

The Analysis on the Relationship between Firms' Exposures to SNS and Stock Prices in Korea*

Taehwan Kim**, Woo-Jin Jung**, Sang-Yong Tom Lee***

Can the stock market really be predicted? Stock market prediction has attracted much attention from many fields including business, economics, statistics, and mathematics. Early research on stock market prediction was based on random walk theory (RWT) and the efficient market hypothesis (EMH). According to the EMH, stock market are largely driven by new information rather than present and past prices. Since it is unpredictable, stock market will follow a random walk. Even though these theories, Schumaker (2010) asserted that people keep trying to predict the stock market by using artificial intelligence, statistical estimates, and mathematical models. Mathematical approaches include Percolation Methods, Log-Periodic Oscillations and Wavelet Transforms to model future prices. Examples of artificial intelligence approaches that deals with optimization and machine learning are Genetic Algorithms, Support Vector Machines (SVM) and Neural Networks. Statistical approaches typically predicts the future by using past stock market data.

Recently, financial engineers have started to predict the stock prices movement pattern by using the SNS data. SNS is the place where peoples opinions and ideas are freely flow and affect others' beliefs on certain things. Through word-of-mouth in SNS, people share product usage experiences, subjective feelings, and commonly accompanying sentiment or mood with others. An increasing number of empirical analyses of sentiment and mood are based on textual collections of public user generated data on the web. The Opinion mining is one domain of the data mining fields extracting public opinions exposed in SNS by utilizing data mining. There have been many studies on the issues of opinion mining from Web sources such as product reviews, forum posts and blogs.

In relation to this literatures, we are trying to understand the effects of SNS exposures of firms on stock prices in Korea. Similarly to Bollen *et al.* (2011), we empirically analyze the impact of SNS exposures on stock return rates. We use Social Metrics by Daum Soft, an SNS big data analysis company in Korea. Social Metrics

* This Work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012-2012S1A3A2033291).

** Doctoral students, School of Business, Hanyang University

*** Corresponding Author, Professor, School of Business, Hanyang University

provides trends and public opinions in Twitter and blogs by using natural language process and analysis tools. It collects the sentences circulated in the Twitter in real time, and breaks down these sentences into the word units and then extracts keywords.

In this study, we classify firms' exposures in SNS into two groups: positive and negative. To test the correlation and causation relationship between SNS exposures and stock price returns, we first collect 252 firms' stock prices and KRX100 index in the Korea Stock Exchange (KRX) from May 25, 2012 to September 1, 2012. We also gather the public attitudes (positive, negative) about these firms from Social Metrics over the same period of time. We conduct regression analysis between stock prices and the number of SNS exposures. Having checked the correlation between the two variables, we perform Granger causality test to see the causation direction between the two variables.

The research result is that the number of total SNS exposures is positively related with stock market returns. The number of positive mentions of has also positive relationship with stock market returns. Contrarily, the number of negative mentions has negative relationship with stock market returns, but this relationship is statistically not significant. This means that the impact of positive mentions is statistically bigger than the impact of negative mentions. We also investigate whether the impacts are moderated by industry type and firm's size. We find that the SNS exposures impacts are bigger for IT firms than for non-IT firms, and bigger for small sized firms than for large sized firms.

The results of Granger causality test shows change of stock price return is caused by SNS exposures, while the causation of the other way round is not significant. Therefore the correlation relationship between SNS exposures and stock prices has uni-direction causality. The more a firm is exposed in SNS, the more is the stock price likely to increase, while stock price changes may not cause more SNS mentions.

Keywords: SNS, Stock Market Prediction, Opinion Mining, Thesaurus, Granger Causality

기업의 SNS 노출과 주식 수익률간의 관계 분석

김 태 환, 정 우 진, 이 상 용

I. 서 론

주식 수익률의 변동에 대한 예측은 경제 분야 뿐만 아니라 수학, 통계, 전산 분야에 걸쳐 오랜 기간 매우 중요한 연구과제로 인식되어 왔고 학계에서만이나 실무에서도 높은 관심거리가 되어 왔다. 이러한 주식 수익률 예측에 관한 과거 연구의 상당 부분은 랜덤워크 이론(Random walk theory, 이하 RWT)과 효율적 시장가설(Efficient Market Hypothesis: 이하 EMH)을 근간으로 하고 있다[Bollen *et al.*, 2011]. EMH란 증권 시장에서의 모든 주식 시장가격은 이용 가능한 모든 정보를 충분히 반영하는 것으로서 새로운 정보가 발생되면 증권 시장의 가격이 즉시 이를 반영한다는 의미로 주식 수익률과 정보와의 관계를 설명하는 이론이다. 특히 앞서 언급한 정보라 함은 증권 가격에 충분히 반영되었기 때문에 이러한 정보를 이용해서 투자자는 초과 이윤을 얻을 수 없음을 의미한다[김중호, 2009]. 한편, RWT의 관점에서는 주식 수익률은 과거의 변화 이력이나 어떤 패턴에 영향을 받지 않고 그 자체로서 독립적으로 움직인다는 이론이다. RWT에서 개개의 가격 변동은 시계열(time series) 상에서 서로 상관관계가 없이 독립적이며, 과거의 가격 변동이 어떠한가를 분석하더라도 그것이 미래의 가격 변동 예측에 도움을 줄 수 없다는 이론이다[안성원, 2010].

