

주식시장 네트워크에서 클러스터링 기법을 이용한 포트폴리오 구성 방법

천봉환, 김은경, 정인준, 우균
부산대학교 컴퓨터공학과
e-mail: cot2000@pusan.ac.kr

A Method for Portfolio Construction Using a Clustering Technique on the Stock Market Networks

Bong-Hwan Chun, Eun-Kyung Kim, In-Jun Jung, Gyun Woo
Dept of Computer Engineering, Pusan National University

요 약

본 논문은 주식 투자 포트폴리오를 구성하기 위해 클러스터링 기법을 이용하는 방법을 제안한다. 클러스터링 기법은 패턴 공간 상의 특징 벡터로 표현된 패턴 데이터를 몇 개의 부분집합으로 나누는 작업을 의미한다. 본 연구에서는 주식시장 네트워크에 클러스터링 기법을 적용하여 안정성과 수익률이 높은 포트폴리오를 구성하는 방법을 제안한다. 그리고 추천 클러스터의 투자 적합여부를 데이터를 통해 확인한다. 2007년 주식 데이터를 대상으로 실험한 결과, 추천 클러스터의 수익률이 전체 수익률을 상회함을 확인할 수 있었다.

1. 서론

주식 투자에 있어 포트폴리오를 구성하는 일은 투자자에게 매우 중요한 일이다. 주식 포트폴리오란 자신이 투명한 투자 영역 및 상품에 대한 것을 정리한 것이다. 포트폴리오는 시장 상황의 위험 요인을 줄이고 높은 수익을 얻는 것이 목표이다. 포트폴리오를 어떻게 구성하느냐에 따라 주식 시장이 안 좋을 때 손실을 줄이고 주식 시장이 좋을 때 수익을 극대화할 수 있다.

본 연구에서는 주식시장 네트워크에 클러스터링 기법을 적용하여 안정성과 수익률이 높은 포트폴리오를 구성하는 방법을 제안하고 그 결과를 바탕으로 주식 포트폴리오를 추천하는 시스템을 구현한다. 본 논문에서 구현된 시스템은 인터넷으로 주가 정보를 수집하여 데이터베이스를 만들고 그 데이터를 바탕으로 일정 기간 동안 주식시장 네트워크를 클러스터링한다. 이후 각 클러스터의 수익률과 클러스터 내 기업 개수를 토대로 추천 포트폴리오를 구성한다.

2. 관련 연구 - 주식시장 네트워크 분석 방법

주식시장 네트워크를 분석하는 방법은 각 기업의 내재 가치를 분석하기 보다는 기업간의 네트워크를 토대로 주가의 변동성을 계산하는 기술적인 분석 방법이다. 이는 주식시장을 네트워크로 간주하고 복잡계 연구를 응용함으로써 복잡계 이론과 모형을 금융시장에 적용하려는 연구다. 이 분야의 연구는 시계열 분석 분야, 행위자 기반 모형화 분야, 복잡 네트워크 분야의 3가지분야로 나누어 설명할

수 있다[1].

먼저 시계열 분석의 예를 보면, 두터운 꼬리(fat tail) 현상의 발견과 변동성 군집(volatility clustering) 현상, 다중 프랙탈 분석(multifractal analysis), 거품(bubble)의 로그 주기성(log-periodicity) 연구가 있다. 두터운 꼬리 현상이란 금융상품 수익률 분포에서 power-law distribution을 발견한 것을 말하며 이것이 빈번한 폭등, 폭락 현상의 배경이 된다고 분석하고 있다. 변동성 군집을 측정하는 연구로 장기 상관관계(long range correlation)를 측정하고 분석하는 것이 있다. 또한 위험을 예측하기 위해 변동성 군집을 모형화하는 연구, GARCH(Generalized Autoregressive Conditional Heteroscedasticity)와 같은 모형을 적용하려고 하는 연구등이 있다[2, 3, 4]. 다중 프랙탈 분석 연구는 금융시장 가격 변동의 예측을 시도하는 차트 분석가의 기술적 분석과 관련이 있는 것으로 자기유사성의 정도를 프랙탈 차원으로 측정하려는 시도에서 확장된 연구이다. 마지막으로 거품의 로그주기성이란 시간이 진행될수록 진동주기가 짧아지며 진동하는 현상을 말하며 극단적인 변화가 있기 전에는 로그주기성의 전조현상이 대부분 관찰된다고 한다.

