

인터넷 주식 토론방과 주식 시장의 상관관계 분석을 통한 투자 종목 선정 시스템

이윤정*, 김건우**, 우균**

*부산대학교 IT기반 융합산업 창의인력양성사업단

**부산대학교 컴퓨터공학과

e-mail:{leeyj01, gunwoo, woogyun}@pusan.ac.kr

The Stock Portfolio Recommendation System based on the Correlation between the Internet Stock Message Board and the Stock Market

Yun-Jung Lee*, Gunwoo Kim**, Gyun Woo**

*Creative Human Resource Development Program for IT Convergence, Pusan Nat'l Univ.

**Dept of Computer Engineering, Pusan Nat'l Univ.

요 약

인터넷게시판이나 트위터 같은 온라인 매체는 쉬운 접근성과 실시간 특성으로 어떤 사건에 대한 사용자들의 반응이 즉각적으로 나타난다. 또한, 실시간으로 엄청난 양의 데이터가 생성되고 있어 이 데이터를 잘 분석한다면 실제 사회에서 나타나는 다양한 현상들에 대해 파악할 수 있다. 최근 주식 시장에서도 이러한 온라인 데이터들을 분석하여 주가 변동이나 주식 시장 상황을 이해하려는 연구가 시도되고 있다. 이 논문에서는 주식 토론방의 게시물과 주가 사이에 어떤 상관관계가 있는지를 분석하고, 이를 이용한 주식 투자 종목 추천 시스템을 제안하고자 한다. 먼저 주가와 주식 토론방 게시물 사이의 상관관계를 분석하기 위해서 KOSPI200에 속한 회사 중 55개의 회사를 대상으로 주가와 주식 토론방 게시물을 분석하였다. 2008년부터 2013년까지 6년 동안 각 회사의 주가와 게시물의 상관관계를 분석한 결과 개별 주가와 게시물 수 사이에는 특별한 상관관계가 나타나지 않았다. 하지만 주가와 게시물 수의 상관관계가 높을수록 주식 수익률이 높은 경향을 보였다. 이 논문에서는 주가와 게시물 수의 상관관계 정보를 이용한 투자 종목 추천 알고리즘을 제안하였고, 모의투자 실험을 통해 제안 방법의 효율성을 보였다. 2008년 1월부터 2013년 12월까지의 주가와 주식 토론방 데이터를 이용한 모의투자 실험에서 제안 방법으로 구성된 포트폴리오의 1개월 평균 수익률은 약 1.82%로, 주식 네트워크 특성을 이용한 기존 방법보다 약 0.64% 높은 수익률을 보였다. 또한, 마코위츠의 효율적 포트폴리오와 KOSPI200 수익률보다 각각 약 0.85%와 1.48% 높게 나타났다.

1. 서론

주식시장에서 주가에 영향을 주는 요인으로는 유가나 환율과 같이 직접적인 측정이 가능한 요인에서부터 시장 상황이나 외부 위험과 같은 눈에 보이지 않는 요인들까지 수없이 많다. 따라서 주식시장은 시간에 따라 계속 변하고 특별한 이유 없이 주가가 급등하거나 급락하는 사건들이 발생하기도 하여 주가를 예측하는 것은 매우 어려운 일이다.

주식 시장을 이해하기 위해서 금융 관련 분야뿐만 아니라 통계나 전산 등 다양한 분야에서 주식 시장을 복잡계로 인식하고 복잡계에서 나타나는 특징들을 찾아내고 분석하는 연구도 이루어지고 있다[1-2]. 만테냐(Mantegna)는 주식 시장을 구성하는 개별 종목들의 상관관계를 기반으로 하여 주식 시장을 최소신장 트리(Minimum

Spanning Tree, MST)로 나타내어 전체 주식 시장의 위상 구조를 이해하려고 시도하였다[3-4].

그뿐만 아니라 최근에는 트위터나 인터넷 검색, 인터넷 뉴스 등과 같은 온라인 매체를 분석함으로써 주식 시장과의 상관관계를 분석하려는 연구가 진행되고 있다[5-6]. Preis et al.는 금융과 관련된 검색어의 검색량 변화가 미래의 주가 변동의 조기 신호로 해석될 수 있다고 제시하였다[7]. 98개의 검색어를 이용한 모의 투자 실험을 통해 구글 검색 데이터가 현재의 경제 상황을 반영할 뿐만 아니라 미래의 경제 활동에 관한 추세를 파악하는데 사용될 수 있음을 발견하였다. 이외에도 인터넷을 통해 실시간으로 생성되는 빅데이터를 분석함으로써 주식시장의 변화를 이해하려는 연구가 계속되고 있다.

