



사이버 공간의 정보가 주가에 미치는 영향 : 인공지능 알고리즘 기법을 이용하여*

우 민 철** (한국거래소)

김 지 현*** (한림대학교)

본 연구는 사이버 공간에서의 개인투자자들에 의해 생성되고 공유되는 정보가 주식 시장에 미치는 영향을 살펴보았다. 2013년 9월부터 2015년 6월까지 국내 주요 증권계시판인 네이버, 다음, 팍스넷을 대상으로 한국거래소에 상장된 종목의 7,054,316개 게시물과 1,662,840,007개의 조회건수를 집계하였다. 이를 위해 서포트 벡터 머신(SVM)이라는 인공지능 기법을 활용하여 게시물이 가장 많은 일자를 사전일로 하여 사전연구를 시행하였다. 주요 분석결과는 다음과 같다. 첫째, 증권계시판의 게시물 건수의 증가와 조회건수의 증가는 주가 상승을 동반하였다. 둘째, 주가에 긍정적인 키워드 혹은 부정적인 키워드가 포함된 게시물은 주가에 더 큰 영향을 주었으며 하락 키워드의 영향이 더 크게 나타났다. 셋째, 유가증권 종목보다 코스닥 종목에서 게시물 및 조회 건수가 주가에 더 많은 영향을 주었다. 이는 코스닥 시장에서 증권계시판을 이용한 부정거래 및 시세조종을 할 개연성이 더 크게 존재함을 시사한다. 넷째, 주가 수준에 따라 증권계시판의 키워드가 주가에 미치는 영향이 다르게 나타났다. 이는 개인투자자 중에서 사이버 정보를 주된 정보원으로 이용하는 투자자 그룹과 증권사 애널리스트 보고서 등 유료 정보원의 활용가능성이 큰 투자자 그룹 간에 증권계시판 영향력 차이가 존재함을 시사한다.

[1] 서 론

개인투자자 매매가 주가에 영향을 미치는가에 대한 질문은 학계의 주요 연구주제이다. 최근, IT 기술의 폭발적 성장과 함께 개인투자자들이 활발한 인터넷 및 모바일 활동을 하게 되면서, 디지털 세계에서 표현

* 본 논문은 2017년 하반기 펀드평가3사(한국펀드평가, FnGuide, 제로인)의 성균관대학교 자산운용연구센터(CAPM) 연구비 지원으로 수행되었습니다.

주제어 : 사이버정보, 증권계시판, SVM(Support Vector Machine), 개인투자자, 정보생성자
JEL 분류기호 : G12, G14

** 주저자, 한국거래소 시장감시부

*** 교신저자, 한림대학교 경영대학 재무금융학과 (E-mail: jhyunkim@hallym.ac.kr)

된 개인투자자의 활동이 향후 주가를 예측할 수 있는지에 대한 학계와 산업계의 관심이 높아지고 있다.

재무학계의 이론에서 개인투자자는 비정보 거래자 또는 잡음 거래자(noise trader)로 정의되어 왔으며, 행태적 오류를 겪는다고 전제하였다. 중요 정보가 주식시장에 미치는 영향을 분석한 기존 연구들은 공통적으로 개인투자자를 정보수령자로 간주한다. 그러나, 2009년 스마트폰의 개발 이후 급속도로 확산된 사회관계망 서비스(SNS), 블로그 등을 통하여 개인투자자들은 동질집단을 형성하고 정보교류가 활발해지면서, 정보수령자인 동시에 정보생성자가 되고 있다.

딜로이트(2007)에 따르면, 미국 인터넷 소비자의 82%가 동질집단 평가(peer-review)에 의해 직접적인 영향을 받는다고 한다. 데이터모니터(2010)는 미술랭가이드나 소비자보고서와 같은 전문매체에 대한 정보의존도는 감소하는 반면, 아마존닷컴이나 엘프닷컴에서 사용자들이 작성한 사용후기와 같은 동질집단 추천에 대한 의존도는 증가하고 있다고 했다.

Keller and Tetlock(2013)은 개인투자자는 지속적인 학습과 손실과정을 통해 투자능력이 향상되고 있으며, 사회관계망 서비스가 이들을 정보거래자로 만드는 채널 중 하나라고 하였다. Chen et al.(2014)은 개인투자자들은 주식에 대한 자신들의 분석을 소셜 미디어를 통해 공유하고, 소셜 미디어에 모여진 투자자들의 의견은 미래 주가 및 비기대 이익을 예측할 수 있다고 주장한다.

우민철과 이우백(2013)에 따르면, 국내 주식시장도 투자자들의 주된 주문매체였던 홈트레이딩 시스템(HTS)의 비중은 점차 감소하는 반면, 모바일트레이딩 시스템(MTS)의 비중이 증가하고 있으며, 증권사 IT 개발부서의 주된 업무도 홈트레이딩 시스템보다는 모바일트레이딩 시스템의 성능개선으로 옮겨지고 있다. 스마트폰을 기반으로 한 모바일트레이딩 시스템의 발달은 풍부한 정보의 수집과 빠른 정보 서핑이 장소에 구애 받지 않는 유비쿼터스 환경에서 가능하도록 구현

하였다. 또한, 다양한 사회적 관계망을 통해 자신의 정보를 공유함으로써 새로운 정보를 창출하고 이를 바탕으로 투자의사결정을 내리는 수준까지 발전하였다.

Barber and Odean(2002)이 주장한 바와 같이, 정보 수집 능력과 분석 능력이 확대됨에 따라 개인투자자들이 과도한 자신감에 빠지고 이로 인해 투자수익에 악영향을 준다는 연구 결과도 있지만, 증권 게시판의 활성화가 주식 시장을 움직일 수 있다는 연구들도 지속적으로 발표되고 있다. Wysocki(1999)는 게시물이 많은 상위 50개 종목을 분석한 결과, 게시물이 익일 거래량 및 주가에 대한 예측력이 있음을 확인하였다. Tumarkin and Whitelaw (2001), Antweiler and Frank(2004), Harris and Raviv(1993), Das and Chen(2007), Gilbert and Karahalios(2009) 등도 게시물에 의한 주식시장 영향을 분석하고 있다.

본 연구는 2013년 9월부터 2015년 6월까지 증권 게시판이 가장 활성화 되어 있는 네이버, 다음, 팍스넷의 전체 게시글을 크롤링(crawling)¹⁾ 방식을 이용하여 자료를 수집하고, 인공지능 방법의 하나인 서포트 벡터 머신(Support Vector Machine, 이하 SVM) 기법을 이용하여 분류·집계하고 이를 근거로 증권 게시판의 활동이 주식시장에 미치는 영향을 분석하였다.

국내 주요 3사의 사이버 공간에는 2,120개의 종목 게시판이 개설되어 있었고, 일반주식 이외에도 ETF, 신주인수권증서, 선박투자회사, 부동산투자회사 등에 대한 게시판도 다수 존재하고 있었다. 분석 기간 중에 집계한 게시물은 약 705만 건이며, 게시물에 대한 조회건수는 약 16억 6,284만 건으로 게시물 당 평균 236번 조회건수를 나타냈다. 코스닥 종목의 게시물에 대한 조회건수가 유가증권시장 종목의 게시물에 대한 조회건수보다 상대적으로 많았다.

주요 분석결과는 다음과 같다. 첫째, 증권 게시판의 활동이 주가에 통계적으로 유의한 영향을 주는 것을

1) 크롤링(crawling) 방식이란 무수히 많은 컴퓨터에 분산·저장되어 있는 문서를 수집하여 검색 대상의 색인으로 포함시키는 기술이다. 어느 부류의 기술을 얼마나 빨리 검색 대상에 포함시키나 하는 것이 우위를 결정하는 요소로서 최근 웹 검색의 중요성에 따라 발전되고 있다.(IT용어사전, 한국정보통신기술협회)



확인하였다. 게시물 건수가 증가함에 따라 주가가 상승했으며 조회건수의 증가 역시 주가 상승을 동반하였다. 둘째, 긍정적인 키워드 혹은 부정적인 키워드²⁾가 명확히 포함된 게시물의 건수는 주가에 더욱 영향을 주는 것으로 나타났다. 주가에 영향을 주는 절대 수준은 부정적 키워드가 긍정적 키워드보다 더 컸으며, 일부 부정적 키워드는 지속적인 주가 하락을 수반하고 있었다. 셋째, 유가증권시장 종목보다 코스닥시장 종목에서 게시물 건수 및 조회 건수가 주가에 미치는 영향이 더 크게 나타났다. 증권계시판 활동이 주가에 영향을 준다면, 이를 악용하는 경우가 발생할 수 있다. 따라서 이러한 결과는 증권계시판을 악용한 부정거래 및 시세조종 등의 불공정거래가 코스닥 종목에서 발생할 개연성이 더 높다는 것을 시사한다. 넷째, 가격대별로 구분하여 동일한 분석을 진행한 결과, 주식 가격대에 따라 증권계시판 활동이 주가에 미치는 영향에 차이가 있음을 확인하였다. 특히, 1천원 이상 1만원 이하의 저가 종목 및 1천원 미만의 초저가 종목 등 개인투자자의 투자 비중이 많은 종목에서 게시물 건수 및 조회건수와 주가와 관계가 더 뚜렷하게 나타났다.

