

웹 뉴스의 양과 주가의 관계에 관한 연구*

김상수** · 남달우*** · 조 현**** · 김성희*****

A Study on the Relation of Web News and Stock Price*

Sang Soo Kim** · Dal-Woo Nam*** · Hyeon Jo**** · Soung Hie Kim*****

■ Abstract ■

In the stock market, the investors rely on stock information to trade. Good information may stimulate buying, raising the stock prices and the bad information may result in selling, decreasing the stock prices. In terms of the relationship between information and stock prices, stock prices can be viewed as reaction of investors to all the information flowing into the market. The significant increase of web stock news volume is often associated with the significant changes of stock prices. When the web stock news volume for a firm increases significantly, the stock price movement is often oscillatory. This paper attempts to investigate the relationship between volumes of information from Korean web IT and stock prices in Korean stock market. This research shows that when the web stock news volume increases significantly, volatility, trading volumes and rate of returns are increase too. The results of the study provide us with the new clues to the microstructure of the stock market from the perspective of the web news.

Keyword : Web Stock News, News Volume, Stock Price, Volatility, Trading Volumes, Rate of Returns

논문투고일 : 2012년 07월 27일 논문수정완료일 : 2012년 09월 10일 논문게재확정일 : 2012년 09월 17일

* 이 논문은 주저자의 석사학위 논문에 근거하여 작성되었음.

** 한국과학기술원 경영대학 공공혁신 전자정부 연구센터 선임연구원

*** 한국과학기술원 경영대학 박사과정

**** 한국과학기술원 경영대학 테크노 경영 연구소 박사 후 연구원, 교신저자

***** 한국과학기술원 경영대학 교수

1. 서 론

오늘 날과 같이 정보가 넘쳐나는 시대에 주가는 수많은 요인에 영향을 받는다. Fama[9]는 현재 주가는 과거의 정보를 모두 반영하고 있다는 효율적 시장가설을 주장하였고, Ross[22]는 주가가 시장에 반영된 정보흐름에 따라 변한다고 했다. 이들의 주장에 의하면 시장의 모든 정보는 주가에 영향을 준다고 할 수 있다.

따라서 투자자들은 주가에 영향을 줄만한 정보를 찾기 위해 많은 노력을 들인다. 투자자들이 주변에서 쉽게 정보를 얻을 수 있는 수단으로는 TV, 라디오, 신문, 대인관계 등을 있지만, 인터넷이 발달한 지금 많은 사람들은 가장 쉽게 접할 수 있는 포털 사이트의 웹 뉴스, 인터넷 주식게시판에서의 개인들의 의견, 기업들의 전자 공시 등을 많이 참고한다[2]. 순장부거나 순이익과 같은 기업의 다양한 회계 정보를 활용하여 의사결정을 수행하기도 한다[1]. 또한 이러한 웹 정보기술 서비스를 활용하면서 이용자들 기업 및 주식에 대한 정보를 습득하고, 이를 활용하여 투자 의사결정을 내리게 된다.

본 연구에서는 각 기업의 웹 뉴스의 양과 주가를 일일 단위로 측정하여 웹 뉴스량과 주가와 관계에 대해 알아본다. 특히 웹 뉴스의 양을 측정하는 과정에서 인터넷 정보기술의 전형적인 유형인 포털 서비스를 이용하고자 한다. 또한 주가 관련 변수 중 변동성, 거래량, 수익률에 초점을 맞추어 웹 뉴스의 양과의 관계를 살펴본다. 특히 본 연구에서는 주가를 설명하기 위해 재무나 금융 분야에서 대표적으로 사용해왔던 자기 상관 방법(AR)이나 이동 평균법(MA)등의 시계열 방법인 아닌 웹 뉴스를 활용하였기 때문에 학술적인 차별점이 있으며, 분석 대상이 일반 투자자들이 쉽게 접할 수 있는 웹 뉴스라는 점에서 실무적인 독창성을 지닌다.

웹 뉴스라고 해서 모두 주가를 변화시키는 것이 아닐 것이기 때문에, 뉴스를 주가 변화를 가져올 수 있는 뉴스와 주가 변화를 가져 올 수 없는 뉴스로 구분한다. 흔히 경제학에서 주장하는 것처럼 인간

을 모두 이성적이라고 가정하면, 주가에 영향을 줄 수 있는 정보를 가진 뉴스만이 실제로 주가를 변동시킬 것이지만, 현실은 그렇지 않기에 주가와 무관한 뉴스들도 주가에 영향을 끼쳐 주가의 변동성을 크게 할 수 있다. 비합리적인 투자자들이 존재한다는 가정 하에 뉴스가 담고 있는 정보의 의미와는 별개로, 웹 뉴스의 양이 증가함에 따라 주가의 변동성, 거래량, 수익률이 변할 것이라고 예상할 수 있다.

본 논문의 구성은 다음과 같다. 제 2장에서는 관련 연구를 고찰하고, 제 3장에서는 연구가설, 자료수집, 분석 방법을 포함하는 전체적인 연구 방법을 제시한다. 제 4장에서는 통계적 분석 결과 및 가설 검증 내용을 담았으며, 제 5장에서는 결론을 제시한다.

2. 선행 연구

2.1 이론적 배경

2.1.1 효율적 시장가설

Fama[9]는 효율적 시장가설을 통해 자본시장의 가격이 현재 시점에서 이용 가능한 정보를 충분히, 즉각적으로 반영하고 있다고 주장했다. Fama는 효율적 시장가설을 관련되는 정보의 범위에 따라 약형 효율적 가설시장, 준강형 효율적 가설시장, 강형 효율적 가설시장으로 구분하고 있다. 약형 효율적 가설시장에서는 어떤 투자자도 과거의 정보에 기초한 거래에 의하여 초과 수익을 얻을 수 없다. 준강형 효율적 시장가설에서는 어떤 투자자도 공개적으로 이용 가능한 정보를 기초로 한 거래에 의하여 지속적으로 시장 초과 수익률을 달성할 수 없다. 마지막으로 강형 효율적 시장가설에서는 어떤 투자자도 모든 공식적으로 이용 가능한 정보뿐만 아니라 내부정보를 사용함으로써 초과 수익을 달성할 수 없다.

본 논문에서는 효율적 시장가설에서 이야기하는 '효율적'인 시장에서의 정보적 효율성 측면을 이용

하여 인터넷 뉴스가 주가 변동에 반영된다고 가정한다.