이 논문에서는 인터넷 주식 토론방의 게시물들과 주가와 어떤 관계가 있는지를 분석하고자 한다. 또한, 그 결과를 이용한 주식 투자 종목 제안 시스템을 제안한다.

이 논문은 2013년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2013S1A5B6053791)

2. 주식 토론방 데이터 분석

이 논문에서는 주식 토론방 게시물과 주가와와의 상관관계를 분석하기 위하여 국내 유명 포털 사이트인 ‘다음’에서 제공하는 주식 토론방의 게시물들을 수집하였다. ‘다음’의 주식 토론방은 회사별로 따로 토론방이 마련되어 있어 게시물의 내용을 보지 않아도 게시물의 회사별 구분이 가능하다.

토론방이 개설된 회사 중 2005년 1월 1일부터 2013년 12월 31일까지 800건 이상의 게시물이 등록된 회사의 게시글들을 분석대상으로 하였다. 수집된 데이터는 표 1과 같다.

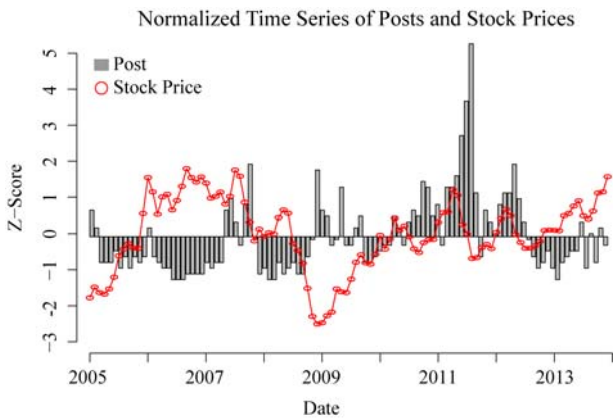
<표 1> ‘다음’ 주식 토론방 게시물 통계
(데이터 수집 기간:2005.01.01.~2013.12.31.)

구분	개수
주식 종목	55
총 게시물	166,352
회사별 1일 평균 게시물 수	0.92

대상 회사는 모두 55개이고, 9년 동안 이들 회사의 주가와 관련된 게시물들은 총 166,352개로 회사별 하루 평균 약 0.92개의 게시물이 등록되었다고 볼 수 있다.

55개의 회사 중 게시물이 가장 많은 것은 S사로 9년 동안 17,962개의 게시물이 등록되었다. 게시물이 가장 많이 생성된 기간은 2011년 8월로 한 달 동안 918개의 게시물이 생성되었다. 반면에 2006년 6월에는 5개의 게시물밖에 등록되지 않아 기간별로 게시물 수의 변동이 큰 것으로 나타났다.

주식 토론방은 자신의 관심 주식 종목에 관한 의견들을 주고받는 게시판이다. 따라서 주가에 영향을 줄 만한 사건이 발생하거나 주가의 변동에 따라 게시물의 내용이나 양도 영향을 받을 것이다. 그림 1은 S사의 주가 변동과 토론방 게시물 수의 변화를 보여준다.

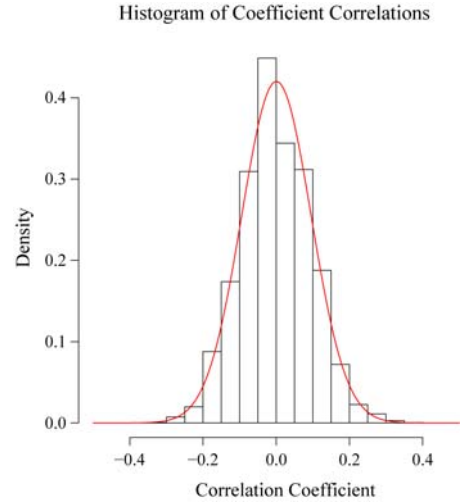


(그림 1) S사의 주가와 게시물 수의 변화 데이터

그림은 S사의 주가와 게시물 수의 시계열 데이터를 정규화하여 나타낸 것이다. 막대그래프와 실선그래프는 각각

z-score로 변환한 게시물 수와 주가를 나타낸다. 그림에서 주가가 하락하는 시기에 게시물 양이 증가하는 경향이 나타나 있는 것을 볼 수 있다.