본 연구는 국내 주식시장에 상장된 종목들의 주가에 사이버 공간의 정보가 어떠한 영향을 주는지 분석한 최초의 연구라는데 의미가 있다. IT 시대 속에서 개인들은 언제 어디서나 사이버 공간에서 모바일 텍스트를 통해서 정보를 교류하고 자신을 표현한다. 따라서 모바일 텍스트는 이제 개인의 감성을 표현하는 가장 중요한 매체로 자리매김하고 있다. 금융공학계는 투자

자 감성이 주식시장에 미치는 영향을 알아내기 위해 투자심리지수 등을 개발하여 관련 연구들을 진행해왔다. 따라서 본 연구는 4차 산업혁명 시대를 앞둔 시점에서 개인들의 새로운 정보 소통 환경인 사이버 증권계시판의 주식시장에 대한 영향을 최초로 분석하였을 뿐만 아니라, 긍정성과 부정성이라는 이진 감성 분류를 통해 텍스트에 표현된 개인들의 감성을 분류함으로써 그 영향력을 세분하여 분석했다는 점에서도 의의가 있다.

크롤링 방법으로 사이버 공간의 정보를 수집하고, 인공지능 알고리즘인 SVM 기법을 활용한 점은 저자들이 아는 범위에서는 일부 미국 시장을 대상으로 한 연구가 존재할 뿐, 국내 선행연구를 찾기 어렵다. 본 기법은 기존에 많이 사용되던 머신러닝 기법인 나이브 베이즈 보다 발전된 인공지능 분류 알고리즘으로 인터넷 게시판과 같은 빅데이터 분류에 용이하다. 인터넷 및 모바일을 통해 텍스트를 입력하는 환경에서는 글의 오타·탈자가 많이 발생하고 띄어쓰기의 오류 및 약어, 은어 등도 많기 때문에 기존의 감성 분류 알고리즘의 경우, 분류 정확도가 저하된다. 따라서 텍스트의 감성을 긍정성과 부정성으로 구분하여 분석한 본 연구의 경우, 이러한 모바일 텍스트 환경에서 보다 높은 성능을 보이는 SVM 기법을 이용함으로써 분석의 정확도를 높였다는 점에서 의의가 있다.

이하의 논문 구성은 다음과 같다. 제 2장은 연구자료 및 방법론을 설명하고 선행연구를 제시하며, 제 3장에서 분석 결과를 제시한다. 제 4장은 결론 및 시사점이다.

[2] 연구자료

2.1 연구자료 및 방법론

국내 주식시장에 참여하고 있는 많은 투자자들은

자신이 투자한 기업 또는 투자하기 위해 분석 중인 기업에 대한 정보를 얻기 위해 증권사 애널리스트 보고서 및 거래소의 공시자료 등을 활용한다. 더불어,

2) SVM기법을 이용하여 키워드의 긍정성과 부정성을 분류하였다. 예를 들어, 주가가 상승한 경우에 나타나는 주요 키워드와 주가가 하락한 경우에 나타나는 주요 키워드를 분류하고, 동 키워드가 포함된 호재성 게시물과 악재성 게시물을 각각 구분하였다.

사회관계망 서비스(SNS)에 게시된 글이나 평가를 수집하면서 자신의 투자 의사결정에 활용하게 된다. 본 연구는 수많은 SNS 중에서 참여자 수가 가장 많은 것으로 알려진 네이버, 다음, 팍스넷의 게시글을 이용하였다.

무수히 많은 데이터를 수집·분류하여 색인화하는 크롤링 방식을 활용하여 2013년 9월부터 2015년 6월까지의 주요 3개 사이트의 증권게시판을 대상으로 실시간으로 게시글을 수집하여 데이터베이스를 구축하였다. 또한, 구축된 게시글 데이터베이스를 대상으로 인공지능 기법인 SVM을 활용하여 호재성 단어와 악재성 단어를 분류하였다. 인공지능 기법은 오늘날 다양한 분야에서 활용되고 있으며 활용 범위도 급속히 확산되고 있다.

본 연구에서 활용된 인공지능과 관련된 기본적인 개념들은 다음과 같다. 딥러닝이란 사람의 학습 능력을 모방하기 위해 다양한 컴퓨터 알고리즘과 방법론을 연구하는 분야로 인공신경망의 한계를 극복하기 위해 제안된 머신러닝 방법이다. 컴퓨터는 스스로 생각하여 판단하는 능력을 보유하고 있지 않기 때문에 수많은 데이터를 통해 사람의 정보처리능력을 모방하게 된다. 딥러닝의 핵심은 분류를 통한 예측이며, 무수히 많은 데이터 속에서 패턴을 발견해 인간이 사물을 구분하듯이 컴퓨터가 데이터를 분류하는 것이다. 분류 방식은 지도 학습과 비지도 학습의 두 가지 방법이 있다.

머신러닝 알고리즘은 대부분 지도 학습을 기초로 하고 있다. 예컨대, 수백만 장의 고양이 사진과 강아지 사진을 보여주면서 학습을 시키면, 컴퓨터는 학습된

결과를 바탕으로 고양이와 강아지를 구분하는 것이다. 이렇게 데이터를 통해 학습할 때, 데이터에 포함된 공통된 속성을 파악하여 통계적 해답을 찾게 되기 때문에, 데이터를 빠르고 정확하게 분류하고 그 속성을 제대로 파악하는 것이 필수적이다. 비지도 학습은 배움의 과정이 없다는 점이 다르다. 특정 사진이 고양이라는 것을 컴퓨터가 스스로 학습한다는 것이다. 따라서 지도 학습에 비해 진보한 기술이며, 컴퓨터의 높은 연산 능력이 요구된다. 이세돌과 대국을 했던 알파고는 과거 수백만 개의 기보를 학습했고 학습된 결과를 바탕으로 최적의 위치에 돌을 놓아 갔다면, 최근 커제와 대국을 했던 알파고 2.0은 기보를 보여주지 않는 비지도 학습을 통한 머신러닝을 한 것으로 알려졌다.

본 연구에서 활용된 SVM 기법은 지도 학습으로 KNN(K-Nearest Neighbors) 모델과 더불어 내용 분류를 위한 대표적인 최신 분류 알고리즘에 속한다.³⁾ 과거에는 분류 알고리즘으로 나이브 베이지안 방법을 많이 사용하였으나, 최근에는 Antweiler and Frank(2004)에서 언급된 것처럼 새로운 분류 알고리즘이 등장하여 다양한 분야에서 활용되고 있다.

