩 윈도우 방식으로 성과를 관측했다. 즉 3개월간의 주가와 거래량 정보를 바탕으로 포트폴리오를 구성하고, 이를 1개월 마다 재조정하며 포트폴리오 투자전략의 성과를 측정하고 비교했다.

3.1. 학습구간

Figure 3.1은 학습구간에서 기관투자자의 거래비중이 높은 종목으로 구성된 포트폴리오의 연평균 수익률 (annual return)을 보여준다. Equal weight, Cap weight는 각각 종목별 투자 비중을 동일하게 할당된 포트폴리오, 시가총액 비중으로 할당된 포트폴리오의 수익률을 나타내며, GA weight는 유전자 알고리즘을 활용하여 구성된 포트폴리오의 수익률을 나타낸다. 학습구간에서 측정된 KOSPI200 지수의 연평균 수익률은 약 6.1%다. 반면에 군집분석을 활용하여 선정한 종목으로 구성된 포트폴리오 중에서 동일비중 포트폴리오, 유전자 알고리즘 비중 포트폴리오의 연평균 수익률은 7.4%와 22.5% 사이에서 관측되었다. 이러한 결과는 기관투자자의 거래 비중이 높은 종목으로 포트폴리오를 구성하면, 시장 수익률보다 높은 수익률을 얻을 수 있음을 보여준다. 즉, 군집분석을 활용하여 기대수익이 높은 종목을 선정하기 위한 소기의 목적이 달성되었음을 시사한다.

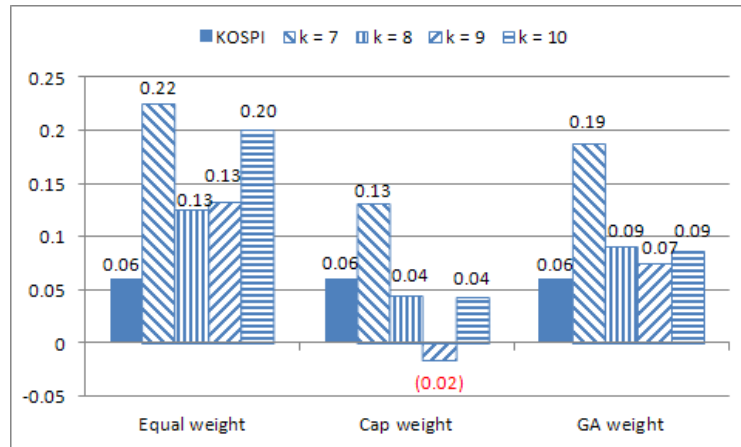


Figure 3.1 Annual returns during the training period (institutional investor)

Figure 3.2는 학습구간에서 외국인투자자의 거래비중이 높은 종목으로 구성된 포트폴리오의 연평균 수익률 (annual return)을 보여준다. 전반적으로 외국인투자자의 거래비중이 높은 종목으로 구성된 포트폴리오의 수익률 역시 KOSPI200 지수의 연평균 수익률보다 높다. 포트폴리오간 수익률을 비교하여 살펴보면, 유전자 알고리즘을 활용하여 종목별 투자 비중을 최적화한 포트폴리오의 수익률이 투자 비중을 동일하게 구성한 포트폴리오와 시가총액 비중으로 구성된 포트폴리오의 수익률보다 높은 것을 확인할 수 있다. 군집의 분류 개수 k 의 변화에 따라서 포트폴리오의 수익률이 달라지지만, 전반적으로 유전자 알고리즘을 활용하여 구성된 포트폴리오의 수익률이 더 높다.