2.1.2 행동 재무론

실제로는 과거 정보를 이용해 초과수익을 얻을 수도 있는데, 이는 효율적 시장가설에 정면으로 배치되는 현상이다. 앞서 말한 효율적 시장 가설을 비롯한 전통적인 재무이론들은 실제 투자자들의 행동과 그로인한 수익률과는 차이가 있다. 행동 재무론은 이러한 투자자들의 실제 행동에 대한 의문에서 시작되었는데, 많은 행동 재무론 학자들은 효율적 시장가설과 반대되는 이상 현상들을 비합리적인 결정을 내리는 현실 세계의 투자자들의 행동을 이용해 설명하였다. 투자자들의 실제 행동이 비합리적인 이유는 정보 처리 측면의 비합리성 때문이다. 즉, 투자자들이 언제나 올바르게 정보를 처리하지는 않는다. 그 예로 Kahneman and Tversky [14]는 위험 상황에서의 의사 결정에 대한 한 연구에서 사람들이 미래를 예측할 때, 최근의 경험에 크게 의존하며 불확실성 하에서는 극단적인 결론을 내리는 경향이 크다는 것을 밝혀냈다. 몇몇 연구에서는 투자자들이 새로운 정보가 나타났을 때 그들의 과거 믿음을 수정하는데 느리다는 것을 밝혀냈으며[5, 12], 새로운 정보가 과거의 정보를 확인해줄 때 과거 정보에 대한 과잉반응이 나타남을 보인 연구도 있었다[8]. 또한 투자자들이 정보를 합리적으로 처리했다고 하더라도 비합리적인 결정을 내릴 수 있다. 즉, 이는 행동적 편의(behavioral bias)를 의미하며 투자자들이 정보를 완벽하고 정확하게 처리했지만, 실제 행동으로 옮기는 과정에서 비이성적일 수 있다는 것을 의미한다. 일부 연구에서는 투자자들이 손실이 난 주식을 오래 보유하고 이익이 난 주식을 일찍 팔아버리는 경향이 있음을 보였다[19, 23]. Thaler and Johnson[25]의 연구에서는 주식시장이 일정기간 상승했을 때 투자자의 자본 이득이 증가하여 이 소득에 대하여는 잃게 되더라도 아깝지 않으므로 위험한 투자를 반복해서 계속할 수 있음을 보였다.

본 논문에서는 행동 재무론을 통해 투자자들이 이용 가능한 정보를 주가에 반영할 때, 항상 이성적이고 정확한 결정을 내리는 것이 아니라고 가정하였다.

2.2 뉴스를 활용한 주가 연구

2.2.1 뉴스와 주가

주식 뉴스와 주가와와의 관계에 대한 연구는 오랜 기간 진행되었고, 대부분 특별한 뉴스에 대한 주식 시장의 반응에 중점을 두고 있다. 주가에 영향을 미칠 수 있는 대표적인 뉴스는 경제지표 발표 소식이 있다. 무역수지 발표가 주가에 미치는 영향에 대한 실증분석을 시도한 연구들이 있었는데 Hardouvelis[11]는 1980년대 전반 미국의 무역수지 발표가 S&P500 지수에 유의한 영향을 미친다는 증거를 발견하지 못하였으나 Kretzmer[16]는 1980년대 후반에는 주가에 대한 유의한 효과가 존재하였음을 밝혔다. Sun and Tong[24]의 연구에 따르면 미국의 무역수지 발표가 일본의 자동차 회사 주가에는 영향을 미치지 않지만 미국의 자동차 회사의 주가에는 영향을 미치지 않았다. 또한 경제지표 이외에도 정치, 사회적 사건의 발생 등 다양한 종류의 뉴스가 주가에 영향을 미칠 수 있다. Kamin-sky and Schmukler[15]는 1990년대 후반 동아시아의 위기기간 중 동아시아 국가들의 주식수익률이 다른 금융 위기국가에서 발생한 해외 뉴스에 민감하게 반응했다는 것을 보였다. Baig and Goldfajn [3]는 실증분석을 통해 한국, 태국, 인도네시아, 필리핀, 말레이시아 등 5개국의 주가변화가 금융위기 기간 중 해외에서 발생한 새로운 뉴스에 영향을 받았음을 밝혔다. Ganapolsky and Schmukler [10]의 연구에 따르면 멕시코 금융위기 기간 중 아르헨티나 정부의 통화위원회제도 유지, 지급 준비율 변화, IMF와의 합의, 새로운 재무장관 지명 등의 뉴스가 아르헨티나의 주가를 안정시켰다.

2.2.2 뉴스량과 주가

정보의 내용 측면에서는 활발한 연구와는 달리 정

보의 양 측면에서는 거의 연구가 이루어지지 않았다. 정보의 내용 측면에서의 연구는 각 정보의 의미를 분석하는 과정이 반드시 필요하다. 특정 정보가 담고 있는 의미가 긍정적인지 부정적인지에 따라 주가에 미칠 수 있는 영향이 다르기 때문이다. 따라서 대부분의 연구에서는 특정 사건을 임의로 선정하여 그 사건이 주가에 미치는 영향을 주로 분석해왔다. 그러나 이러한 접근 방식의 연구 결과는 미래의 투자에 적용하기에는 어려움이 있다.

과거 다수의 연구에서 정보의 양과 주가와와의 관계를 파악하는 과정에서 정보의 양을 직접 측정하기가 어려워 거래량을 정보의 양으로 정의하였다 [6, 13, 17, 21]. 그러나 정보의 양과 거래량은 동일하게 움직이지 않으며 서로 다른 특성을 가진 변수이다. 이는 웹을 통한 정보의 획득이 쉽지 않았던 과거 연구에서의 임시적인 선택으로 판단된다[18].

정보의 내용이 아닌 정보의 양적인 측면에서의 연구의 대표적인 예는 Wysocki[26]의 웹 주식 메시지의 양과 주가와와의 연관성에 대한 연구가 있다. 그의 연구에서는 웹 주식 메시지의 증가가 당일 주가 및 전일 주가 변화와 관련이 있음을 입증되었다. Liang[18]는 Wysocki의 연구를 발전시키면서 주식 메시지에 비해 주식 뉴스는 투자자에게 더 큰 영향을 미친다고 판단하여 웹 주식 뉴스의 양과 주가와와의 관계를 밝혔다. 추가적으로 정보기술의 전형적인 유형인 검색엔진의 검색량과 주가의 관계를 규명하는 연구들이 있어왔다[4, 7, 20].