두 번째로 행위자 기반 모형화(agent-based model:ABM) 분야가 있다. 행위자 기반 모형이란 시스템을 구성하는 행위자들의 미시적인 상호작용을 시뮬레이션 기법으로 처리하여 거시적인 변동을 설명하는 것을 말한다. 여기서 행위자는 경제활동의 주체가 되는 투자자, 기업, 국가를 말하며 전략은 행위자의 경제활동 규칙을 말한다. 행위자 기반 모델 분야의 대표적인 연구로는 산타페 연구소 인공

주식시장과 양의 되먹임 연구가 있다[5, 6].

마지막으로 복잡 네트워크 분야가 있다. 복잡 네트워크는 자연현상과 사회현상 모두에서 흔히 관찰된다. 이는 구성원 사이의 협동현상을 통해 끊임없이 변화하며 더욱 복잡한 구조로 진화해 나가는 것을 말한다. 복잡 네트워크는 척도없는 네트워크 구조를 가지며 매우 불균일한 분포를 가진다. 허브노드가 존재하지만 연결 예지가 적은 노드도 다수가 존재한다. 복잡 네트워크는 무작위적인 오동작이나 고장에는 매우 강하지만 허브에 대한 계획적인 공격에는 취약한 단점을 가지고 있다.

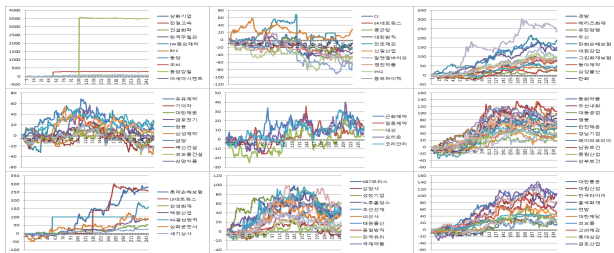
복잡계 네트워크에서는 미국의 주식시장 변화가 국내 주가에 영향을 미치는지 여부에 대한 연구들이 있다. 이 연구들은 주로 성숙한 시장이 신흥 시장의 가격변화에 얼마나 영향을 미치는지에 대해 연구하고 있다[7, 8, 9].

주식시장에서 금융 데이터와 같은 시계열 자료로부터 정보흐름을 관찰하고자 할 때 정보흐름의 검증과정을 효과적으로 수행하기 위하여 최소 비용 신장 트리(minimum spanning tree: MST)방법에 의하여 생성되는 주식 네트워크 접근법을 사용하고 있다[10, 11, 12, 13]. MST 방법에서는 의존성 정도를 거리 정보로 취급한다.

3. 클러스터링 기법

3.1 클러스터링 전략

주식시장 네트워크에 클러스터링 기법을 적용하면 주가의 패턴이 비슷한 기업들이 각각의 클러스터에 모이게 된다. 이를 검증하기 위해 2007년 1월 2일 부터 12월 28일까지 약 1년 동안 약 120개의 기업의 주가정보를 클러스터링하여 보았다. 각 클러스터별 기업들의 주가 변화를 그래프로 나타내면 그림 1과 같다.



(그림 1) 클러스터별 기업의 주가 변화 그래프: 총 10개의 클러스터

그림 1을 보면 각 클러스터에 속한 기업들의 주가 변화가 아주 유사한 것을 확인할 수 있다. 본 논문에서는 동일한 클러스터에 속한 기업은 동일한 성향을 가진 것으로 간주한다. 그래서 가장 안정적인 수익률을 얻는 클러스터를 분석하고 그 클러스터를 중심으로 포트폴리오를 추천한다.

3.2 추천 클러스터 선정

이 절에서는 안정적인 수익률을 가지는 클러스터를 선

정하는 방법에 대해 설명한다. 주식시장의 행태를 분석하는 연구는 여러 가지가 있는데 그 중에서도 주식시장 폭락의 전조에 대해 분석한 연구에 초점을 맞추었다. 이 연구에서는 주식시장의 폭락에 전조가 있을 것이라는 가정하에 이를 찾고자 했고 그들은 주식시장 패닉의 눈이 점점 모여서 눈사태가 되는 현상과 통한다고 보았다.