이러한 과거 이론에도 불구하고, 통계학의 기본 개념을 바탕으로 과거 데이터를 이용해 미래를 예측하는 주식 수익률 예측에 관한 많은 연구들이 진행되어 왔다. 특히, Fama and French[1988]는 장기 주식 수익률이 부정적 자기상관관계(negative autocorrelation)가 유의함을 보임으로써 작게는 25%, 높게는 45%까지를 과거 수익률을 통하여 주식

수익률을 예측할 수 있다고 하였다[김영규, 1994].

주식 수익률 예측에 대한 연구는 Schumaker [2010]의 분류에 따라 수학적 예측, 통계학적 예측, 인공지능적 예측 등의 세 유형으로 분류 할 수 있다. 수학적 예측 방법은 기존의 수학적 모델을 기초로 정량적 평가를 통해 미래 가치를 예측하여 투자 여부를 결정하는 기법이다. 대표적인 수학적 예측 모델에는 여과 방법(percolation method)과 웨이블릿 변환 방법(Wavelet transform)이 있다. 통계학적 예측 방법은 과거의 주식시장 데이터로부터 미래를 예측하는 접근법으로 이동평균(moving average) 분석과 몬테카를로(Monte Carlo) 시뮬레이션과 같은 확률 모델이 있고, 주식 수익률 예측에 대한 최적화와 기계 학습에 주로 사용되는 인공지능적 예측 방법에는 SVM(Support Vector Machine), 신경망(Artificial Neural Network), 유전 알고리즘(Genetic Algorithm) 등이 있다[안성원, 2010]. 이와 같이 다양한 주식 수익률 예측에 대한 방법론이 존재 하지만, 최근에는 기존의 주식 수익률 예측의 유형과는 별개로 금융공학적인 접근법을 통해 주식 수익률 예측과 활용에 대한 연구가 진행 되어 오고 있다. 이러한 접근은 빅데이터의 활용이 활발해지면서 더욱 관심을 끌고 있는데, 빅데이터의 대표적 대상 중의 하나인 SNS를 이용해 주식 수익률을 예측 하여 실제 투자결정에 활용한 결과 더 높은 수익을 올리게 되었다는 사례들이 보고되고 있다.¹⁾ 이러한 결과는 SNS를 이용한 주식 수익률 예측 방법이 가능하다는 것을 뜻할 수도 있는 것이다.

주식 수익률은 수많은 참가자의 위험에 대한 서로 다른 대처 방안, 투자자들의 투자 기간의 차이,

1) 조선비즈, [소셜미디어투자] 트위터 분석투자 5개월 수익률 9.7%, 2013. 04. 25.

새로운 정보에 대한 또 다른 대응이나 동기 유발에 따라 매우 복잡한 양상을 보이고 있는데 이는 연속 발생 데이터라는 특성에 기인한다. 주식 수익률의 이러한 특성들로 인해 빅데이터 분석을 이용한 주식 수익률 예측 방식에 대한 관심이 대두되고 있는데, 그 중에서도 스마트 시대와 맞물려 개인의 오피니언 생산과 파급 속도가 빠른 SNS를 통해 방대한 양의 정보가 발생되고 전달되면서, 빅데이터를 이용한 주식 수익률 예측은 그 의미가 점차 확대되고 있다고 할 수 있을 것이다. 최근 이러한 경향에 맞추어 Bollen *et al.*[2011]은 특정 기간에 상장된 기업 주식의 증가를 추출하여 같은 기간에 SNS상에서 소통된 불특정 대중들의 감정적 표현을 추출하여 특정 이벤트(추수감사절, 대선)에 변화된 대중들의 감정 변화에 따라 주식 수익률이 어떻게 변하는지를 실증 분석하였다.

본 논문은 Bollen *et al.*[2011]의 연구와 유사하게, 국내에서 서비스되고 있는 블로그와 트위터 등과 같은 SNS에서 특정 기업의 노출의 변화와 주식 수익률 흐름의 관계를 알아보고, 이러한 SNS상의 노출의 변화가 기업의 주식 수익률에 얼마나 영향을 미치는지를 파악해 보고자 하는 실험적 연구이다. 그러나, 과거 연구에서는 상관관계 분석에 중심이 맞춰져 있다면, 본 연구에서는 인과관계 분석도 함께 시도하고자 한다. 본 논문의 구성은 다음과 같다. 제 II장에서는 연구배경이 되는 SNS, 빅데이터 분석, 오피니언 마이닝 등을 다룰 것이다. 제 III장에서는 연구방법에 대하여 논의하고, 제 IV장에서는 연구결과에 대한 논의할 것이다. 마지막으로 본 논문을 요약하고 결론을 도출할 것이다