본 논문에서는 일련의 선행연구들 바탕으로 한국에서의 웹 정보기술을 통해 수집할 수 있는 개별 종목 웹 뉴스의 양과 주가의 관계에 대해 알아본다. 또한 기존의 연구를 확장하여 웹 뉴스의 양은 변동성뿐만 아니라 각 종목의 거래량, 수익률과도 관계를 살펴 볼 것이다.

3. 연구 가설

3.1 웹 뉴스의 양과 주가

특정 기업에 대한 웹 뉴스의 양에 큰 변화가 없

을 때 해당 종목의 주가에도 큰 변화가 없으며, 웹 뉴스의 양이 증가했을 경우 해당 종목의 주가가 큰 변화를 보이는 경우를 쉽게 볼 수 있다. 효율적 시장 가설의 정보 측면의 효율성에 따라 모든 웹 뉴스는 주가에 반영되어야 한다. 이에 투자자들이 언제나 이성적이고 합리적인 결정을 내리지는 않는다는 행동 재무론을 더하여 주가와 직접적인 관련이 없는 웹 뉴스라고 할지라도 투자자들이 이에 반응하여 주가에 영향을 미칠 수 있다고 추론할 수 있다.

뉴스는 크게 직접 뉴스와 간접 뉴스로 나눌 수 있다. 직접 뉴스는 특정 기업에 대한 정보를 담고 있는 뉴스를 의미하고, 간접 뉴스는 타 기업의 뉴스의 내용 중 언급된 뉴스를 의미한다[15]. 각 기업의 직접 뉴스는 각 기업에 대한 정보를 담고 있으므로 주가에 직접적인 영향을 미쳐 긍정적이거나 부정적인 효과를 낼 것이다. 또한 각 기업의 간접 뉴스는 그 기업에 대한 정보를 담고 있는 것은 아니지만 동종 업계 또는 연관 기업에 대한 정보를 담고 있으므로 주가에 미치는 영향이 있을 수 있다.

웹 뉴스의 영향을 받은 투자자들의 거래 활동을 통해 주가가 변화할 때 이 변화를 측정할 수 있는 지표 중 대표적인 변수인 변동성과 거래량을 토대로 가설 1-1과 가설 1-2를 설정하였다.

가설 1-1 : 일일 웹 직접 뉴스의 양, 일일 웹 간접 뉴스의 양은 일일 주가 변동성에 영향을 미칠 것이다.

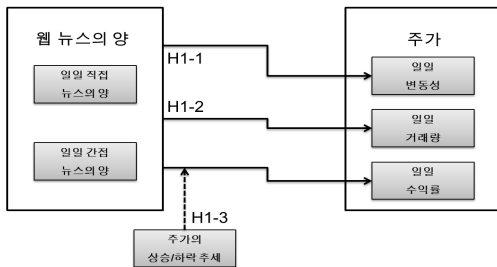
가설 1-2 : 일일 웹 직접 뉴스의 양, 일일 웹 간접 뉴스의 양은 일일 거래량에 영향을 미칠 것이다.

변동성과 거래량에 비해 수익률은 투자자들이 보다 민감하게 생각하는 지표이다. 웹 뉴스의 양은 항상 양의 값을 갖게 되고, 변동성과 거래량 또한 항상 양의 값이므로 쉽게 그 관계를 파악할 수 있다. 그러나 수익률은 경우에 따라 양 또는 음의 값을 갖게 되므로 웹 뉴스의 양과의 직접적인 관

계를 파악하기에는 어려움이 있다. 따라서 웹 뉴스의 양과 더불어 주가의 상승/하락 추세를 고려하여 수익률과의 관계를 파악해보고자 하였다. 즉, 주가가 상승 추세에 있을 경우와 주가가 하락 추세에 있을 경우를 구분하여 웹 뉴스의 양과 수익률과의 관계를 파악하기 위한 가설을 설정하였다.

가설 1-3 : 일일 웹 직접 뉴스의 양, 일일 웹 간접 뉴스의 양, 주가의 상승/하락 추세는 일일 주가 수익률에 영향을 미칠 것이다.

위의 가설 1-1~가설 1-3을 요약하면 [그림 1]과 같다.



[그림 1] 연구 가설 1

3.2 웹 뉴스의 양의 극단적인 변화와 주가

웹 뉴스의 양의 유의한 증가는 특별한 사건이 특정 기업에 발생하였음을 의미하고 이는 그러한 특정 사건에 대한 주가의 변화로 나타나게 된다[18].

웹 뉴스의 양과 주가 정보의 관계에 대한 가설 1을 더욱 확대하여 웹 뉴스의 양이 극단적으로 변화했을 때의 상황을 살펴보고자 한다. 기업과 관련된 웹 뉴스의 양이 일일 평균 웹 뉴스의 양과 비교하여 유의하게 증가하였을 경우, 해당 종목 주가의 변동성과 거래량도 일일 평균 주가 변동성과 일일 평균 거래량에 비해 유의하게 증가할 것이라는 추론으로 다음과 같은 가설을 설정하였다.

가설 2-1 : 일일 웹 뉴스의 양이 유의하게 증가한 날에는 그렇지 않은 날보다 일일 주가

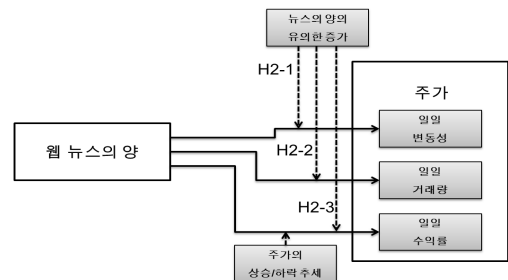
변동성이 더 클 것이다.

가설 2-2 : 일일 웹 뉴스의 양이 유의하게 증가한 날에는 그렇지 않은 날보다 일일 거래량이 더 클 것이다.

수익률의 경우 첫 번째 가설에서와 마찬가지로 주가가 상승 추세에 있을 경우와 주가가 하락 추세에 있을 경우를 구분하여 실험하였다. 기업과 관련된 웹 뉴스의 양이 일일 평균 웹 뉴스의 양과 비교하여 유의하게 증가하였을 경우, 해당 종목 주가의 수익률도 일일 평균 주가 수익률에 비해 유의하게 증가하거나(상승 추세인 경우) 감소할 것(하락 추세인 경우)이라는 추론으로 다음과 같은 가설을 설정하였다.