SVM 기법의 분류 방법은 다음과 같다. 수집된 게시글을 분류한 후, 분류된 작업을 근거로 감성어(Lexicon) 사전⁴⁾을 작성한다. 단어를 분류하는 과정에는 감성 분석의 정확성이 가장 높다고 평가되는 Lexicon Semantic Pattern(이하, LSP)를 활용하였다. LSP는 구문 분석의 중의성, 모호성 문제를 형태소 분석과 자연어 처리, 동의어/유의어/원형복원 등을

3) KNN 알고리즘은 새로운 데이터가 들어와서 분류를 해야 할 때 가까운 k개의 데이터들을 찾아서 어디에 분류할지를 결정한다. 예를 들어 크기가 작고 무게가 가벼우면 곱이고 크기가 크고 무게가 무거우면 오렌지로 분류할 때, 새로운 데이터의 크기와 무게를 보고 곱에 가까운지 오렌지에 가까운지를 이용하여 분류하는 기법이다. SVM 알고리즘은 이질적인 그룹 간의 거리를 최대화 하는 기준을 나누고 이 기준을 이용하여 분류를 하는 방법이다. 따라서 기준은 이질적인 그룹 사이의 거리인 벡터를 최대한 멀리 나누어 주는 것이 좋다. 예를 들어, 곱의 무게 중 가장 무거운 것이 200그램이고 오렌지 중 가장 가벼운 것이 300그램이라면 분류 기준을 250그램으로 나눌 수 있다. 한편, 나이브 베이지안 방법은 크기와 무게를 독립적으로 보아서 크기와 무게 사이에 관계가 없다고 가정하는 단순한 방법이다. 보다 자세한 내용은 Vapnik(1995), Vapnik and Chervonenkis(1964), Joachims(1999) 등을 참고하기 바란다.

4) 감성 분석(Sentiment Analysis)이란 SNS 게시글, 모바일 메시지 등의 텍스트에 어떠한 감성이 나타나 있는지를 분석하는 기법으로, 데이터 수집, 형태소 분석, 극성 탐지의 3단계로 이루어진다. 텍스트의 감정 및 의견을 추출하는 극성 탐지를 위해서는 문서 단위의 극성 분석, 속성 단위의 극성 분석, 사전 단위의 극성 분석이 있다. 사전 단위의 극성 분석의 경우, 지정 단어에 따라 미리 감성어 사전(Lexicon 사전)을 만들고, 이 사전에 있는 단어와 텍스트를 분석하여 텍스트의 극성을 긍정, 중립, 부정적으로 분류할 수 있다. 한국정보통신기술협회의 IT용어사전에서는 '소비자의 감성과 관련된 텍스트 정보를 자동으로 추출하는 텍스트 마이닝(Text Mining) 기술의 한 영역'으로 '문서를 작성한 사람의 감정을 추출해 내는 기술로 문서의 주제보다 어떠한 감정을 가지고 있는가를 판단하여 분석한다.'고 설명하고 있다.



통해 해결하여 분석결과가 정확한 것으로 알려졌다. 또한, 의미 자질에 431,706개의 단어와 400개의 특허 사전을 이용하여 분석함으로써 분석결과와 정확성을 보완하였다. 선행 연구들이 게시물의 건수를 설명 변수로 사용한 반면, 본 연구는 Gilbert and Karahalios(2010)가 제시한 바와 같이 특정 어휘가 종목에 대한 매수·매도 추천 의도를 명확히 표현한다는 측면에서 질적 설명변수를 추가했다는 점에서 차별성을 갖는다.

2.2 선행연구

증권 게시판이 주가에 미치는 영향에 대한 이론적 배경은 다음과 같다.

정보 가설에 따르면, 증권 게시판에 새로운 정보가 포함되어 있어 게시물은 향후 주가에 대한 예측력을 가지고 있게 된다. 투자심리 가설에 따르면, 증권 게시판에 새로운 정보는 없지만 현재 주가 및 거래량 등에 반영된 정보 이상으로 시장에 형성된 투자 심리를 나타내는 지표로서 역할을 한다. 군집 가설에 따르면, 증권 게시판에 펀더멘탈과 관련된 정보가 없음에도 불구하고 투자자들이 게시판의 매수 추천이나 매도 추천을 따르기 때문에 주가가 효율적인 가격에서 이탈하게 된다고 주장한다. 마지막으로, 데이터레이더가 증권 게시판에 형성된 일시적 쓸림현상을 인지하고 이를 추종하면서 주가 및 거래량에 영향을 준다는 데이터레이더 가설도 있다.

증권 게시판의 글들이 주가 등에 미치는 영향을 분석한 선행연구들 중에서 일부를 소개하면 다음과 같다. Wysocki(1999)는 증권사이트인 야후 파이낸스(Yahoo! Finance)에 게시된 3,000여 종목을 분석하여 게시물의 건수가 익일 거래량과 수익률에 대한 예측력이 있다는 결과를 제시하였다.

Tumarkin and Whitelaw(2001)는 레이징닷컴(Raging.com)이라는 증권 게시판을 대상으로 프로그래밍 언어의 일종인 펄 스크립트를 이용하여 1999년 4월 7일부터 2000년 2월 18일 사이에 게시된 73종목의 181,633개 게시물을 수집하였다. 이렇게

생성된 데이터베이스를 통해 해당 종목의 주가 및 거래량과의 관계를 분석한 결과, 게시물의 건수가 많은 사건일 이후에는 유의적인 수익률 증가는 발견되지 않았다. 그러나, 긍정적 게시물 건수가 많은 사건일 이전에 게시물 건수와 수익률 간에 유의적인 관계가 있음을 보였다. 또한 사건일 당일에도 게시물 활동량과 수익률 간에 유의적 관계가 발견되었다.

Antweiler and Frank(2004)는 인기 증권사이트인 야후 파이낸스 및 Raging Bull에 2000년 1년 동안 게시된 45개 종목의 150만개 이상의 게시물을 컴퓨터 언어학적 방법을 사용하여 분석하였다. 사용된 방법은 기존에 많이 사용되던 나이트 베이지안 알고리즘 방법과 최근에 분류 작업에 많이 사용되고 있는 SVM 알고리즘 방법이다. 그 결과, 증권 게시판의 게시물이 주가 변동성을 예측하는 반면, 주가 수익률을 예측하지는 못한다고 하였다. 또한, 긍정 게시물의 건수와 부정 게시물의 건수가 유사한 경우는 거래량이 감소한다는 것을 보였다. 이는 투자자의 의견이 비동질화 되었을 때 거래량이 감소한다는 Harris and Raviv(1993)의 이론을 뒷받침한다. Das and Chen(2007)은 증권사이트인 야후 파이낸스에 2011년 6월부터 8월까지 게시된 기술 섹터에 속한 24종목의 145,110개 게시물을 이용하여 감성 지수가 익일의 주가 및 거래량과 양의 관계가 있음을 제시하였다. 이들은 게시물에 반영된 심리를 도출하기 위한 자연어 처리를 하기 위해서, 다양한 전통적 분석 기법을 활용한 하이브리드 분류 알고리즘을 도입하여 투자 심리 지수를 더욱 정교하게 도출하고자 하였다.

Hanke and Hauser(2008)는 'Crummy database'에 있는 2005년의 235개 종목의 주식 스팸 메일을 분석한 결과, 스팸 메일이 주식수익률, 변동성, 장중 스프레드 및 거래량을 증가시키며, 유동성이 적은 종목일수록 스팸 메일을 통한 주가조작이 쉽다는 것을 확인하였다. 한편, 스팸 메일의 내용 자체가 수익률에 영향을 미치지 않는다고 보고하였다. Frieder and Zittrain (2007)는 인터넷 토론 게시판인 'NANAS'에서 얻은 1,775,743개의 주식호봉성 스팸 메일을 분석한 결과, 과도하게 부풀린 스팸이 특정

종목의 거래량에 영향을 주었으며 과장 선전한 날 매수한 종목을 익일에 매도하여 평균 5.5%의 이익을 얻었다는 결과를 제시하였다.

Gilbert and Karahalios(2010)는 라이브저널(LiveJournal) 블로그에 있는 20,110,390개 게시물을 이용하여 'Anxiety'라는 감정 게시물이 증가할 때마다 S&P500 지수가 하락하는 것을 예측할 수 있다는 결과를 제시하였다. Tetlock(2007)은 1984년부터 1999년 동안 월스트리트 저널의 칼럼에서 추출한 콘텐츠를 이용하여 미디어와 주식시장의 관계를 연구한 결과, 부정적 심리가 기업 가치 하락에

대한 예측력을 갖는다고 주장하였다. 이들은 이러한 분석을 위한 감성어 사전을 만들기 위해 'General Inquirer'라는 내용 분석 프로그램을 사용하였다.

이처럼 증권 게시판에서 정보의 생성 및 공유 활동이 주식 시장에 미치는 영향에 대한 연구 결과들은 혼재되어 있다. 한편, 미국 시장을 대상으로 한 관련 연구는 활발히 진행되고 있는 반면, 국내 주식시장을 대상으로 한 연구는 거의 전무한 실정이다. 이에 본 연구에서는 국내에서 활발한 활동을 보이는 3개사의 증권 게시판을 이용하여 국내 투자자들의 사이버 공간에서의 활동이 주식 시장에 어떠한 영향을 미치는지를 살펴보고자 한다.