이 중 주식시장 폭락의 가장 큰 전조는 동조화 현상이었다. 2008년까지 주식시장의 동조화 수준이 극단적으로 높아졌고 결국 글로벌 금융위기가 왔다. 동조화는 주식시장에서 많은 투자자들이 더는 독립적으로 자신의 투자의사를 결정하지 않고 남들을 따라하기 시작했으며 그 양상이 극단적으로 높아지면 결국 주식시장에 패닉이 나타나 적절한 조건이 형성된다는 것이다.

동조화 현상에 근거하여 우리는 각각의 클러스터에 속한 기업의 수에 관심을 가졌다. 동조화현상에 의해 투자자들이 비슷한 패턴의 투자를 한다면 그에 해당하는 기업 또한 비슷한 주가 변화를 가질 것이라고 가정하였다. 동조화 현상을 가정할 때 같은 클러스터에 속한 기업의 수가 아주 많으면 그 기업의 수익률이 떨어질 것이라고 예상할 수 있다.

이 가정을 검증하기 위하여 실제 주식 데이터를 이용하여 수익률을 살펴보았다. 2005년도 실제 주식 데이터를 기준으로 각각 클러스터의 수익률과 개수를 비교하여 보았다. 투자 시점 이전 6개월 동안의 주식 데이터를 클러스터링 알고리즘을 통해 10개의 클러스터로 분류하고 투자 시점 이후 6개월간의 수익률을 살펴보았다. 이 때 각 클러스터의 개수도 함께 살펴보았다.

<표 1> 클러스터별 수익률과 기업 개수

| 클러스터 번호 | 이전 수익률 | 이후 수익률 | 기업 수 |
|---------|--------|--------|------|
| 1 | 139.14 | 124.00 | 18 |
| 2 | 112.77 | 125.64 | 21 |
| 3 | 102.93 | 111.97 | 3 |
| 4 | 130.44 | 128.72 | 12 |
| 5 | 157.66 | 138.02 | 15 |
| 6 | 134.64 | 158.41 | 3 |
| 7 | 146.91 | 120.14 | 6 |
| 8 | 145.34 | 123.92 | 15 |
| 9 | 84.10 | 90.41 | 9 |
| 10 | 152.37 | 118.79 | 16 |

수익률은 표 1과 같다. 표 1은 KOSPI 2005년 2월부터 7월까지의 자료를 통해 클러스터링하고 각 클러스터의 기업에 2005년 8월부터 6개월간 투자했을 때 얻는 수익률을 계산한 표이다. 이런 방식으로 2004년, 2006년 또한 수익률 계산을 하였다. 그리고 2005년 주식 데이터를 기반으로 각 클러스터에 속한 기업의 개수에 대한 수익률을 살펴보았다. 그 결과는 표 2와 같다.

표 2에서 기업의 수가 아주 적은 1행을 살펴보면 수익률이 안정적이지 않고 상승과 하락의 차이가 아주 큰 것을 볼 수 있다. 또한 기업의 수가 많은 클러스터의 경우는 뛰어난 수익률을 보이지 못했다. 수익률에 강세를 보인것

은 기업의 수가 가장 적은 것에서 한 단계 높은 2행과 기업의 수가 가장 많은 클러스터에서 한 단계 낮은 9행이었다. 이를 토대로 우리는 클러스터에 속한 기업의 개수가 2번째로 적은 것과 2번째로 많은 클러스터를 추천 클러스터로 선정하였다.

<표 2> 클러스터에 속한 기업의 개수에 대한 수익률

| 기업 수 순위 (오름차순) | 수익률 순위 분포 (단위 : 개) | | | | | | | | | |
|-------------------|--------------------|----|----|----|----|----|----|----|----|-----|
| | 1등 | 2등 | 3등 | 4등 | 5등 | 6등 | 7등 | 8등 | 9등 | 10등 |
| 1 | 3 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 3 | 2 | 0 | 2 | 1 | 0 | 2 | 1 | 1 |
| 3 | 2 | 1 | 1 | 2 | 0 | 2 | 1 | 1 | 2 | 0 |
| 4 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 0 | 0 | 2 |
| 5 | 1 | 1 | 1 | 0 | 2 | 3 | 0 | 1 | 1 | 1 |
| 6 | 0 | 2 | 1 | 0 | 0 | 2 | 0 | 1 | 2 | 2 |
| 7 | 2 | 1 | 1 | 2 | 1 | 0 | 1 | 0 | 1 | 2 |
| 8 | 0 | 0 | 3 | 1 | 1 | 1 | 2 | 2 | 0 | 0 |
| 9 | 0 | 2 | 1 | 4 | 2 | 0 | 2 | 0 | 0 | 1 |
| 10 | 0 | 1 | 0 | 1 | 1 | 1 | 4 | 0 | 0 | 0 |