가설 2-3 : 일일 웹 뉴스의 양이 유의하게 증가한 날에는 그렇지 않은 날보다 일일 주가 수익률이 더 크게 변화할 것이다.

위의 가설 2-1 : 가설 2-3을 요약하면 [그림 2]와 같다.



[그림 2] 연구 가설 2

4. 자료 수집 및 분석 방법

4.1 자료 수집

4.1.1 웹 뉴스의 양

본 논문의 웹 주식 뉴스는 전형적인 인터넷 정보기술 서비스인 포털 네이버 뉴스(<http://news.na>)

ver.com) 서비스를 통해 각 기업의 명칭을 검색했을 때의 결과를 날짜별로 수집하였다. 네이버의 뉴스 중 제목에 기업의 명칭이 포함된 뉴스를 직접 뉴스, 제목에는 기업의 명칭이 없으나 본문 내용 중 기업의 명칭이 포함된 뉴스를 간접 뉴스로 처리하였다. 즉, 다른 기업의 직접 뉴스의 본문에 포함된 뉴스를 간접 뉴스로 정의하였다.

4.1.2 주가 정보

표본이 될 종목을 선정하는 데에는 KOSPI를 대표하는 50개의 종목을 의미하는 KOSPI50 지수와 KOSDAQ을 대표하는 30개 종목을 의미하는 KOSDAQ STAR INDEX를 기준으로 하였다. 데이터 처리 과정에서 신규 상장 또는 거래정지 등의 사유로 인하여 1년간의 표본 기간 중 주가가 하락된 기업들은 제외하였다. 그 결과 KOSPI에서는 KOSPI50에 포함된 50개의 종목 중 49종목, KOSDAQ에서는 KOSDAQ STAR INDEX에 포함된 30개의 종목 중 27종목, 총 76종목이 선택되었다.

4.2 분석 방법

4.2.1 웹 뉴스의 양

첫 번째 가설을 검증하기 위한 변수인 각 기업의 일별 직접, 간접 웹 뉴스의 양을 76개의 기업 249 거래일을 대상으로 정의하였다($D_{i,t}$ = 기업 i 의 t 시점에서의 직접뉴스, $U_{i,t}$ = 간접뉴스). 기업의 규모에 따라 일별 직접, 간접 웹 뉴스의 양의 크기의 편차가 심하기 때문에 이를 정규화 해야 할 필요가 있었다. 이를 위해 각 기업의 1년 동안의 직접, 간접 웹 뉴스의 양의 평균과 표준편차를 구하였다. 일별 직접, 간접 웹 뉴스의 양은 서로 독립적이고 랜덤하다고 가정하였고, 이를 토대로 각 기업의 직접, 간접 웹 뉴스의 양을 정규분포를 가정하고 점수화하였다. 각 기업의 일별 직접, 간접 웹 뉴스양의 점수는 다음과 같다.

$$Z \sim N(0, 1^2) \quad (1)$$

$$DScore_{i,t} = P\left(Z < \frac{D_{i,t} - \mu_{d,i}}{\sigma_{d,i}}\right),$$

$$(i = 1, \dots, 76 \quad t = 1, \dots, 249)$$

$$UScore_{i,t} = P\left(Z < \frac{U_{i,t} - \mu_{u,i}}{\sigma_{u,i}}\right),$$

$$(i = 1, \dots, 76 \quad t = 1, \dots, 249)$$

$\mu_{d,i}(\mu_{u,i})$ 와 $\sigma_{d,i}(\sigma_{u,i})$ 는 직접(간접)뉴스의 평균과 표준 편차를 의미한다. 이와 같은 정규화와 점수화 과정을 거쳐 DScore와 UScore는 각각 249일의 표본 기간과 76개의 종목을 곱한 18,924개의 데이터를 얻었다.

두 번째 가설에서는 직접 웹 뉴스와 간접 웹 뉴스의 양을 가중 평균한 웹 뉴스의 양을 변수로 사용하였다. 각 기업의 일별 웹 뉴스의 양의 정의는 다음과 같다.

$$N_{i,t} = D_{i,t} \times \alpha + U_{i,t} \times \beta, \quad (2)$$

$$(\alpha = 1, \beta = 0.1, i = 1, \dots, 76 \quad t = 1, \dots, 249)$$

기존 연구를 참고하여 α 와 β 의 값은 각각 1과 0.1을 사용하였다. 즉, 기업에 큰 영향을 미칠 수 있는 직접 뉴스의 양의 비중과 타 기업과의 관계에 따라 달라질 수 있는 간접 뉴스의 양의 비중이 가중 평균 된 웹 뉴스의 양을 정의하였다[15].

기업에 따라 일별 웹 뉴스의 양의 크기의 편차가 심하기 때문에 첫 번째 가설에서와 마찬가지로 이를 정규화 해야 할 필요가 있었다. 이를 위해 각 기업의 1년 동안의 웹 뉴스의 양의 평균과 표준편차를 구하였다.

일별 웹 뉴스의 양은 서로 독립적이고 랜덤하다고 가정하였고, 이를 토대로 각 기업의 웹 뉴스의 양을 정규분포를 가정하고 점수화하였다. 각 기업의 일별 웹 뉴스양의 다음과 같다.

$$Z \sim N(0, 1^2) \quad (3)$$

$$NScore_{i,t} = P\left(Z < \frac{N_{i,t} - \mu_{n,i}}{\sigma_{n,i}}\right),$$

$$(i = 1, \dots, 76 \quad t = 1, \dots, 249)$$

웹 뉴스의 양의 극단적인 변화를 살펴보기 위하여 NScore의 통계적으로 유의한 변화를 다음과 같이 집합으로 정의하였다.

$$NScore0 = \{NScore | NScore < 0.95\}$$

$$NScore1 = \{NScore | NScore \geq 0.95\}$$

즉, 95%의 신뢰구간을 기준으로 웹 뉴스의 양이 유의하게 변화 한 날과 일반적인 날의 집합인 NScore1과 NScore0을 정의하였다.