[3] 분석 결과

3.1 분석 자료

한국 주식시장에 상장된 종목들에 대한 증권 게시판의 게시물이 주식시장에 미치는 영향을 분석하기 위해 2013년 9월부터 2015년 6월까지 네이버, 다음 및 팍스넷의 증권 게시판에 있는 게시물을 크롤링 방식으로 수집하여 데이터베이스를 구축하였다. 전체 게시물을 대상으로 SVM 기법을 이용하여 게시물의 내용을 분석하고, 우호적 게시물과 비우호적 게시물로 구분하고 게시물건수와 조회건수 등이 주가에 미치는 영향을

분석하였다.

증권 게시판은 종목별로 개설되어 있으며, 개설된 종목은 2,120개였다. 유가증권 종목이 1,037종목으로 49%를 차지하였고, 코스닥 종목은 1,080종목으로 51%를 차지하였다. 2013년에 개설된 코넥스 시장에 상장된 종목의 게시판도 3개 포함되어 있었다. 증권 게시판이 개설된 종목들을 종류별로 구분하면, 주권이 87.2%로 가장 많았고 ETF 및 신주인수권증서, 선박투자회사 등에 대한 게시물이 12.8%였다. 특히, 외국주권은 중국원양자원만 게시물이 형성되었다.

〈표 1〉 품목별, 시장별 종목 현황

| | 유가증권 종목수 | 코스닥 종목수 | 코넥스 종목수 | 합산 종목수 | 전체대비 비중(%) |
|-------------|-------------|------------|------------|-----------|---------------|
| 주권 | 796 | 1,050 | 3 | 1,849 | 87.2 |
| ETF | 165 | | | 165 | 7.8 |
| 신주인수권증서 | 20 | 30 | | 50 | 2.4 |
| 선박투자회사 | 33 | | | 33 | 1.6 |
| 부동산투자회사 | 9 | | | 9 | 0.4 |
| 투자회사 | 6 | | | 6 | 0.3 |
| 수익증권 | 4 | | | 4 | 0.2 |
| 신주인수권증권 | 2 | | | 2 | 0.1 |
| 외국주권 | 1 | | | 1 | 0.0 |
| 사회간접자본투융자회사 | 1 | | | 1 | 0.0 |
| 합 계(종목) | 1,037 | 1,080 | 3 | 2,120 | |



일별, 종목별로 수집한 게시물 건수와 게시물에 대한 조회건수를 분석한 결과, 총 7,054,316건의 게시물이 있었으며 게시물에 대한 조회건수의 합은 1,662,840,007건이었다. 주권에 대한 게시물이 7,004,338건으로 전체 게시물의 99.29%를 차지했으며, 게시물에 대한 조회건수 중에서 주권 게시물에 대한 조회건수가 1,650,930,747건으로 전체의 99.28%를 차지하였다.

증권 게시판에 게시된 게시물을 불특정인이 열람할 때마다 조회건수가 증가하며 당일 자정을 기준으로 합산하였다. 특정 종목에 대한 게시물 건수가 동 종목에 대한 직접적인 관심 지표라고 한다면, 특정 종목에 대한 게시물의 조회건수는 동 종목에 대한 간접적인 관심 지표로 평가할 수 있다.

게시물 건수 대비 조회건수는 평균 237배였으며, ETF 게시물에 대한 조회건수가 294배로 주권 게시물에 대한 조회건수 비율인 235배 보다 높았다. 이는 ETF에 대한 직접적인 관심 지표보다 간접적인 관심 지표가 상대적으로 더 높았음을 의미한다.

유가증권 종목에 대한 게시물 건수 대비 조회건수 비중은 253배, 코스닥 종목에 대한 게시물 건수 대비 조회건수 비중은 223배로, 유가증권 종목에 대한 간접적 관심 지표가 상대적으로 높게 나타났다. 선박투자회사의 경우, 게시물 건수보다 조회건수가 적었으며 수익증권의 경우에는 게시물에 대한 조회건수가 전혀 없었다. 이들 품목의 경우, 일부 투자자들에게만 관심이 있는 종목이라는 것을 방증하는 결과라 하겠다.

〈표 2〉 품목별, 시장별 게시물수/조회건수

| 종 류(종목수) | | 유가증권시장 | 코스닥시장 | 코넥스시장 | 총합계 | 비 중 |
|-----------------|-----|-------------|-------------|-------|---------------|-------|
| 주권 (1,849) | 게시물 | 3,011,272 | 3,993,058 | 8 | 7,004,338 | 99.29 |
| | 조회수 | 761,599,699 | 889,329,872 | 1,176 | 1,650,930,747 | 99.28 |
| ETF (165) | 게시물 | 35,885 | | | 35,885 | 0.51 |
| | 조회수 | 10,565,971 | | | 10,565,971 | 0.64 |
| 신주인수권 증서(50) | 게시물 | 2,958 | 2,921 | | 5,879 | 0.08 |
| | 조회수 | 706,826 | 623,149 | | 1,329,975 | 0.08 |
| 선박투자 회사(33) | 게시물 | 1,068 | | | 1,068 | 0.02 |
| | 조회수 | 596 | | | 596 | 0.00 |
| 외국주권 (1) | 게시물 | 244 | | | 244 | 0.00 |
| | 조회수 | 12,718 | | | 12,718 | 0.00 |
| 수익증권 (4) | 게시물 | 6,902 | | | 6,902 | 0.10 |
| | 조회수 | | | | | |
| 전 체 (2,102) | 게시물 | 3,058,329 | 3,995,979 | 8 | 7,054,316 | |
| | 조회수 | 772,885,810 | 889,953,021 | 1,176 | 1,662,840,007 | |

Gilbert and Karahalios(2010)가 제시한 바와 같이 특정 어휘는 종목에 대한 매수 또는 매도와 관련한 추천 의도를 보다 명확히 표현하고 있다. 본 연구에서도 분석 자료를 대상으로 주가 상승 또는 하락 시점에 많이 언급되는 키워드를 분석하였다. SVM 기법을 통해 파악한 결과, 주가가 상승한 경우에 많이 언급되는 키워드로 “대장, 상한, 고점, 홀딩” 등이 있었으며,

주가가 하락한 경우에 많이 언급되는 키워드로는 “손실, 대기, 개인, 유증” 등이 있었다.

주가 상승 키워드와 주가 하락 키워드가 혼재되어 호재성 또는 악재성 게시물의 분류가 모호한 경우와 달리, 호재성 게시물로 분류되면서 주가 상승 키워드가 포함된 게시물은 전체의 5.65%이며, 악재성 게시물로 분류되면서 주가 하락 키워드가 포함된 게시물⁵⁾은

전체의 5.16%였다. 유가증권 시장에 속한 주권의 경우, 상승 키워드가 포함된 게시물 건수(6.22%)가 하락 키워드가 포함된 게시물 건수(4.94%) 보다 더 많았으며, 코스닥 시장에 속한 주권의 경우 상승 키워드가 포함된 게시물 건수(5.22%) 보다 하락 키워드가 포함

된 게시물 건수(5.37%)가 더 많았다. 선박투자회사의 경우 1,068개 게시물 중에서 상승 키워드가 포함된 게시물 건수가 단 2개였으며, 하락 키워드가 포함된 게시물 건수는 전무하였다.