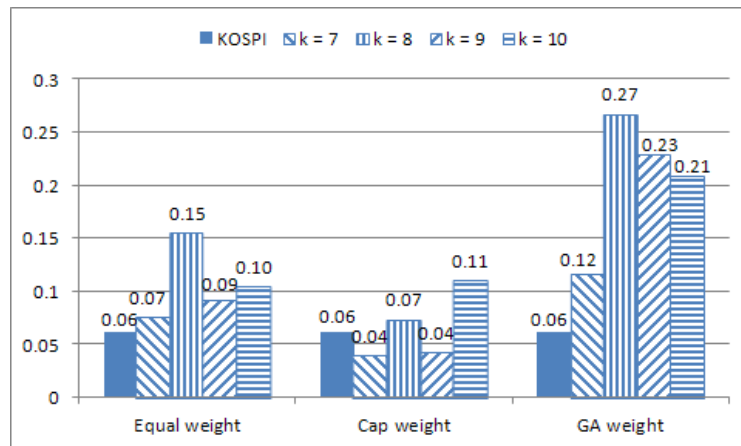


Figure 3.2 Annual returns during the training period (foreign investor)

Figure 3.3과 Figure 3.4는 학습구간에서 군집분석과 유전자 알고리즘을 단계적으로 활용하여 구성된 포트폴리오의 누적 수익률 변화 추이를 보여준다. Figure 3.3은 기관투자자의 거래비중이 높은 포트폴리오, Figure 3.4는 외국인투자자의 거래비중이 높은 포트폴리오의 수익률이다. 군집의 분류 개수 k 값에 관계없이 포트폴리오의 수익률은 KOSPI200 지수와 유사하게 오르내린다. 하지만 최종 누적 수익률은 KOSPI200 지수를 매수 후 보유 (buy and hold)한 전략보다 높은 것을 살펴볼 수 있다. 이러한 결과는 기관투자자의 거래 비중이 높은 종목들의 주가 상승률이 KOSPI200 지수의 상승률보다 높다는 것을

시사한다. 그리고 미국발 금융위기로 인하여, 전 세계 경제가 침체를 거듭하던 2008년의 움직임을 살펴 보면, 외국인투자자의 거래비중 높은 종목으로 구성된 포트폴리오 투자전략의 낙폭이 KOSPI200 지수의 낙폭보다 대체적으로 작은 것을 확인할 수 있다. 이는 유전자 알고리즘으로 포트폴리오의 분산을 최소화하여, 위험을 감소시키려는 시도가 성공적이었음을 보여준다.

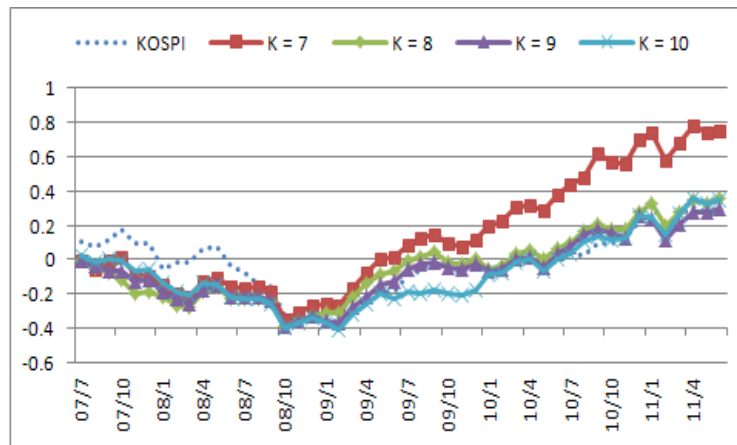


Figure 3.3 Cumulative returns during training period (institutional investor)