4.2.2 추가정보

종목의 일별 수익률을 구하기 위한 기준으로는 일별 증가를 이용하였다. $P_{i,t}$ 는 기업 i 의 t 시점에서의 증가를 의미한다. 또한 KOSPI50과 KOSDAQ STAR INDEX를 지수의 기준으로 하였다.

$$I_{KOSPI,t} = KOSPI의 t시점에서의 지수, \quad (5)$$

$$(t = 1, \dots, 249)$$

$$I_{KOSDAQ,t} = KOSDAQ의 t시점에서의 지수,$$

$$(t = 1, \dots, 249)$$

시계열의 효과와 종목별 시가총액 규모에 따른 편차를 제거하기 위해, 일반적으로 각 기업의 일별 수익률은 다음과 같이 정의한다.

$$R_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}} \quad (6)$$

$$(i = 1, \dots, 76 \quad t = 1, \dots, 249)$$

또한 지수의 수익률도 동일한 방법으로 정의한다. 본 연구에서는 시장 전체의 움직임을 제거하기 위하여 각 종목의 수익률에서 지수의 수익률을 뺀 추가대비수익률을 분석에 이용하였다. 즉 각 종목의 일별 수익률을 다음과 같이 다시 정의하였다.

$$RValue_{i,t} = \begin{cases} R_{i,t} - R_{KOSPI,t} & (i \in KOSPI) \\ R_{i,t} - R_{KOSDAQ,t} & (i \in KOSDAQ) \end{cases}, \quad (7)$$

$$(i = 1, \dots, 76 \quad t = 1, \dots, 249)$$

각 종목의 일별 변동성을 구하기 위한 기준으로는 수익률과 마찬가지로 일별 증가를 이용하였다. 시계열의 효과와 종목별 시가총액 규모에 따른 편차를 제거하기 위해, 기존의 연구를 참고하여 각 종목의 일별 변동성을 다음과 같이 정의하였다.

$$V_{i,t} = |R_{i,t}| \quad (i = 1, \dots, 76 \quad t = 1, \dots, 249) \quad (8)$$

일반적인 재무, 금융 연구에서의 변동성은 최근 20일, 60일, 120일 등의 정해진 기간 동안의 증가를 기준으로 정의하는 것에 반해 본 연구에서는 일별 데이터를 얻기 위해 하루간의 주가의 변동 폭을 변동성으로 정의하였다. 또한 시장 전체의 움직임을 제거하기 위하여 각 종목의 변동성에서 지수의 변동성을 뺀 추가대비변동성을 분석에 이용하였다. 즉 각 종목의 일별 변동성을 다음과 같이 다시 정의하였다.

$$VValue_{i,t} = \begin{cases} |R_{i,t} - R_{KOSPI,t}| & (i \in KOSPI) \\ |R_{i,t} - R_{KOSDAQ,t}| & (i \in KOSDAQ) \end{cases}, \quad (9)$$

$$(i = 1, \dots, 76 \quad t = 1, \dots, 249)$$

거래량은 각 종목의 하루 동안 거래된 주식의 수를 의미한다. 각 종목의 일별 거래량을 직접 이용하기에는 변동성, 수익률과 마찬가지로 종목의 규모에 따른 편차가 있기 때문에 재무, 금융 연구에서의 거래회전율의 개념을 이용하였다. 각 종목의 일별 거래회전율($QValue_{i,t}$)은 일별 거래량($q_{i,t}$)을 상장 주식수($Q_{i,t}$)로 나눈 값을 의미한다.

즉, 본 연구에서는 위의 거래회전율 $QValue$ 를 거래량 분석에 이용하였다. 각 기업 주가의 상승과 하락 추세를 구분하기 위하여 20일 이동 평균($M_{i,t}$)을 이용하였다. 20일 이동 평균은 각 기업의 최근 20일 증가의 평균을 의미한다. 주가가 20일 이동 평균보다 큰 날은 상승 추세에 있는 날, 주가가 20일 이동 평균보다 작은 날은 하락 추세에 있는 날로 정의하였다.

따라서 각 기업의 일별 상승과 하락 추세의 정

의는 다음과 같다.

$$TValue_{i,t} = \begin{cases} 1 & (\text{if } P_{i,t} > M_{i,t}) \\ -1 & (\text{if } P_{i,t} \leq M_{i,t}) \end{cases} \quad (10)$$

(i = 1, ..., 76 t = 1, ..., 249)

위와 같은 과정을 거쳐 RValue, VValue, QValue, TValue는 각각 249일의 표본 기간과 76개의 종목을 곱한 18,924개의 데이터를 얻었다.

5. 분석 결과

5.1 웹 뉴스의 양과 주가

가설 1-1을 검증하기 위해 일일 웹 직접 뉴스의 양, 일일 웹 간접 뉴스의 양을 독립변수로 설정하고 일일 주가 변동성을 종속변수로 설정하여 단순 회귀 분석을 실시하였다.

<표 1> 가설 1-1 모델(요약, 분산 분석, 회귀 분석)

모형	R	R ²	수정된 R ²	추정값의 표준오차
1	.204a	.042	.042	1.58845

a : 예측값 : (상수)/UScore_웹 간접 뉴스의 양/
DScore_웹 직접 뉴스의 양.

모형	제공합	자유도	평균 제공	F	Sig.
회귀 모형	2,078.148	2	1,039.074	411.813	.000a
잔차	47,740.917	18,921	2.523		
합계	49,819.066	18,923			

a : 예측값 : (상수)/UScore_웹 간접 뉴스의 양/
DScore_웹 직접 뉴스의 양.

b : 종속변수 : VValue_변동성.

모형	비표준화 계수		표준화 계수	t	Sig.
	B	표준오차	베타		
(상수)	1.055	.029		36.392	.000
DScore	.258	.054	.036	4.742	.000
UScore	1.177	.048	.188	24.497	.000

a : 종속변수 : VValue_변동성.

<표 1>에서와 같이 t-검정과 F-검정 모두 통계적으로 유의한 것으로 나타났다. 주가에는 수많은 요인들이 영향을 미치게 되므로 회귀분석의 설명력이 높지는 않으나 통계적으로 유의한 것으로 보아 일일 웹 직접 뉴스의 양과 일일 웹 간접 뉴스의 양은 일일 주가 변동성에 영향을 미치는 변수이다. 따라서 가설 1-1을 지지한다.

<표 2> 가설 1-2 모델(요약, 분산 분석, 회귀 분석)

모형	R	R ²	수정된 R ²	추정값의 표준오차
1	.155a	.024	.024	1.22220

a : 예측값 : (상수)/UScore_웹 간접 뉴스의 양/
DScore_웹 직접 뉴스의 양.