〈표 3〉 품목별, 시장별 주요 키워드를 포함한 게시물건수

| 종 류(종목수) | 키워드 | 유가증권시장 | | 코스닥시장(건, %) | | 총합계(건, %) | |
|-----------------|---------------------------------|-----------|-------|-------------|-------|-----------|-------|
| | | 건수 | % | 건수 | % | 건수 | % |
| 주권 (1,849) | 상 승 | 187,207 | 6.22 | 208,605 | 5.22 | 395,812 | 5.65 |
| | 하 락 | 148,756 | 4.94 | 214,298 | 5.37 | 363,054 | 5.18 |
| | 게시물 | 3,011,272 | | 3,993,058 | | 7,004,338 | |
| ETF (165) | 상 승 | 1,691 | 4.71 | 0 | | 1,691 | 4.71 |
| | 하 락 | 1,026 | 2.86 | 0 | | 1,026 | 2.86 |
| | 게시물 | 35,885 | | | | 35,885 | |
| 신주인수권 증서(50) | 상 승 | 373 | 12.61 | 437 | 14.96 | 810 | 13.78 |
| | 하 락 | 373 | 12.61 | 330 | 11.30 | 703 | 11.96 |
| | 게시물 | 2,958 | | 2,921 | | 5,879 | |
| 외국주권 (1) | 상 승 | 11 | 4.51 | 0 | | 11 | 4.51 |
| | 하 락 | 3 | 1.23 | 0 | | 3 | 1.23 |
| | 게시물 | 244 | | | | 244 | |
| 선박투자 회사(33) | 상 승 | 2 | 0.19 | 0 | | 2 | 0.19 |
| | 하 락 | | | | | | |
| | 게시물 | 1,068 | | | | 1,068 | |
| 총합계 (2,098) | 상 승 | 189,284 | 6.19 | 209,042 | 5.23 | 398,326 | 5.65 |
| | 하 락 | 150,158 | 4.91 | 214,628 | 5.37 | 364,786 | 5.17 |
| | 게시물 | 3,058,329 | | 3,995,979 | | 7,054,316 | |
| 주가 상승시 | 대장, 상한, 고점, 홀딩, 사업, 인수, 감사, 게임 | | | | | | |
| 주가 하락시 | 손실, 대기, 개인, 유증, 대주주, 문제, 악재, 감사 | | | | | | |

3.2 증권 게시판이 주식 시장에 미치는 영향 분석

증권 게시판의 게시물이 주식 시장에 미치는 영향을 분석하기 위하여 게시물 건수가 가장 많았던 시점을 사건일로 하여 t-5일에서 t+5일까지의 초과수익률을 산출하였다. 초과수익률은 Fama and French 3요인 모형으로 추정된 수익률을 차감하여 산출했으며, 시장 대표지수 수익률을 이용하여 산출한 시장 대비 초과수익률은 주가 방향성에 있어 큰 차이가 없기 때문에

별도의 표로 제시하지는 않았다.

〈표 4〉는 종목별, 일별로 게시물 건수 및 게시물에 대한 조회건수를 측정된 데이터베이스와 주가 자료를 이용하여 사이버 공간의 정보가 주식 시장에 미치는 영향을 분석한 결과를 보여주고 있다. 게시물 건수는 00시 00분 00초부터 24시간 동안 게시물 건수를 집계하였고, 조회 건수는 게시물이 게재된 시각을 기준으로 10일간 조회된 건수를 집계하였다. 10분 간격으로 크롤링 방식으로 게시물을 집계하면서 중복된

5) 표현의 중복을 줄이기 위해 상승(하락) 키워드가 포함된 게시물로 서술한다.



게시물은 제외하였고 조회건수는 최근 값으로 업데이트하는 방식을 사용하였다. 예를 들어 10시 00분에 삼성전자의 게시물이 A, B, C의 3건이 있었고, 조회건수가 각각 10건, 20건, 30건이었다고 가정하고, 23시

50분에 삼성전자의 게시물이 A, B, C, D의 4건이고, 조회건수가 각각 11건, 20건, 32건, 1건 이라면, 당일의 삼성전자 게시물은 4건, 게시물 당 조회건수는 11건, 20건, 32건, 1건으로 기록된다.

〈표 4〉 게시물 건수 및 조회 건수에 대한 주가반응

| 수익률 | T-5 | T-4 | T-3 | T-2 | T-1 | T-0 | T+1 | T+2 | T+3 | T+4 | T+5 |
|------------------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Panel A : 게시물 건수 | | | | | | | | | | | |
| 전체 | 0.21 | 0.35 | 0.63* | 0.44* | 0.63* | -0.16 | -0.31 | -0.23 | -0.29 | -0.22 | 0.06 |
| 유가 | 0.10 | 0.59* | 0.59* | 0.43* | 0.39 | -0.04 | -0.12 | -0.01 | -0.10 | 0.06 | 0.18 |
| 코스닥 | 0.28 | 0.20 | 0.65* | 0.45* | 0.79* | -0.23 | -0.43 | -0.38 | -0.42 | -0.40 | -0.02 |
| Panel B : 조회건수 | | | | | | | | | | | |
| 전체 | 0.26 | 0.31 | 0.46 | 0.54* | 0.62* | 0.80* | -0.21 | -0.19 | -0.30 | -0.24 | 0.06 |
| 유가 | 0.29 | 0.44* | 0.47* | 0.50* | 0.57* | 0.71* | 0.11 | 0.03 | -0.17 | -0.02 | 0.08 |
| 코스닥 | 0.23 | 0.19 | 0.46 | 0.57* | 0.66* | 0.88* | -0.47 | -0.38 | -0.42 | -0.41 | 0.05 |

* 5% 수준에서 통계적으로 유의

패널 A에서는 종목별 게시물 건수가 가장 많은 날을 사건일로 하고, 패널 B에서는 종목별 게시물에 대한 조회건수가 가장 많은 날을 사건일로 하여 주가와 의 관계를 보여준다. 패널 A를 살펴보면, 게시물 건수가 가장 많은 사건일 이전까지는 양의 초과수익률을 보였으나 사건일 당일에는 주가가 하락 반전하였다. 이후에는 지속적인 음의 초과수익률을 보였다. 시장별로 구분한 경우, 코스닥시장 종목의 게시물에 대한 주가반응이 유가증권시장 종목의 게시물에 대한 주가반응 보다 상대적으로 높았으며 하락 반전시의 하락폭도 상대적으로 크게 나타났다.

조회건수에 대한 주가 반응은 게시물 건수에 대한 주가 반응과 유사한 패턴을 보이지만, 사건일 당일에는 큰 차이를 보인다. 게시물 건수의 경우, 사건일 당일에 하락 반전하는 양상을 보였으나, 조회건수의 경우에는 사건일 당일까지 양의 초과수익률을 보이고 사건일 익일부터 하락 반전하는 양상을 보였다. 이러한 결과는 게시물의 건수나 게시물에 대한 조회건수의 증가가 동 종목들에 대한 투자자들의 관심이 커지고

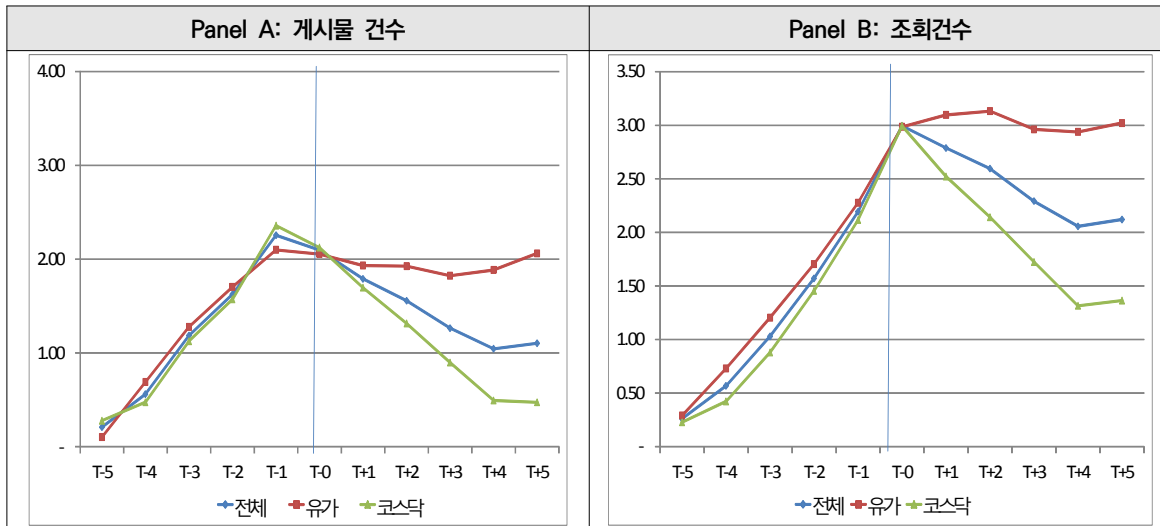
있음을 나타내는 지표로 해석할 수 있고, 투자자들의 관심 증가와 주가간에 양의 관계가 있음을 보여주고 있다. 투자자들은 관심을 끄는 종목에 대해 매매를 집중한다는 Barber and Odean(2008) 등의 연구들은 관심과 매매간 직·간접적인 관계를 증명하고 있지만, 본 연구는 투자자의 관심이 증가함에 따라 주가도 지속적으로 상승하고 있다는 것을 보여주고 있다.⁶⁾

〈그림 1〉은 게시물 건수 및 조회건수에 대한 주가 반응을 누적 초과수익률로 나타낸 것이다. 게시물 건수의 증가에 따라 누적 초과수익률도 지속적으로 증가하다가 사건일 전일을 고점으로 하락 반전하는 양상을 보였다. 유가증권시장에 속한 종목의 경우, 고점 이후 하락폭이 그다지 크지 않은 반면, 코스닥시장에 속한 종목의 경우, 사건일 이후 급격한 하락을 보였다.