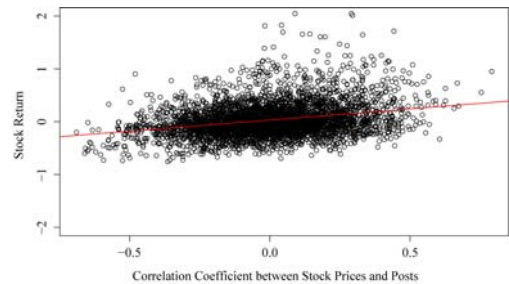
주가와 토론방 게시물의 상관관계가 어느 정도인지를 살펴보기 위해 각 회사의 주가와 게시물 수의 상관계수를 분석하였다. 그림 2는 55개 회사의 주가와 게시물 수 사이의 상관계수 분포를 보여준다. 상관계수는 6개월 간격으로 측정하였으며 피어슨 상관계수를 이용하였다.



(그림 2) 주가와 게시물 수 사이의 상관계수 분포(상관계수 측정 단위: 6개월)

상관계수 분포를 볼 때 주가와 게시물에 유의미한 상관관계를 보이는 기간도 있으나 대부분 경우는 주가와 게시물 수의 상관관계가 낮은 것으로 보인다.

다음으로 주가와 게시물의 상관관계와 수익률과의 관계를 분석하였다. 그림 3은 <주가, 게시물>의 상관계수와 그 기간의 주식 수익률 분포를 보여준다.



(그림 3) 주가와 게시물 수의 상관관계에 따른 주식 수익률 분포

그림에서 막대그래프는 상관계수의 히스토그램을 나타내고, 실선 그래프는 정규분포 곡선을 나타낸다. 6개월 단위로 각 주식의 주가와 게시물의 상관관계를 측정하고 해당 기간의 수익률을 분석한 결과 상관계수 값과 수익률 사이에 약한 선형관계가 나타났다. 그림 3에서 주가와 게시물 사이에 강한 양의 상관관계가 나타나는 경우 양의 수익률을 보일 확률이 높으며, 강한 음의 상관관계가 있는

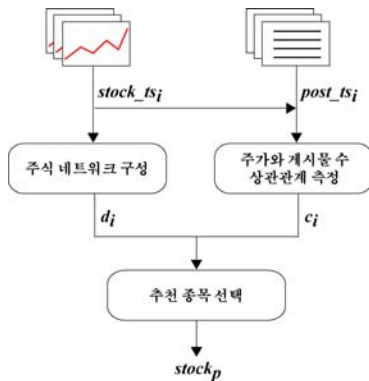
경우 수익률도 음일 확률이 높은 것으로 추정된다.

위의 분석 결과를 살펴볼 때 주가와 토론방 게시물 간에 직접적인 상관관계는 없으나 주가가 급락하거나 급등하는 시기에는 토론방 사용자들의 반응이 증가하며, 이러한 게시판의 반응이 주가의 흐름을 어느 정도 반영한다고 생각할 수 있다.

3. 투자 종목 추천 시스템

이 논문에서는 시간에 따라 변하는 주식시장에서 주가와 게시물의 상관관계를 이용하여 투자 종목을 추천하는 방법을 제안한다. 우리는 기존에 주식 시장의 네트워크 특성을 이용하여 주식 포트폴리오를 구성하는 알고리즘을 제안하였다[8]. 기존 방법에서는 효율적 포트폴리오 구성 종목들이 주식 네트워크에서 외곽에 위치하는 경향이 있다는 연구 결과를 기반으로 한다[5]. 이 방법에서는 K-means 알고리즘으로 주식들을 클러스터링하고, 각 클러스터에서 차수가 1인 종목을 하나씩 선택한다. 따라서 K-means 알고리즘의 특성상 시뮬레이션을 수행할 때마다 클러스터의 결과가 달라질 수 있으며 또한 차수가 1인 종목이 여러 개일 경우 무작위로 선택하므로 같은 기간에도 실행할 때마다 포트폴리오 구성 결과가 달라질 수 있다.

제안 방법에서는 주가와 게시물 수 간의 상관관계와 차수 정보를 함께 이용하므로 종목 선택에 좀 더 많은 정보를 제공할 수 있다. 또한, 상관계수가 임계치 이상이고 차수가 1인 종목을 모두 선택하므로 결과 포트폴리오는 변하지 않는다. 제안 방법의 알고리즘은 그림 4와 같다.