모형	제공합	자유도	평균 제공	F	Sig.
회귀 모형	696.581	2	348.291	233.161	.000a
잔차	28,263.730	18,921	1.494		
합계	28,960.312	18,923			

a : 예측값 : (상수)/UScore_웹 간접 뉴스의 양,
DScore_웹 직접 뉴스의 양.

b : 종속변수 : QValue_거래량.

모형	비표준화 계수		표준화 계수	t	Sig.
	B	표준오차	베타		
(상수)	.524	.022		23.496	.000
DScore	.238	.042	.044	5.688	.000
UScore	.636	.037	.133	17.211	.000

a : 종속변수 : QValue_거래량.

다음으로 가설 1-2를 검증하였다. 이를 위해 일일 웹 직접 뉴스의 양, 일일 웹 간접 뉴스의 양을 독립변수로 설정하고 일일 거래량을 종속변수로 설정하여 단순 회귀 분석을 실시하였다. <표 2>에서와 같이 t-검정과 F-검정 모두 통계적으로 유의한 것으로 나타났다. 가설 1-1에서의 변동성과의 회귀분석 결과와 마찬가지로 회귀분석의 설명력이 높지는 않으나 통계적으로 유의한 것으로 보아 일일 웹 직접 뉴스의 양과 일일 웹 간접 뉴스의 양

은 일일 거래량에 영향을 미치는 변수이다. 따라서 가설 1-2를 지지한다.

마지막으로 가설 1-3을 검증하기 위해 일일 웹 직접 뉴스의 양, 일일 웹 간접 뉴스의 양 그리고 주가의 상승하락 추세를 독립변수로 설정하고 일일 수익률을 종속변수로 설정하여 단순 회귀 분석을 실시하였다.

<표 3>에서와 같이 t-검정과 F-검정 모두 통계적으로 유의한 것으로 나타났다. 가설 1-1과 가설 1-2에서의 변동성, 거래량과의 회귀분석 결과와 마찬가지로 회귀분석의 설명력이 높지는 않으나 통계적으로 유의한 것으로 보아 일일 웹 직접 뉴스의 양과 일일 웹 간접 뉴스의 양 그리고 주가의 상승과 하락 추세는 일일 수익률에 영향을 미치는 변수이다. 따라서 가설 1-3를 지지한다.

<표 3> 가설 1-3 모델(요약, 분산 분석, 회귀 분석)

모형	R	R ²	수정된 R ²	추정값의 표준오차
1	.240a	.058	.058	2.29929

a : 예측값 : (상수), TValue_상승 하락 추세, UScore_웹 간접 뉴스의 양, DScore_웹 직접 뉴스의 양.

모형	제곱합	자유도	평균 제곱	F	Sig.
회귀 모형	6,127.156	2	2,042.385	386.323	.000a
잔차	100,024.852	18,920	5.287		
합계	106,151.008	18,923			

a : 예측값 : (상수), TValue_상승 하락 추세, UScore_웹 간접 뉴스의 양, DScore_웹 직접 뉴스의 양.

b : 종속변수 : RValue_수익률.

모형	비표준화 계수		표준화 계수	t	Sig.
	B	표준오차	베타		
(상수)	-.362	.042		-8.606	.000
DScore	.421	.079	.041	5.344	.000
UScore	.395	.070	.043	5.666	.000
TValue	.531	.017	.224	31.646	.000

a : 종속변수 : RValue_수익률.

5.2 웹 뉴스의 양의 극단적인 변화와 주가

가설 2-1을 검증하기 위해 독립표본의 평균을 비교하는 t-검정을 실시하였다. 즉, 웹 뉴스의 양이 유의하게 증가한 날의 집단과 그렇지 않은 날의 집단으로 나누어 변동성의 평균을 비교하였다. <표 4>와 같이 웹 뉴스의 양이 유의하게 증가한 데이터는 전체 표본 18,924개 중 1,220개였다. 웹 뉴스의 양이 유의하게 증가한 날의 변동성은 평균 2.36%로 일상적인 날의 평균 변동성인 1.68%보다 높음을 알 수 있다. 표에서 이러한 평균 차이에 대한 통계적인 분석을 위하여 t-검정을 실시한 결과, 유의함을 알 수 있었다. 따라서 다음의 가설 2-1을 지지한다.

<표 4> 가설 2-1 집단통계량, 독립표본 검정

	NScore01 변화	N	평균	표준 편차	표준 오차
VValue 변동성	1	1,220	2.3629	2.4086	.0689
	0	17,704	1.6816	1.5442	.0116
평균의 동일성에 대한 t-검정					
	t	자유도	Sig.	평균차	
VValue 변동성	9.743	1,288.966	.000	.68135	

다음으로 가설 2-2를 검증하기 위해 가설 2-1에서와 마찬가지로 독립표본의 평균을 비교하는 t-검정을 실시하였다. 즉, 웹 뉴스의 양이 유의하게 증가한 날의 집단과 그렇지 않은 날의 집단으로 나누어 거래량의 평균을 비교하였다.

<표 5>와 같이 웹 뉴스의 양이 유의하게 증가한 데이터는 전체 표본 18,924개 중 1,220개였다. 웹 뉴스의 양이 유의하게 증가한 날의 거래량은 평균 1.34로 일상적인 날의 평균 거래량인 0.9보다 높음을 알 수 있다. 이러한 평균 차이에 대한 통계적인 분석을 위하여 t-검정을 실시한 결과, 유의함을 알 수 있었다. 따라서 다음의 가설 2-2를 지지한다.

〈표 5〉 가설 2-2 집단통계량, 독립표본 검정

	NScore01 변화	N	평균	표준 편차	표준 오차
QValue 거래량	1	1,220	1.3403	1.9112	.0547
	0	17,704	.9033	1.1713	.0088
평균의 동일성에 대한 t-검정					
	t	자유도	Sig.	평균차	
QValue 변동성	7.884	1,282.86	.000	.43695	

마지막으로 가설 2-3을 검증하기 위해 가설 2-1과 2-2에서와 마찬가지로 독립표본의 평균을 비교하는 t-검정을 실시하였다. 먼저 주가의 상승과 하락추세를 고려하여 상승중인 경우와 하락중인 경우 두 번으로 나누어 각각 평균 비교 검정을 실시하였다. 즉, 상승추세, 하락추세 각각의 상황에서 웹 뉴스의 양이 유의하게 증가한 날의 집단과 그렇지 않은 날의 집단으로 나누어 수익률의 평균 비교를 실시하였다.