게시물에 대한 조회건수 대비 주가 반응을 분석한 결과, 사건일과 누적 초과수익률의 고점이 일치했으며 사건일 이후 하락 반전하는 양상을 보였다. 시장별로 볼 때 유가증권시장에 속한 종목의 경우, 사건일 이후에도 소폭이나마 상승하는 양태를 보였다면 코스닥시장에 속한 종목의 경우 급격한 하락 반전 양상을 보였다

6) 게시물을 게재하는 투자자 또는 게시물을 조회하는 투자자가 동 종목을 매매하는지 검증할 수 없다는 것이 본 연구의 한계이나, 다수 투자자들의 관심이 집합된 게시물 건수와 조회건수가 다수 투자자들의 매매결과인 주가와 의 양의 관계를 확인한 것에서 의미를 찾을 수 있다.

〈그림 1〉 게시물 건수 및 조회건수에 대한 주가반응



는 점은 매우 흥미롭다.

〈표 4〉가 전체 종목을 대상으로 게시물 건수 및 조회 건수와 주가간의 관계를 분석했다면 〈표 5〉는 상승 키워드 및 하락 키워드가 있는 종목만을 대상으로 동일한 분석을 시행한 결과이다. 패널 A에서는 상승 키워드가 있는 종목을 대상으로 게시물이 가장 많았던 일자를 사건일로 하여 분석하고, 패널 B에서는 하락 키워드가 있는 종목을 대상으로 게시물이 가장 많았던 일자를 사건일로 주가를 분석하였다.

패널 A는 상승 키워드가 있는 종목을 분석한 바, 일별 초과수익률도 양의 값을 보였으며 사건일 당일의 초과수익률이 3.72%로 상대적으로 가장 큰 값을 보였다. 사건일 이후, 하락 반전한 주가는 지속적으로 하락하는 양상을 보였다. 전체 종목을 대상으로 분석했던 〈표 4〉에 비해 사건 당일까지 주가는 양의 초과수익률을 보였고, 수익률의 절대값도 상대적으로 더 큰 것을 알 수 있다. 유가증권시장에 속한 종목과 코스닥시장에 속한 종목으로 구분하여 분석한 결과, 초과수익률의 절대 수준은 코스닥에 속한 종목이 더 높았다. 사건일 당일, 유가증권시장에 속한 종목의 경우 평균 초과수익률이 2.54%인 반면 코스닥시장에 속한 종목은 4.30%로 큰 차이를 보였다. 흥미로운

것은 사건일 이후 하락 반전하는 과정에서 유가증권시장에 속한 종목이 더 큰 폭으로 하락하는 것으로 나타났다. 사건 기간인 [t-5, t+5] 동안의 누적 초과수익률은 전체 종목의 경우 7.05%였으며, 유가증권시장에 속한 종목의 경우 4.94%, 코스닥시장에 속한 종목의 경우 8.09%로 나타나 코스닥시장 종목에 대한 주가 반응이 더 큰 것을 확인하였다.

패널 B는 하락 키워드가 있는 종목을 대상으로 게시물 건수가 가장 많은 일자를 사건일로 하여 사건연구 방법에 따른 주가 움직임을 나타낸 것이다. t-5일 이후 일별 초과수익률이 양의 값을 보였으나, 게시물이 가장 많았던 사건일에 하락한 이후 지속적으로 하락하는 양상을 보였다. 유가증권시장에 속한 종목의 사건일 초과 하락율은 -4.36%이고, 코스닥시장에 속한 종목의 사건일 초과 하락율은 -4.41%로 큰 차이를 보이지는 않았다. 다만, 사건일 이후의 하락폭은 코스닥시장에 속한 종목이 상대적으로 큰 것으로 나타났다. 사건 기간([t-5, t+5]) 동안의 누적 초과수익률은 전체 종목의 경우 -3.70%를 보였으며, 유가증권시장에 속한 종목의 경우 -2.71%, 코스닥시장에 속한 종목의 경우 -4.34%로 나타나 코스닥시장 종목의 하락폭이 상대적으로 더 크다는 것을 알 수 있다.



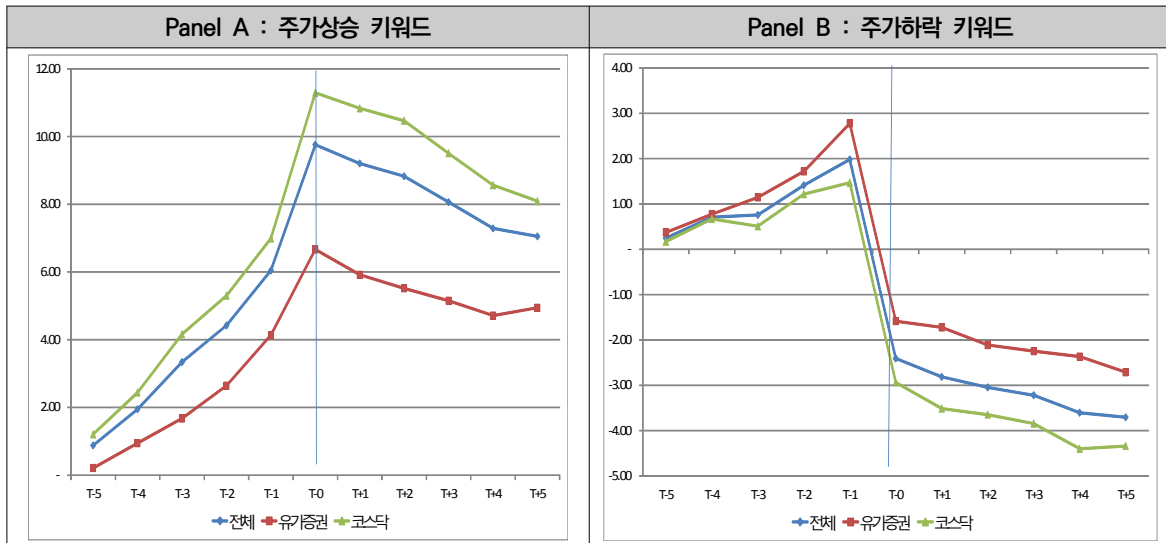
〈표 5〉 게시물과 관련된 주가 움직임

| 수익률 | T-5 | T-4 | T-3 | T-2 | T-1 | T-0 | T+1 | T+2 | T+3 | T+4 | T+5 |
|--------------------|------|------|-------|------|-------|--------|--------|-------|-------|-------|-------|
| Panel A : 주가상승 키워드 | | | | | | | | | | | |
| 전체 | 0.87 | 1.07 | 1.39 | 1.08 | 1.62 | 3.72* | -0.55 | -0.38 | -0.76 | -0.78 | -0.24 |
| 유가증권 | 0.21 | 0.73 | 0.73 | 0.97 | 1.49 | 2.54* | -0.75 | -0.40 | -0.37 | -0.44 | 0.23 |
| 코스닥 | 1.20 | 1.24 | 1.72* | 1.14 | 1.69* | 4.30* | -0.46 | -0.37 | -0.96 | -0.94 | -0.47 |
| Panel B : 주가하락 키워드 | | | | | | | | | | | |
| 전체 | 0.25 | 0.46 | 0.05 | 0.66 | 0.57 | -4.39* | -0.41 | -0.23 | -0.17 | -0.39 | -0.10 |
| 유가증권 | 0.37 | 0.40 | 0.37 | 0.57 | 1.06* | -4.36* | -0.14 | -0.39 | -0.14 | -0.12 | -0.34 |
| 코스닥 | 0.17 | 0.50 | -0.16 | 0.71 | 0.25 | -4.41* | -0.58* | -0.13 | -0.20 | -0.56 | 0.06 |

〈그림 2〉는 〈표 5〉에 나타난 일별 초과수익률을 이용한 누적초과수익률을 그래프로 나타내어 보여주고 있다. 주가상승 키워드인 경우 사건일을 중심으로 지속적으로 상승하는 양태를 보이고 있으며, 코스닥 시장에 속한 종목들의 상승폭이 유가증권시장에 속한 종목들의 상승폭에 비해 높았다. 사건일 이후의 주가 하락폭은 주가 상승폭에 비해 완만하게 이뤄져 게시물

에 따른 추격 매수자들이 존재하고 있음을 알 수 있다. 패널 B의 하락 키워드의 경우, 사건일 전일까지 2% 내외의 누적 초과수익률을 보였으나, 하락키워드가 급등한 사건 당일에 누적 초과수익률이 급격히 하락하는 것을 알 수 있다. 사건일 이후에도 하락 추세는 지속되고 있으나, 하락폭은 사건일에 비해 완만한 양상을 보인다.