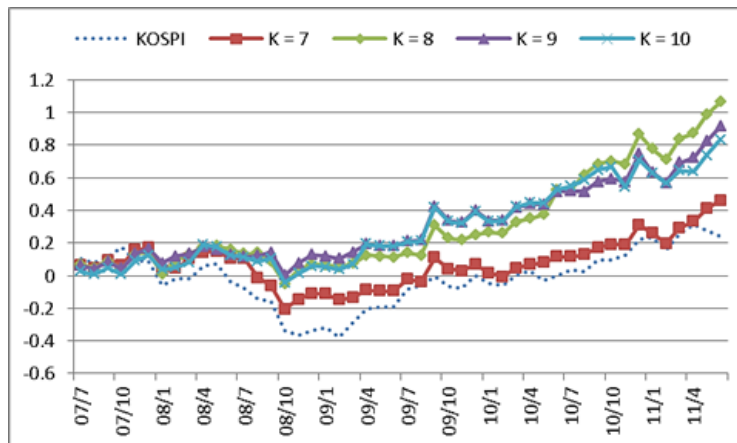


Figure 3.4 Cumulative returns during the training period (foreign investor)

샤프지수 (Sharpe ratio)는 투자안의 성과를 측정하기 위한 지표로써 (Sharpe, 1994), 기본적인 위험 보상비율의 하나다. 포트폴리오의 수익률을 R_p , 무위험 수익률을 R_f , 포트폴리오의 표준편차를 σ_p 로 표기하면, 다음과 같이 계산할 수 있다.

$$\text{Sharpe ratio} = \frac{R_p - R_f}{\sigma_p}$$

샤프지수의 크기는 단위 위험을 부담하는 대가로 얻은 수익의 크기이며, 포트폴리오를 보유함으로써 부담한 위험 프리미엄 (risk premium)을 의미한다. 이러한 샤프지수의 비교를 통하여 위험을 반영한 포트폴리오의 투자성적을 살펴볼 수 있다.

Figure 3.5와 Figure 3.6은 각각 학습구간에서 기관투자자, 외국인투자자의 거래비중이 높은 종목으로 구성된 포트폴리오의 샤프지수 값을 보여준다. 샤프지수는 CD 91일률을 무위험 수익률의 대용치로 사용하여 계산했다. KOSPI200 지수의 샤프지수가 약 0.097인 반면, 전반적으로 동일비중 포트폴리오, 유전자 알고리즘 비중 포트폴리오의 샤프지수는 이보다 높다. 그리고 Figure 3.6에서 포트폴리오간 샤프지수를 비교하여 살펴보면, 군집분석과 유전자 알고리즘을 단계적으로 활용하여 외국인투자자의 거래비중이 높은 종목으로 구성된 포트폴리오가 수익률뿐만 아니라 샤프지수도 우수함을 확인할 수 있다.

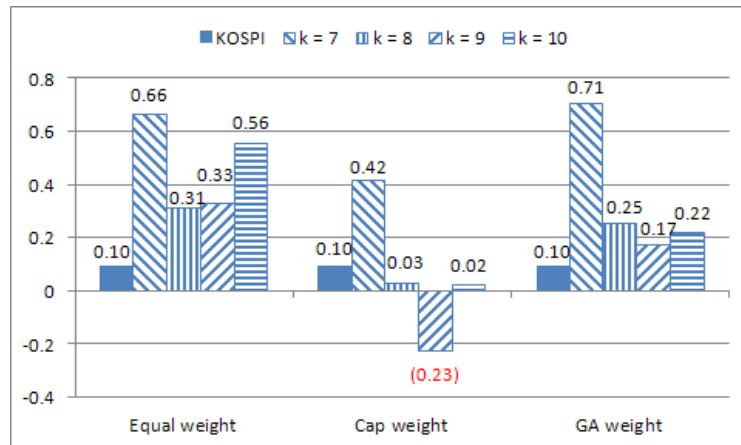


Figure 3.5 Sharpe ratios during the training period (institutional investor)