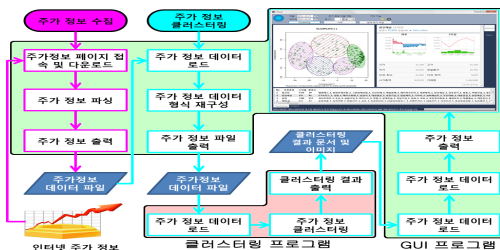
4. 구현

4.1 PSS의 설계

PSS의 목표는 본 논문에서 제시한 클러스터링 기법을 이용하여 최적의 투자를 위한 종목을 추천하여 사용자가 보기 편하게 보여주는 것이다. 이러한 목표를 만족시키기 위하여 시스템을 클러스터링 연산을 수행하는 클러스터링 프로그램과 GUI 프로그램으로 나누어 구현하였다.

클러스터링 프로그램은 주가 데이터를 입력받아 클러스터링 연산을 수행하고 결과를 출력하는 기능을 한다. 이 프로그램은 R 프로그래밍 언어[14]를 이용하여 구현하였다. 클러스터링 프로그램은 CSV형식의 주가 정보 파일을 입력 받는다. 이 파일에는 일정 기간 동안의 기업들의 주가 정보와 기업명이 저장된다. 클러스터링 프로그램은 주가 정보 클러스터링 연산을 수행한 이후 클러스터링 결과를 텍스트 파일과 이미지 파일로 출력한다.

GUI 프로그램은 주가 정보 수집 및 파싱, 클러스터링 결과를 입력받아 화면에 출력하여 주는 기능을 한다. GUI 프로그램은 입력된 주가 정보를 클러스터링 프로그램의 입력파일 형식에 맞추어 가공하여 출력한다. 그리고 클러스터링 프로그램이 수행된 다음 클러스터링 프로그램의 수행 결과를 이용하여 화면에 표시한다.



(그림 3) PSS의 구동 과정

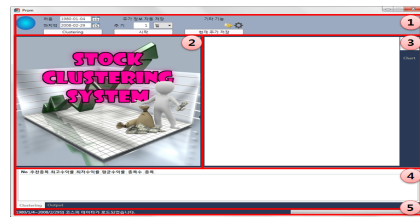
PSS의 구동 과정은 그림 3과 같다. 주가 정보 클러스터링을 수행하기 위해 시스템이 시작되면 GUI프로그램이 주가 정보의 데이터 파일을 불러온다. 시스템이 불러오는

주가 정보 데이터 파일은 GUI 프로그램이 인터넷으로는 일정한 주기 혹은 즉석으로 주가 정보를 받아와 저장한 데이터들로 이루어진 파일이다. GUI 프로그램은 사용자가 원하는 기간의 데이터를 클러스터링 프로그램의 입력에 맞는 형식으로 바꾸어 파일로 출력한다. 클러스터링 프로그램은 GUI 프로그램이 출력한 파일을 이용하여 클러스터링 연산을 수행한 뒤 클러스터링 결과를 이미지 파일과 텍스트 파일로 출력하게 된다. GUI 프로그램은 다시 그 파일을 입력받아 분석한 뒤 정보를 화면에 출력하여 준다.