〈표 6〉은 상승 추세일 경우의 집단 통계량을 나타낸다. 상승 추세일 경우 웹 뉴스의 양이 유의하게 증가한 데이터는 672개, 그렇지 않은 데이터가 8,904개였다. 웹 뉴스의 양이 유의하게 증가한 날의 수익률 평균은 1.4%로 일상적인 날의 평균 수익률인 0.50%보다 높음을 알 수 있다. 이러한 평균 차이에 대한 통계적인 분석을 위하여 t-검정을 실시한 결과, 유의함을 알 수 있었다.

〈표 6〉 가설 2-3 집단통계량, 독립표본 검정(상승)

	NScore01 변화	N	평균	표준 편차	표준 오차
RValue 수익률	1	672	1.4016	3.2145	.1240
	0	8,904	.4967	2.2875	.0242
평균의 동일성에 대한 t-검정					
	t	자유도	Sig.	평균차	
RValue 수익률	7.162	723.190	.000	.90490	

즉, 주가가 상승 추세인 경우 일일 웹 뉴스의 양

이 유의하게 증가한 날에는 그렇지 않은 날보다 일일 추가 수익률이 더 크다.

〈표 7〉은 하락 추세일 경우의 집단 통계량을 나타낸다.

(표 7) 가설 2-3 집단통계량, 독립표본 검정(하락)

	NScore01 변화	N	평균	표준 편차	표준 오차
RValue 수익률	1	548	-.884	3.08	0.131
	0	8800	-0.51	2.16	0.023
평균의 동일성에 대한 t-검정					
	t	자유도	Sig.	평균차	
RValue 수익률	-2.815	581.047	.005	-.3764	

하락 추세일 경우 웹 뉴스의 양이 유의하게 증가한 데이터는 548개, 그렇지 않은 데이터가 8,800개였다. 웹 뉴스의 양이 유의하게 증가한 날의 수익률 평균은 -0.88%로 일상적인 날의 평균 수익률인 -0.51%보다 낮음을 알 수 있다. 이러한 평균 차이에 대한 통계적인 분석을 위하여 t-검정을 실시한 결과, 유의함을 알 수 있었다. 즉, 주가가 하락 추세인 경우 일일 웹 뉴스의 양이 유의하게 증가한 날에는 그렇지 않은 날보다 일일 추가 수익률이 더 작다. 상승 추세일 경우와 하락 추세일 경우의 각각의 결과에 따라 다음의 가설 2-3을 지지한다.

6. 결 론

본 논문에서는 기존의 정보와 추가와의 관계 또는 뉴스와 추가와의 관계에 관한 연구들을 확장하여 웹 뉴스의 양과 추가와의 관계를 살펴보았다.

효율적 시장 가설과 행동 재무론을 이용하여 웹 뉴스의 양이 추가에 영향을 줄 것이라고 가정하였다. 이를 바탕으로 직접, 간접 웹 뉴스의 양이 변동성, 거래량, 수익률에 영향을 줄 것이라는 가설을 설정하고 회귀분석을 하였다. 이를 정리해보면

다음과 같다. 일일 웹 직접 뉴스의 양, 일일 웹 간접 뉴스의 양은 일일 평균 변동성, 일일 평균 거래량에 영향을 미친다. 또 일일 웹 직접 뉴스의 양, 일일 웹 간접 뉴스의 양, 상승/하락 추세는 일일 평균 수익률에 영향을 미친다.

이는 뉴스의 의미와 내용과 상관없이 단순히 뉴스를 통한 정보의 양만을 측정하더라도 주가와 관계가 존재함을 의미한다.

또한 직접, 간접 웹 뉴스의 양이 평균과 비교하여 극단적으로 변화한다면 변동성, 거래량, 수익률에도 변화가 있을 것이라고 가정하였고 이를 바탕으로 평균 비교 검정을 실시하였다. 이를 정리해보면 다음과 같다.

일일 웹 뉴스의 양이 유의하게 증가한 날에는 그렇지 않은 날보다 평균 주가 변동성, 평균 거래량이 더 크다. 또 일일 웹 뉴스의 양이 유의하게 증가한 날에는 그렇지 않은 날보다 평균 주가 수익률이 더 크게 변화한다.

즉, 주가에 영향을 미칠 수 있는 기업에 대한 특정 사건이 발생하였을 때 이는 웹 뉴스의 양의 유의한 증가로 나타나고, 이러한 증가는 투자자들의 심리에 영향을 끼쳐 주가의 매수와 매도 활동을 활발하게 하여 변동성, 거래량, 수익률을 더 크게 변화시킴을 알 수 있다. 인터넷 정보기술의 발달로 인해 적극적으로 활용되고 보급되는 다양한 정보들이 실제 기업의 주가 및 가치에 영향을 미침을 실증적으로 연구하였다는 점에 본 연구의 의의가 있다.

하지만 본 연구는 단순히 웹 뉴스의 양만을 측정했기 때문에, 기사의 내용의 의미를 분석할 수 없었다. 이는 문장 또는 제목에 포함된 감정 분석을 필요로 하는데 이를 위해서는 빠르고 정확한 자연어 처리 과정이 필요하나 현재까지의 기술로는 완벽히 기계적으로 해결할 수 없다는 문제점도 있다. 차후에 기사의 의미를 분석하여 웹 뉴스의 정보로서의 가치를 보다 명확히 계량화할 수 있다면 주가와 관계에 있어서 더욱 정확한 결론을 얻을 수 있을 것으로 생각된다.