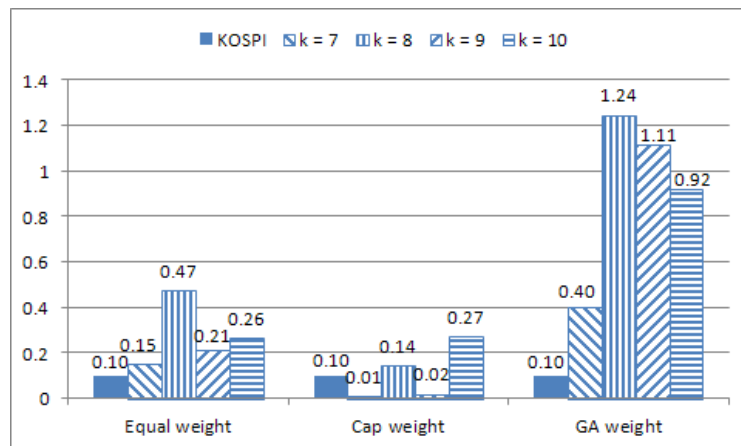


Figure 3.6 Sharpe ratios during the training period (foreign investor)

3.2. 실험구간

지금까지 학습구간의 실험을 통하여, 본 연구에서 제안한 포트폴리오 투자전략의 우수성을 살펴보았다. 다음으로는 실험구간에서 최적의 군집 분류 개수 k 를 선택하여 포트폴리오를 구성하고 투자성과를 검증했다. k 값이 변화하면 군집분석을 활용하여 선택하는 종목이 달라지고, 이에 따라서 포트폴리오의

투자성과가 달라진다. 그리고 Figure 3.3, Figure 3.4에서 확인할 수 있듯이, k 값에 따른 투자성과는 학습구간의 설정에 따라 달라질 수밖에 없다. 그러므로 학습구간에서 수익률과 샤프지수가 가장 높은 k 값을 최적의 k 값으로 선택하기는 어렵다. 따라서 본 연구에서는, 학습구간에서의 수익률과 샤프지수가 평균에 가장 근접한 값을 최적의 k 값으로 선택했다. 위의 학습구간 실험결과를 통하여, 학습구간에서 수익률과 샤프지수가 평균에 근접한 k 값은 9임을 알 수 있다. 즉, 군집 분류 개수를 9로 적용하여 실험구간에서 포트폴리오 전략의 투자성과를 측정했다.

Table 3.1은 실험구간에서 기관투자자의 거래비중이 높은 종목으로 구성된 포트폴리오의 투자성과를, Table 3.2는 실험구간에서 외국인투자자의 거래비중이 높은 종목으로 구성된 포트폴리오의 투자성과를 보여준다. 그리고 Figure 3.7은 실험구간에서 군집분석과 유전자 알고리즘을 단계적으로 활용하여 구성된 포트폴리오의 누적 수익률 변화 추이를 보여준다. 실험구간에서 KOSPI200 지수는 박스권에서 10% 내외의 등락을 거듭하며, 연평균 약 6.0%씩 하락했다. 반면에 군집 분류 개수를 9로 적용하여 구성된 기관과 외국인 거래비중이 높은 포트폴리오는 각각 연평균 3.2%, 5.3%의 수익률을 기록했다. 그리고 샤프지수는 약 -0.002와 0.128로 KOSPI200의 샤프지수보다 높게 관측되었다.

학습구간에서 관찰한 투자성과와 마찬가지로, 실험구간에서도 본 연구에서 제안한 포트폴리오 투자전략의 성과는 매우 뛰어났다. 따라서 기관 및 외국인투자자의 거래 비중이 높은 종목으로 포트폴리오를 구성하고 분산을 최소화하여 투자하는 포트폴리오 전략은 충분히 유용한 투자전략이며, 높은 투자성과를 보장한다는 것이 검증되었다. 또한 실험결과에 비추어보면, 기관투자자보다 외국인투자자의 거래 비중이 높은 종목으로 포트폴리오를 구성하는 것이 조금 더 나은 투자성과를 보이는 것을 확인할 수 있었다.