〈그림 2〉 누적 초과수익률 추이



일반적으로 사이버 공간의 종목게시판은 주로 개인 투자자들 간의 정보를 공유하는 공간이다. 기관투자자나 외국인투자자의 경우 증권사 애널리스트 등을 통해 정보를 공유할 뿐 일반 개인투자자들과 정보를 공유하는 것은 일반적이지 않다.

주식시장에 참여하는 투자자는 개인투자자, 기관투자자 및 외국인투자자로 구분되며, 실무적으로나 학문적으로 개인투자자를 하나의 집단으로 간주하여 분석하고 평가하고 있다. 그러나, 개인투자자들의 매매내역을 살펴보면, 기관투자자와 유사한 규모의 자금

을 투자하는 개인투자자에서부터 소액 자금 투자자로서 일부 종목들은 1주도 살 수 없기에 저가 종목만을 투자 대상으로 하는 개인투자자에 이르기까지 다양한 것이 현실이다.

이에 <표 6>은 사이버 공간의 정보 활동이 주식시장에 미치는 영향을 분석함에 있어 개인투자자들의 거래 규모에 따른 영향을 보여주고 있다. 거래규모는 주가 수준을 대응치로 이용하였다. 패널 A에 따르면, 1만원 미만 저가 종목과 100만원 이상 초고가 종목의 경우, 사건일에 양의 초과수익률을 보인 반면 1만원 이상 100만원 미만인 종목의 경우에는 사건일 전일까지

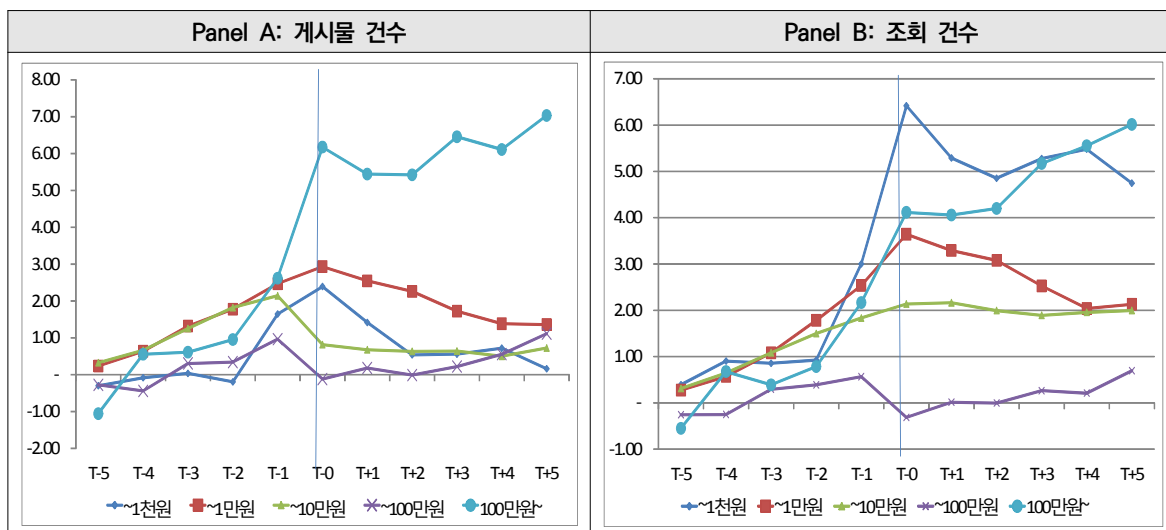
보였던 양의 초과수익률이 사건일에 음의 초과수익률로 하락 반전하였다.

패널 B의 조회건수에 대한 주가 반응의 경우, 1만원 미만 저가 종목에서 조회건수가 최대치인 사건일에 양의 수익률이 나타나고, 특히 1천원 미만 초저가 종목에서 그 반응이 가장 강하게 나타났다. Kumar and Lee(2006), Baker and Wurgler(2006) 등은 소규모 종목들의 경우 개인투자자들에 의한 투자가 많기 때문에 개인의 감성에 더 많은 영향을 받는다고 주장하였다. 본 연구의 결과는 이러한 기존 연구들의 결과를 지지한다고 할 수 있다.

<표 6> 가격대에 따른 주가 반응

| 수익률 | T-5 | T-4 | T-3 | T-2 | T-1 | T-0 | T+1 | T+2 | T+3 | T+4 | T+5 |
|------------------|-------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|
| Panel A : 계시물 건수 | | | | | | | | | | | |
| ~1천원 | -0.30 | 0.22 | 0.12 | -0.23 | 1.84* | 0.75* | -0.97* | -0.88 | 0.02 | 0.16 | -0.56 |
| ~1만원 | 0.23 | 0.40 | 0.68* | 0.46 | 0.69* | 0.46* | -0.39 | -0.29 | -0.54 | -0.34 | -0.02 |
| ~10만원 | 0.32 | 0.34 | 0.59 | 0.57 | 0.32 | -1.33 | -0.14 | -0.04 | 0.01 | -0.14 | 0.22 |
| ~100만원 | -0.28 | -0.16 | 0.75* | 0.04 | 0.62* | -1.08 | 0.30 | -0.19 | 0.23 | 0.34 | 0.56 |
| 100만원~ | -1.06 | 1.61* | 0.06 | 0.34 | 1.66* | 3.56* | -0.73 | -0.02 | 1.03 | -0.34 | 0.92 |
| Panel B : 조회 건수 | | | | | | | | | | | |
| ~1천원 | 0.39 | 0.51 | -0.05 | 0.07 | 2.08* | 3.41* | -1.13 | -0.44 | 0.43 | 0.20 | -0.74 |
| ~1만원 | 0.27 | 0.29 | 0.51 | 0.70* | 0.75* | 1.11* | -0.35 | -0.21 | -0.55 | -0.49 | 0.09 |
| ~10만원 | 0.31 | 0.33 | 0.44* | 0.41* | 0.34* | 0.30 | 0.03 | -0.17 | -0.10 | 0.06 | 0.04 |
| ~100만원 | -0.26 | 0.00 | 0.55* | 0.09 | 0.18 | -0.88 | 0.33* | -0.02 | 0.27 | -0.06 | 0.48 |
| 100만원~ | -0.55 | 1.23 | -0.29 | 0.39* | 1.38* | 1.95* | -0.06 | 0.14 | 0.97* | 0.38 | 0.46 |

<그림 3> 가격대별 누적 초과수익률 추이





〈그림 3〉은 게시물 건수 및 조회건수에 대한 주가 반응을 가격대별로 분석한 〈표 6〉의 결과를 근거로, 누적 초과수익률을 그래프로 나타낸 결과를 보여주고 있다. 패널 A를 통해 게시물 건수의 영향을 보면, 100만원 이상인 종목은 사건일 이후에도 지속적으로 누적초과수익률이 상승하고 있으며, 1만원~100만원 미만인 종목은 사건일 전에 주가가 하락하고 있었다.