(그림 4) 추천 종목 선택 알고리즘

그림에서 $stock_ts_i$ 와 $post_ts_i$ 는 회사 i 의 주가와 주식 토론방에 등록된 게시물의 시계열 자료이다. 주식 네트워크는 개별 주식을 노드로 하고 주식간 피어슨 상관계수에 예지 가중치로 하는 최소신장 트리를 의미한다. 이렇게 주식 네트워크가 구성되면 각 회사의 차수 d_i 가 결정된다.

다음으로 주가 시계열 자료와 주식 토론방 게시물 수의 시계열 자료를 이용하여 주가와 게시물 간의 상관계수 c_i 를 계산한다. 마지막으로 최종 투자 종목 $stock_p$ 는 c_i 값이 임계치 이상이고, d_i 가 1인 종목으로 선택한다.

4. 실험 결과

제안 방법의 효율성을 보이기 위해 실제 주식들을 대상으로 하여 모의 투자 실험을 수행하고 수익률을 살펴보았다. 먼저 실험을 위해 2장에서 설명한 것처럼 코스피200을 구성하는 종목 중 800개 이상의 게시물이 등록된 55개 회사의 2008년 1월부터 2013년 12월까지 약 6년 동안의 일별 주가와 ‘다음’의 주식 토론방 게시물 데이터가 사용되었다.

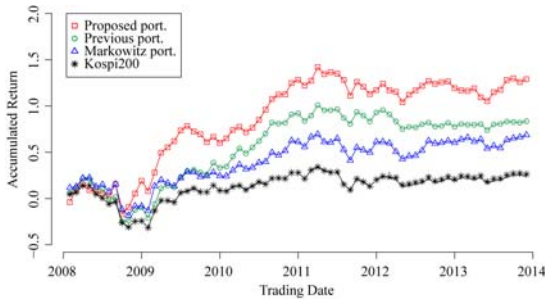
실험은 다음과 같이 구성된다. 실험 데이터를 이용해 직전 1개월의 데이터를 분석하여 다음 1개월의 투자 포트폴리오를 구성하고 수익률을 측정하였다. 포트폴리오 구성을 위해서 주가와 게시물 상관계수가 임계치 이상인 종목 중에서 차수가 1인 주식을 최종 투자 종목으로 선택한다. 실험에서 임계치는 0.3으로 설정하였다. 이때 임계치 이상인 종목 중에서 차수가 1인 종목이 없는 경우에는 차수가 가장 적은 종목을 선택한다.

제안 방법으로 구성된 포트폴리오의 수익률 분석을 위해 포트폴리오 구성은 1개월 단위로 하고, 1개월씩 이동시켜 모의 투자 실험을 수행하였다. 또한, 제안 방법으로 구성된 포트폴리오가 얼마나 효율적인지를 보이기 위해 주식 네트워크 특성만 고려한 기존 방법과 마코위츠의 효율적 포트폴리오 구성 알고리즘을 이용하여 구성된 포트폴리오의 수익률, 같은 기간의 코스피200 평균 수익률과 비교하였다. 각각의 수익률은 표 2에 정리되어 있으며, 지면 관계상 전체 수익률 결과의 일부만 나타내었다.

<표 2> 포트폴리오별 모의 투자 수익률 비교

투자기간	1개월 평균수익률(%)			
	제안방법	기존방법	마코위츠	코스피
2008년	0.48	-1.13	-0.74	-2.24
2009년	5.12	4.26	3.04	3.24
2010년	4.85	4.35	2.86	1.14
2011년	-0.98	-0.62	-1.06	-1.22
2012년	1.12	-0.14	1.04	0.71
2013년	0.21	0.16	0.53	0.22
전체	1.82	1.18	0.97	0.34

모의 투자 실험에서 실험 기간의 1개월 수익률 평균은 약 1.82%로, 기존 방법보다는 약 0.64%, 마코위츠 방법보다는 약 0.85%, KOSPI200 평균 수익률보다는 약 1.48% 높게 나타났다. 실험 결과에서 제안 방법으로 구성된 포트폴리오의 수익률이 가장 높게 나타났다. 그림 5는 모의 투자 실험 기간의 누적 투자 수익률을 보여준다.