4.2 PSS의 구현

4.2.1 GUI 프로그램의 사용자 환경

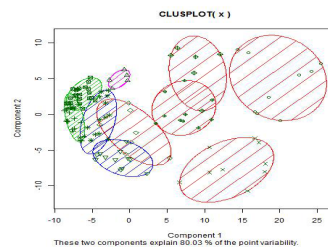
주식 클러스터링 시스템의 GUI는 그림 4와 같이 구성하였다. GUI 프로그램의 화면은 그림 4에서와 같이 크게 5개의 부분으로 나누어진다. 1번 부분은 명령 입력 부분이다. 그리고 2번 부분은 클러스터링 결과 이미지를 출력하는 부분이다. 3번 부분에는 주식 상세 정보 출력되며, 4번 부분에서는 클러스터링 결과가 출력된다. 마지막으로 5번 부분은 상태표시줄이다. 상태표시줄에서는 프로그램의 현재 수행 상태가 표시된다. 명령 입력 부분에서 대부분의 명령이 입력된다.



(그림 4) GUI 프로그램의 초기 화면

4.2.2 클러스터링

GUI 프로그램에서 출력된 주가 정보 파일을 이용하여 클러스터링 프로그램은 주가 정보 클러스터링 연산을 수행한다. 클러스터링은 Hartigan-Wong 알고리즘을 사용하였다.



(그림 4) 클러스터링 결과 이미지

클러스터링 프로그램은 클러스터링 결과를 그림 4와 같은 형태로 출력한다. 그림 4에서 각 점은 종목을 의미하며 원은 클러스터를 의미한다.

5. 주식 포트폴리오 추천

5.1 주식 포트폴리오 구성

이 절에서는 클러스터링 알고리즘을 이용하여 구현한 시스템을 이용하여 포트폴리오를 구성한다. 포트폴리오는 투자 시점 이전의 데이터를 이용하여 구성하며 그 기간은 6개월로 한다. 주식 투자에서 단기 투자보다 6개월 이상의 장기 투자가 더 안정성있는 수익률을 낸다는 연구가 있다. 투자 시점당 2개의 그룹이 추천되며, 추천 클러스터는 각 클러스터의 주식 개수에 따라 선택된다.

<표 3> 시물레이션을 통한 클러스터별 수익률 측정

| 투자시점 | 추천 종목 수익률 (%) | | 포트폴리오 수익률 (%) | 전체 수익률 (%) |
|------------|---------------|--------|------------------|---------------|
| | 클러스터 1 | 클러스터 2 | | |
| 2000-07-01 | 75.58 | 97.85 | 87.10 | 74.66 |
| 2000-09-01 | 92.18 | 92.03 | 92.07 | 89.69 |
| 2000-11-01 | 107.64 | 107.86 | 107.72 | 105.45 |
| 2001-01-01 | 121.37 | 118.62 | 119.15 | 120.52 |
| 2001-03-01 | 99.00 | 118.49 | 102.35 | 95.71 |
| 2001-05-01 | 126.87 | 104.08 | 121.38 | 97.05 |
| 2001-07-01 | 110.35 | 106.19 | 109.42 | 104.35 |
| 2001-09-01 | 141.14 | 165.34 | 142.54 | 140.12 |
| 2001-11-01 | 205.22 | 144.27 | 193.14 | 154.03 |
| 2002-01-01 | 106.33 | 102.79 | 105.99 | 109.24 |
| 2002-03-01 | 95.67 | 93.98 | 95.37 | 90.69 |
| 2002-05-01 | 74.19 | 64.27 | 71.55 | 76.07 |
| 2002-07-01 | 76.98 | 97.33 | 82.32 | 81.12 |
| 2002-09-01 | 81.36 | 89.00 | 82.26 | 78.53 |
| 2002-11-01 | 101.86 | 68.38 | 89.19 | 84.92 |
| 2003-01-01 | 123.11 | 101.43 | 118.50 | 97.94 |
| 2003-03-01 | 115.65 | 92 | 114.74 | 106.76 |
| 2003-05-01 | 120 | 59.29 | 117.66 | 107.57 |
| 2004-07-01 | 130.97 | 205 | 178.35 | 134.4 |
| 2004-09-01 | 154.09 | 212.37 | 164.98 | 159.63 |
| 2004-11-01 | 130.13 | 135.53 | 133.53 | 143.89 |
| 2005-01-01 | 166.2 | 176.77 | 167.8 | 140.33 |
| 2005-03-01 | 105.55 | 111.9 | 111.38 | 110.86 |
| 2005-05-01 | 118.61 | 132.1 | 119.35 | 131.27 |
| 2005-07-01 | 133.44 | 108.26 | 131.55 | 124.39 |
| 2005-09-01 | 171.41 | 125.77 | 135.6 | 122.03 |
| 2005-11-01 | 109.7 | 99.86 | 102.93 | 115.56 |