Table 3.1 Results during the testing period (institutional investor)

	KOSPI200	$k = 9$
Annual return	-0.060	0.032
Sharpe ratio	-0.552	-0.002

Table 3.2 Results during the testing period (foreign investor)

	KOSPI200	$k = 9$
Annual return	-0.060	0.053
Sharpe ratio	-0.552	0.128

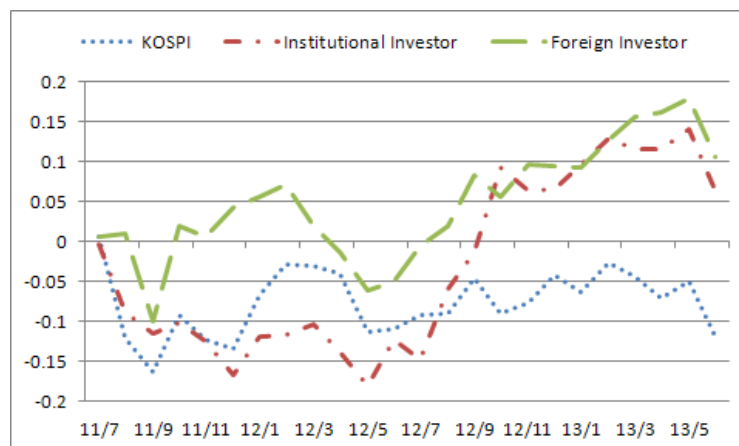


Figure 3.7 Cumulative returns during the testing period

4. 결론 및 향후 연구

주식시장에서 기대수익이 높은 종목을 선정하여 포트폴리오를 구성하려는 투자전략에 대한 연구는 지속적으로 진행되어 왔다. 일반적으로 포트폴리오를 구성하는 종목을 선정하기 위해서는 주로 개별기업의 재무적 요인을 분석하는 기본적 분석을 사용한다. 하지만 본 연구에서는, 기관 및 외국인 투자자가 주식시장에서 우수한 투자성과를 얻는다는 기존의 연구에 주목하여, 개별 주식의 거래 주체를 분석하여 포트폴리오를 구성하는 전략을 제안했다.

본 연구에서 제안한 포트폴리오 투자전략은 다음과 같다. 투자자 유형에 따른 거래 정보를 활용하여 군집분석을 수행하고, 기관투자자 또는 외국인투자자의 거래 비중이 높은 군집에 포함된 종목으로 포트폴리오를 구성한다. 그리고 유전자 알고리즘으로 종목별 투자 비중을 최적화하여, 포트폴리오의 분산을 최소화한다.

학습구간의 실험을 통하여, 본 연구에서 제안한 투자전략의 수익률이 시장 수익률을 상회함을 보여주었다. 그리고 포트폴리오의 분산을 최소화하면, 안정적인 투자성과를 얻을 수 있음을 확인했다. 투자전략의 검증을 위하여 설정한 실험구간에서도, 포트폴리오 투자전략은 뛰어난 성과를 기록했다. 이러한 결과는 특정 투자 주체의 매매행태를 분석하여 투자 의사결정에 이용할 수 있으며, 이를 통하여 높은 투자성과를 얻을 수 있음을 보여준다. 또한 인공지능 기법이 투자 의사결정에 유용하게 사용될 수 있음을 시사한다.

본 연구에서 진행된 실험은 포트폴리오 구성에 있어서 KOSPI200의 90종목을 대상으로 하였다. 향후 KOSPI200을 전체종목을 대상으로 한 연구는 추후 연구로 남기겠다. 또한 포트폴리오의 구성 및 재조정시기를 다변화하여 수익률의 변화 양상을 살펴보도록 하겠다.