(그림 5) 포트폴리오별 누적 수익률

그림을 통해 제안 방법으로 구성된 포트폴리오가 모의 실험의 거의 전 구간에서 높은 수익률을 보이고 있음을 알 수 있다. 따라서 제안 방법으로 생성한 포트폴리오가 효율적으로 구성되었다고 할 수 있다.

5. 결론

이 논문에서는 인터넷 주식 토론방의 이용자 반응과 주가와 상관계수를 분석하였다. 이를 위하여 '다음' 포털의 주식 토론방에 등록된 게시물을 수집하고 게시물 수의 변화와 주가의 상관계수를 측정하였다. KOSPI200을 구성하는 주식 종목 중에서 게시물이 800개 이상 등록된 55개의 주식을 대상으로 분석한 결과 주가 변화와 게시물 수 간에 뚜렷한 상관계수는 나타나지 않았다. 그렇지만 주가가 급락하거나 급등하는 시기에 게시물의 수가 증가하는 것으로 나타났다. 또한, 주가와 게시물 수의 상관계수가 강한 양이거나 강한 음을 나타내는 경우 수익률도 그에 따라 높거나 낮은 경향이 나타났다.

이러한 분석 결과를 토대로 주식 투자 종목 추천 알고리즘을 제안하였다. 제안 알고리즘은 기존 연구에서 제안 하였던 주식 네트워크의 특성과 게시물과의 상관계수를 함께 고려하여 추천 종목을 선택한다. 마코위츠의 효율적 포트폴리오에 속한 종목들은 주식 네트워크에서 외곽에 위치하는 경향이 있다는 점을 이용하여 주식 네트워크에서 해당 종목의 차수가 낮고, 게시물과의 상관계수가 높은 종목을 추천 종목으로 선택하였다.

제안 방법으로 구성된 포트폴리오의 효율성을 보기 위해 55개 회사의 주식을 이용하여 모의투자 실험을 수행하였다. 실험 결과 2008년 1월부터 2013년 12월까지의 데이터를 1개월 단위로 포트폴리오를 구성하여 실험한 결과 제안 방법으로 구성된 포트폴리오의 1개월 평균 수익률은 약 1.82%로 주식 네트워크 특성만을 이용한 기존의 방법 보다는 약 0.64%의 높은 수익률을 기록하였다. 또한, 마코위츠의 알고리즘으로 구성된 포트폴리오의 수익률과 KOSPI200 수익률보다 각각 약 0.85%와 1.48% 높게 나타났다. 따라서 제안 방법으로 구성된 포트폴리오가 효율적으로 구성되었음을 알 수 있다.

제안 방법을 활용한다면 주식시장에 대한 깊은 이해와 정보가 부족한 일반 개인투자자들도 이전 기간의 주가와

주식 토론방 데이터를 이용하여 자신의 관심 종목 분석할 수 있으며, 투자 종목을 선택할 때 활용할 수 있을 것이다.

참고문헌

- [1] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, "Dynamics of market correlations: Taxonomy and portfolio analysis," *Phys. Rev. E*, vol. 68, no. 5, p. 056110, Nov. 2003.
- [2] G. Oh, C. Eom, F. Wang, W.-S. Jung, H. E. Stanley, and S. Kim, "Statistical properties of cross-correlation in the Korean stock market," *Eur. Phys. J. B*, vol. 79, no. 1, pp. 55 - 60, Jan. 2011.
- [3] R. N. Mantegna, "Hierarchical structure in financial markets," *Eur. Phys. J. B*, vol. 11, no. 1, pp. 193 - 197, Sep. 1999.
- [4] 허화, 김승환, 강석규, and 엄철준, "주식간 연결구조와 효율적 포트폴리오," *금융공학연구*, vol. 5, no. 2, pp. 65 - 84, Dec. 2006.
- [5] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1 - 8, 2011.
- [6] T. Preis, H. S. Moat, and H. E. Stanley, "Quantifying Trading Behavior in Financial Markets Using Google Trends," *Sci. Rep.*, vol. 3, Apr. 2013
- [7] T. Preis, H. S. Moat, and H. E. Stanley, "Quantifying Trading Behavior in Financial Markets Using Google Trends," *Sci. Rep.*, vol. 3, Apr. 2013.
- [8] 이윤정, 우균, "주식 포트폴리오 추천을 위한 주식 시장 네트워크 분석," *한국콘텐츠학회논문지*, vol. 13, no. 11, pp. 48 - 58, Nov. 2013.