특히, 1천원 이상 1만원 미만인 저가 종목 및 1천원 미만의 초저가 종목의 경우, 사건일을 기준으로 주가의 상승과 하락 양상이 정확히 대비되는 모습을 보이고

있다. 이러한 주가 반응은 특히 1천원 미만의 초저가 종목에서 강하게 나타난다. 이는 미국 주식 시장을 대상으로 스캠 메일의 영향을 분석, 저가 종목에서 사건일 이전의 주가의 상승 및 하락 반전 모습이 강하게 나타남을 발견한 Hanke and Hauser(2008)의 연구 결과와 일치한다. 패널 B의 조회건수에 대한 주가 반응의 경우도, 1천원 미만 초저가 및 1천원 이상 1만원 미만 저가 종목에서 사건 당일의 급격한 수익률 증가와 그 이후의 뚜렷한 수익률 반전 현상이 나타나고 있다.

[4] 결론 및 시사점

본 논문은 증권 게시판의 정보 활동이 주식 시장에 미치는 영향에 대하여 살펴보았다. 분석을 위하여 2013년 9월부터 2015년 6월까지 국내 주요 증권 게시판의 게시물 건수 및 조회건수, 긍정성 게시물 및 부정성 게시물을 머신 러닝 기법인 SVM이라는 인공지능 알고리즘을 사용하여 파악한 후, 사이버 정보를 통한 정보의 수령 및 생성 행위가 주가와 어떠한 관계가 있는지 살펴보았다.

주된 분석 결과는 다음과 같다. 첫째, 증권관련 사이트의 게시물은 주가에 통계적으로 유의한 영향을 주는 것이 확인되었다. 게시물 건수의 증가와 조회건수의 증가는 주가 상승을 동반하였다. 둘째, 특히 주가에 긍정적인 키워드 혹은 부정적인 키워드가 포함된 게시물의 건수는 주가에 더욱 영향을 주는 것으로 나타났다. 하락 키워드가 상승 키워드보다 더 큰 영향을 보였으며, 주요 하락 키워드는 지속적인 주가 하락에 영향을 주는 것으로 나타났다. 셋째, 유가증권 시장 종목보다 코스닥시장 종목에서 게시물 건수 및 조회건수가 주가에 미치는 영향이 더 크게 나타났다. 이는 코스닥시장에서 증권 게시판을 이용한 부정거래 및 시세조종을 할 개연성이 더 크게 존재한다는 것을 시사한다. 넷째, 주식 종목의 가격대에 따라서 증권계

시판의 키워드가 주가에 미치는 영향이 다르게 나타났다. 특히, 개인투자자 비중이 많은 1천원~1만원 이하 저가 종목 및 1천원 미만의 초저가 종목에서 게시물 건수 및 조회건수와 주가의 관계가 더욱 뚜렷하게 나타났다. 아울러, 100만원 이상의 초고가 종목에서는 사건일 전후의 주가 모습이 저가 종목과는 다르게 나타남으로써, 개인투자자 중에서 사이버 정보를 주된 정보원으로 이용하는 투자자 그룹과 증권사 애널리스트 보고서 등 유료의 다른 정보원도 활용할 가능성이 큰 투자자 그룹의 차이 간에 있어서 증권 게시판의 영향력 차이를 간접적으로 보여주고 있다.

4차 산업혁명 시대에 진입하면서, 정보가 주가에 반영되는 통로도 다변화되고 있으며, 기존에 정보수령자로 간주되던 개인투자자 그룹은 정보 수령과 동시에 사이버 공간에서 정보를 생성하는 그룹으로 변화하고 있다.

본 연구는 사이버 공간에서의 개인들에 의해 생성되고 공유되는 정보가 주식 시장에 어떠한 영향을 미치는지를 살펴보았다는 점에서 그 의의를 갖는다. 또한, 사이버 정보의 게시물 건수 및 조회건수가 주가에 영향을 주고 있다는 발견은 급속히 발전하고 있는 모바일 거래 환경에서 증권 게시판을 이용한 부정거래

및 시세조종의 개연성이 상존하고 있음을 시사하여 준다.

따라서, 본 연구의 결과는 증권 게시판을 이용한 주식 가격의 왜곡 현상을 방지하는 제도적 필요성을 인지하고 이러한 행위의 예방·감시 측면에서 시사하

는 바가 크다고 하겠다. 아울러, 정보화 시대를 넘어 유비쿼터스 시대를 맞이한 스마트 자본시장의 투자자들은 사이버 공간을 통해 정보를 소통·공유하는 활동 자체가 주식 시장에 영향을 미칠 수 있다는 점을 인지하여야 할 것이다.

참고 문헌

우민철, 이우백, 2013, 모바일 주식 거래의 행태와 수익성 분석, *한국증권학회지* 42, pp. 133-161.

Antweiler, W. A. and M. Z. Frank, 2004, Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, *Journal of Finance* 59, pp. 1259-1294.

Baker, M. and J. Wurgler, 2006, Investor Sentiment and the Cross-section of Stock Returns, *Journal of Finance* 61, pp. 1645-1680.

Barber, B. M., and T. Odean, 2002, Online Investors: Do the Slow Do First? *Review of Financial Studies* 15, pp. 455-488.

Barber, B. M. and T. Odean, 2008, All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, *Review of Financial Studies* 21, pp. 785-818.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang, 2014, Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media, *Review of Financial Studies* 27, pp. 1367-1403.

Das, S. R. and M. Y. Chen, 2007, Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web, *Management Science* 53, pp. 1375-1388.

Datamonitor, 2010, Social media in financial services: The customer as the advisor.

Deloitte, 2007, Most consumers read and rely on online reviews; companies must adjust.

Frieder, L. and J. Zittrain, 2007, Spam Works: Evidence from Stock Touts and Corresponding Market Activity, *Hastings Communications and Entertainment Law Journal* 30, pp. 479.

Gilbet, E. and K. Karahalios, 2010, Widespread Worry and the Stock Market, working paper, Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC.

Hanke, M. and F. Hauser, 2008, On the Effects of Stock Spam E-mails, *Journal of Financial Markets* 11, pp. 57-83.

Harris, M. and A. Raviv, 1993, Differences of Opinion Make a Horse Race, *Review of Financial Studies* 6, pp. 473-506.

Joachims, T., 1999, Making Large-Scale SVM Learning Practical. In B. Scholkopf, C. Burges, A. Smola (Eds.), *Advances in Kernel Methods*, The MIT Press.

Kelley, E. K. and P. C. Tetlock, 2013, How Wise Are Crowds? Insights from Retail Orders and Stock Returns, *Journal of Finance* 68, pp. 1229-1265.

Tetlock, P. C., 2007, Giving Content to Investor Sentiment: The Role of Media in the



Stock Market, *Journal of Finance* 62, pp. 1139-1168.

Tumarkin, R. and R. F. Whitelaw, 2001, News or Noise? Internet Postings and Stock Prices, *Financial Analysts Journal* 57, pp. 41-51.

Vapnik, V., 1995, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York.

Vapnik, V., A. Chervonenkis, 1964, On the

Uniform Convergence of Relative Frequencies of Events to Their Probabilities, *Theory Probability and Its Application* 16, pp. 264-280.

Wysocki, P. D., 1999, Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards, Working Paper, November, University of Michigan.

Can the Cyber-Information Influence the Stock Price? : Using an Artificial Intelligent(AI) Algorithm Methodology

Min-Cheol Woo* (KRX)

Jee-Hyun Kim** (Hallym University)

Abstract

This paper investigates whether the information in internet stock message boards can influence stock prices. Using the SVM(Support Vector Machine), which is one of artificial intelligent algorithm methodologies, we collect 7,054,316 messages posted on the three major internet stock message boards- Never, Daum, and Paxnet- for the period from September 2013 to June 2015. The total number of posting reads (hits) is 1,662,840,007. The major findings of this paper are as follows. First, we find significant evidence that the number of postings and the number of hits increase stock prices. Second, postings with bullish or bearish key-words have more significant effect on the stock prices. Third, these findings are more prominent with the KOSDAQ-listed stocks, suggesting that the possibility of unfair trading based on posting manipulation is higher with KOSDAQ-listed stocks. Forth, we find that the influence of the number of postings and hits are different according to stock price ranges, indicating that there exists various individual investors with respect to the use of messages on internet stock message boards.

Key words : *Cyber Information, Internet Stock Message Boards, SVM (Support Vector Machine), Individual Investors, Information Provider*

Article history : Received 10 October 2017, Revised 16 November 2017, Accepted 1 December 2017

* The first author, KRX

** Corresponding author. Hallym University (E-mail: jhyunkim@hallym.ac.kr)