References

- Ahn, J. J., Kim, D. H., Oh, K. J. and Kim, T. Y. (2012). Applying option Greeks to directional forecasting of implied volatility in the options market: An intelligent approach. *Expert Systems with Applications*, **39**, 9315-9322.
- Bae, K., Yamada, T. and Ito, K. (2006). How do individual, institutional, and foreign investors win and lose in equity trades? Evidence from Japan. *International Review of Finance*, **6**, 129-155.
- Byun, H. W., Song, C. W., Han, S. K., Lee, T. K. and Oh, K. J. (2009). Using genetic algorithms to develop volatility index-assisted hierarchical portfolio optimization. *Journal of the Korean Data & Information Science Society*, **20**, 1049-1060.
- Choe, H., Kho, B. and Stulz, R. (1999). Do foreign investors destabilize stock markets? The Korean experience in 1997. *Journal of Financial Economics*, **54**, 227-264.
- Grinblatt, M. and Keloharju, M. (2000). The investment behavior and performance of various investor types: A study of Finland's unique data set. *Journal of Financial Economics*, **55**, 43-67.
- Hong, D. H., Hwang, J. H. and Kyung, S. J. (2011). Combined value investment strategy performance. *The Korean Journal of Financial Engineering*, **10**, 59-81.
- Jang, G. S., Lai, F., Jiang, B. W. and Parnig, T. M. (1993). Intelligent stock trading system with price trend prediction and reversal recognition using dual-module neural networks. *Journal of Applied Intelligence*, **3**, 225-248.
- Kamasaka, A., Nofsinger, J. R. and Kawakita, H. (2003). Investment patterns and performance of investor groups in Japan. *Pacific-Basin Finance Journal*, **11**, 1-22.
- Kam, H. and Shin, Y. (2012). The profitability of portfolio investment based on the past trading volume and returns. *Korea Corporation Management Review*, **19**, 243-259.
- Kim, D. S. and Cheon, Y. (2004). Foreign investors vs. domestic investors, Who are better informed investors? *Korean Journal of Financial Studies*, **33**, 1-44.
- Ko, K. and Kim, K. (2004). Portfolio performance and characteristics of each investor type : Individuals, institutions, and foreigners. *Korean Journal of Financial Studies*, **33**, 35-62.

- Kohzadi, N., Boyd, M. S., Kermanshahi, B. and Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, **10**, 169-181.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, **7**, 77-91.
- Oh, K. J., Kim, T. Y. and Min, S. (2005). Using genetic algorithm to support portfolio optimization for index fund management. *Expert Systems with Applications*, **28**, 371-379.
- Oh, K. J. and Kim, Y. M. (2013). An intelligent early warning system for forecasting abnormal investment trends of foreign investors. *Journal of the Korean Data & Information Science Society*, **24**, 223-233.
- Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management*, **21**, 49-58.
- Yi, K. Y. and Lee, Y. G. (2004). The differences in investment behavior and performance by investor types. *Korea Industrial Economics Association*, **17**, 1233-1253.

Using cluster analysis and genetic algorithm to develop portfolio investment strategy based on investor information

Donghyun Cheong¹ · Kyong Joo Oh²

^{1,2}Department of Information and Industrial Engineering, Yonsei University

Received 2 December 2013, revised 26 December 2013, accepted 6 January 2014

Abstract

The main purpose of this study is to propose a portfolio investment strategy based on investor types information. For improvement of investment performance, artificial intelligence techniques are used to construct a portfolio. Among many artificial intelligence techniques, cluster analysis is applied to select securities and genetic algorithm is applied to assign the respective weight within the portfolio. Empirical experiments in the Korean stock market show that proposed portfolio investment strategy is practicable and superior strategy. This result implies that analysis of investor's trading behavior may assist investors to make an investment decision and to get superior performance.

Keywords: Cluster analysis, genetic algorithm, investor types, portfolio investment.

¹ Graduate student, Department of Information and Industrial Engineering, Yonsei University, Seoul 120-749, Korea.

² Corresponding author: Associate professor, Department of Information and Industrial Engineering, Yonsei University, Seoul 120-749, Korea. E-mail: johanoh@yonsei.ac.kr