참 고 문 헌

- [1] 김문현, “코스닥 IT기업의 개발비의 가치관련성”, 『한국IT서비스학회지』, 제8권, 제3호(2009), pp.67-81.
- [2] 남달우, 박진우, 김민경, 조현, 김성희, “인터넷 주식 게시판을 통한 집단지성과 주식시장과의 상관관계 연구”, 『인터넷전자상거래연구』, 제12권, 제2호(2012), pp.149-164.
- [3] Baig, T. and I. Goldfajn, *Financial Market Contagion in the Asian Crisis*, IMF Staff Paper, WP/98/155, 1998.
- [4] Bank, M., M. Larch, and G. Peter, “Google search volume and its influence on liquidity and returns of German stocks”, *Finance Markets and Portfolio Management*, Vol.25, No.3(2011), pp.239-264.
- [5] Barberis, N., A. Shleifer, and R. Vishny, “A Model of Investor Sentiment”, *Journal of Financial Economics*, Vol.49, No.3(1998), pp. 307-343.
- [6] Blume, L., D. Easley, and M. O'hara, “Market statistics and technical analysis : The Role of Volume”, *Journal of Finance*, Vol.49, No.1(1994), pp.153-181.
- [7] Da, Z., J. Engelberg, and P. Gao, “In Search of Attention”, *The Journal of Finance*, Vol. 66, No.5(2011), pp.1461-1499.
- [8] Daniel, K., D. Hirshleifer, and A. Subrahmanyam, “Investor Psychology and Security Market Under and Overreactions”, *Journal of Finance*, Vol.53, No.6(1998), pp.1839-1886.
- [9] Fama, E. F., “Efficient Capital Markets : A Review of Theory and Empirical Work”, *Journal of Finance*, Vol.25, No.2(1970), pp.383-417.
- [10] Ganapolsky, E. and S. Schmukler, *Crisis Management in Capital Markets : the Im-*

- pacts of Argentine Policy during Tequila Effect*, Working Paper, 1998,
- [11] Hardouvelis, G., "Macroeconomic Information and Stock Prices", *Journal of Economics and Business*, Vol.39, No.2(1987), pp.131-140.
- [12] Hong, H. and S. J. C., "A Unified Theory of Under-reaction, Momentum Trading, and Overreaction in Asset Markets", *Journal of Finance*, Vol.54, No.6(1999), pp.2143-2184.
- [13] Jain, P. C. and G. Joh, "The dependence between hourly stock prices and trading volume", *Journal of Financial and Quantitative Analysis*, Vol.23, No.3(1988), pp.269-283.
- [14] Kahneman, D. and A. Tversky, "Prospect Theory : An Analysis of Decision under Risk", *Econometrica*, Vol.47, No.2(1979), pp. 263-292.
- [15] Kaminsky, G. and S. Schmukler, "What Triggers Market Jitters? A Chronicle of Asian Crisis", *Journal of International Money and Finance*, Vol.18, No.4(1999), pp.537-560.
- [16] Kretzmer, P., *Macroeconomic Announcement, Stock Prices and Trading Volume*, Working Paper, 1991.
- [17] Lakonishok, J., "Past stock price changes and current trading volume", *Journal of Portfolio Management*, Vol.15, No.4(1989), pp. 18-24.
- [18] Liang, X., "Mining Associations between Web Stock News Volumes and Stock Prices", *International Journal of Systems Science*, Vol.37, No.13(2006), pp.919-930.
- [19] Odean, T., "Are Investors reluctant to realize their losses?", *Journal of Finance*, Vol. 53, No.5(1998), pp.1775-1798.
- [20] Preis, T., D. Reith, and H. E. Stanley, "Complex dynamics of our economic life of different scales : insights from search engine query data", *Philosophical transaction of the royal society*, Vol.368, No.1933(2010), pp.5707-5719.
- [21] Rogalski, R., "The dependence of stock prices and volume", *The review of economics and statistics*, Vol.60, No.2(1978), pp.268-274.
- [22] Ross, S. A., "Information and volatility : The no-arbitrage martingale approach to timing and resolution irrelevancy", *Journal of Finance*, Vol.44, No.1(1989), pp.1-18.
- [23] Shefrin, H. and M. Statman, "The Disposition to Sell Winners Too Early and Ride Losers Too Long : Theory and Evidence", *Journal of Finance*, Vol.40, No.3(1985), pp. 777-790.
- [24] Sun, Q. and W. Tong, "The Effect of U.S. Trade Deficit Announcements on the Stock Prices of U.S. and Japanese Automakers", *Journal of Financial Research*, Vol.23, No.1 (2000), pp.15-43.
- [25] Thaler, R. H., "Gambling with the House Money and Trying to Break Even : The Effects of Prior Outcomes on Risky Choice", *Management Science*, Vol.36, No.6(1990), pp.643-660.
- [26] Wysocki, P. D., *Cheap Talk on the Web : The Determinants of Postings on Stock Message Boards*, Technical Report, University of Michigan Business School, 1999.

◆ 저 자 소 개 ◆

**김 상 수 (sangsoo@business.kaist.ac.kr)**

현재 카이스트 경영대학 테크노경영연구소 공공혁신연구센터에서 선임 연구원으로 재직중이다. 카이스트 경영대학 경영공학과에서 IT경영 전공으로 경영공학 석사학위를 취득하였으며 주 연구분야는 모바일 광고, 웹 서비스, e-비즈니스이다.

**남 달 우 (ddalwoo@business.kaist.ac.kr)**

연세대학교 전기전자공학과 졸업 후 한국과학기술원 경영대학에서 석사를 하고 현재 동 대학원에서 MIS 분야로 박사과정에 재학 중이다. 주요 관심분야로 고객관계관리, 집단지성, 소셜네트워크서비스, 전자정부 등이 있으며, 인터넷전자상거래연구, e-비즈니스연구 등에 다수 논문을 실었다.

**조 현 (sineoriz@gmail.com)**

현재 카이스트 경영대학 Smart Decision Support System 연구실에서 박사 후 연구원으로 재직 중이며 국제 e-비즈니스 학회 이사로 역임 중이다. 카이스트 경영대학 경영공학과에서 MIS(Management Information System)을 전공으로 경영공학 박사학위를 취득하였으며 주요 관심분야는 소셜 네트워크 서비스, 주가 예측, IT컨설팅, e-비즈니스, 추천 시스템 등이다. e-비즈니스 연구, 전자상거래학회지, 인터넷 전자상 거래 연구, 인터넷 정보 학회 논문지 등에 다수의 논문을 실었다.

**김 성 희 (seekim@business.kaist.ac.kr)**

미국 스탠포드 대학교에서 경영정보공학 박사학위를 현재 한국과학기술원 경영대학에서 정교수로 재직 중이다. 관심분야는 경영정보, 의사결정 지원 시스템, 전자상거래, 모바일 정책이다. 저서는 그린 IT 경제학, 위키매니지먼트, 전자상거래.com 등이 있으며 대표적 논문으로는 “Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations(2010),” “A New Marketing Strategy Map for Direct Marketing (2009)” 등이 있다.