5.2 포트폴리오 유용성 평가

본 연구에서 제시하고 있는 PSS의 유용성을 평가한다. 기준 수익률은 전체 수익률이며 전체 수익률과 비교하여 높은 수익률을 꾸준히 유지하였을 경우 안정적인 포트폴리오로 평가하였다. 포트폴리오 데이터는 앞 절에서 제시한 1년에 6개의 포트폴리오이며 각각 포트폴리오의 수익률을 계산하고 그 결과를 전체 수익률과 비교한 결과는 표 3과 같다. 27회의 실험을 한 결과 21회의 경우 추천 포트폴리오의 수익률이 전체 수익률을 상회하였다. 2005년을 제외한 2000년에서 2004년까지의 추천 포트폴리오의 수익률이 전체 수익률보다 6~7% 높은 것을 확인할 수 있었다. 또한 수익률이 낮은 포트폴리오의 경우에도 수익률이 100%가 넘는 것을 볼 수 있었다.

6. 결론

이 논문에서는 주식 네트워크에 클러스터링 알고리즘을 적용하여 안정적인 포트폴리오를 구성하는 방법을 제안하였다. 포트폴리오는 클러스터 안의 기업 수와 수익률에 따라 추천된 클러스터의 기업으로 구성된다. 또한 클러스터링 결과와 추천 포트폴리오에 대한 정보를 화면에 표

시해주는 PSS를 구현하였다.

그리고 실제 코스피 종목의 주가 데이터를 대상으로 PSS의 성능을 측정하였다. 클러스터링을 통한 포트폴리오 9개를 이용하여 6개월 동안 투자 시물레이션을 시행하였고 전체 기업의 수익률과 포트폴리오의 수익률을 비교하였다. 수익률 결과를 분석해 본 결과 77%의 포트폴리오가 전체 수익률에 비해 안정적인 수익률을 보장하는 것을 알 수 있었다.

참고문헌

- [1] 복잡계 네트워크, 복잡계 워크샵 (복잡계 이론의 사회 과학적 적용), 삼성경제연구소, 2006.
- [2] C. W. Granger, "Investingating causal relations by econometric models and cross-spectral methods," *Econometric*, pp 24-36, 1969.
- [3] R. Engel and C. W. Granger, "Cointegration and error correction: Representation, estimation, and testing," *Econometrica*, 55:252-276, 1987.
- [4] C. Eun and S. Shim, "International transmission of stock market movements," *Rinancial and Quantitative Analysis*, 24:241-256, 1989.
- [5] B. LeBaron, "Building the Santa Fe Artificial Stock Market", Working Paper, Brandeis University, 2002.
- [6] W. B. Arthur, "Positive Feedbacks in the Economy", *Scientific American* 2:92-99, 1990.
- [7] U. B. Dufrene, O. F. Ayaid and A. Chatterjee, "Investment implications of the korean nancial market reform," *International Review of Financial Analysis*, 7:83-94, 1998.
- [8] R. W. Masulis, Y. Hamao and V. Ng, "Correlation in price changes and volatility across international stock markets," *Review of Financial Studies*, 3(2):281-307, 1990.
- [9] M. A. King and S. Wadhvani, "Transmission of volatility between stock markets," *Review of Financial Studies*, 3(1):5-33, 1990.
- [10] R. N. Mantegna, "Hierarchical structure in nancial markets," *The European Physical Journal B - Condensed Matter and Complex Systems*, 11(1):193-197, 1998.
- [11] G. Bonanno, F. Lillo, and R. N. Mantegna, "High-frequency cross-correlation in a set of stocks," *Quantitative Finance*, 1:96-104, 2000.
- [12] G. Bonanno, G. Caldarelli, F. Lillo, S. Micciche, N. Vandewalle, and R. N. Mantegna, "Networks of equities in nancial markets," *The European Physical Journal B*, 38:363-371, 2004.
- [13] R. Coelho, S. Hutzler, P. Repetowicz, and P. Richmond, "Sector analysis for a ftse portfolio of stocks," *Physica A*, 38:5-33, 2006.
- [14] 위키백과, R(프로그래밍 언어), 2011.