II. 연구 배경

2.1 소셜미디어와 SNS

소셜미디어란 '사람들이 연결되어 있는 관계망'이란 개념으로 Social Networking Service(SNS)라 불리 우기도 한다[김지수, 2004]. 소셜미디어의 특징

은 참여(participation), 공개(openness), 대화(conversation), 커뮤니티(community), 연결(connectness) 등이 있는데, 참여란 특정한 주제에 관심이 있는 사람들이 모여 스스로 자신들의 지식, 의견, 및 피드백들을 서로 공유하는 개념이다. 공개란 참여의 활동 자체가 누구에게나 열려있기 때문에 이용자들 사이의 참여활동 등을 촉진한다. 대화는 정보의 제공자와 소비자가 서로 쌍방향 커뮤니케이션 방식을 지향하는 것이고, 커뮤니티는 동일한 주제에 관심을 가진 이용자들끼리 서로의 정보를 공유할 수 있도록 지원하는 것이다. 마지막으로 연결은 이러한 정보 및 이용자들을 서로 연결하여 관계형성을 촉진하는 것이다[한국정보산업연합회, 2007].

Kaplan and Haenlein[2010] 등의 연구에서는 소셜미디어를 웹 2.0의 기본 발상인 사용자들의 참여라는 이념과 기술을 지향하는 인터넷 기반 어플리케이션의 한 그룹이라 정의했고, 이는 이용자들이 UGC(User Generated Contents)를 만들거나 교환 할 수 있도록 그 기능을 수행한다고 명시하였다. 특히 Kaplan and Haenlein[2010]의 논문에서 SNS는 이용자들이 개인의 프로필을 생성하고 이러한 프로필에 접근하기 위해 친구와 동료들을 초대하고 이메일이나 메시지를 전달하게 하는 어플리케이션이라고 정의하면서 소셜미디어는 SNS를 포괄하는 개념이라고 주장하였다. 본 논문에서는 Kaplan and Haenlein[2010]의 분류에 따라 SNS에 초점을 맞추어 연구를 진행할 것이다.

SNS상에서 이용자들이 자신의 경험을 타인과 '공유'하는 것은 정보를 전달하는 커뮤니케이션의 한 일종으로서 구전 커뮤니케이션이라 할 수 있다. 일반적으로 온라인 구전은 다음의 특징들로부터 오프라인 구전과 구별된다. 첫째, 오프라인의 경우 구전의 상대방이 정보를 제공한 자와 대부분 가까운 지인들로 구성되어있지만, 온라인 구전의 구전 대상은 오프라인의 구전과는 달리 그 범위에 한계가 없다[Ratchford, 2001; Sun *et al.*, 2006]. 둘째, 온라인 구전은 '텍스트'를 기반으로 정보의 전달이 이루어지기 때문에 구전 정보의 영구적 보관이 가능하고 정보 전달의 속도 또한 빠르다[성영신 외, 2002].

최근에는 이러한 온라인 구전의 특징을 이용해 정치, 경제 그리고 사회적으로 기업, 개인의 이미지 혹은 제품의 홍보 효과를 내기 위해 다양하게 이용되고 있다. 본 논문에서는 이러한 온라인 구전의 특징을 'SNS 효과'라고 재정의 하고, 이러한 'SNS 효과'가 주식 투자자들의 행동에 변화를 일으켜 궁극적으로 주가에 영향을 미칠 수가 있는지의 여부를 알아보고자 한다.

2.2 빅데이터 분석

초기에 빅데이터의 개념은 구글과 같은 대기업이나 NASA의 연구과학 프로젝트에서 분석하는 단순히 대용량의 데이터를 일컫는 것이었다[Merv Adrian, 2011]. 반면, Manovich[2011]는 데이터를 수집하고 처리하던 기존의 소프트웨어의 능력을 넘어서는 데이터들의 모임을 빅데이터로 정의하였고 더욱이 이런 데이터의 크기는 꾸준히 그 범위가 증가할 것이라고 하였다. 또한, 근래에는 다방면에서 무수히 많은 데이터가 실시간으로 생산되고 있고, 특히 스마트폰의 단말 및 SNS 등으로 대표되는 다양한 정보 채널의 등장과 이로 인한 정보의 생산, 유통, 보유량의 증가는 데이터의 기하급수적인 증가를 촉진하고 있다[조성우, 2011]. 따라서 다방면에서 산재 되어 있는 방대한 양의 데이터를 처리하기 위해서는 진보된 분산 처리 기술과 통계적 기법인 빅데이터 분석이 요구되고 있다. 이러한 관점으로 본 연구에서는 '빅데이터 분석'을 기존의 데이터 처리 방식으로는 처리할 수 없는 막대한 양의 데이터를 분석하여 새로운 데이터, 새로운 인사이트를 창출하는 것으로 정의하였다.

2.3 오피니언 마이닝

오피니언 마이닝은 데이터 마이닝 기술의 한 분야로써 웹 상(특히 블로나 트위터 등)에 게재된 이용자들의 상품평 및 견해 등에 나타난 의견을 추출하는 분야로서, 텍스트의 주제를 판단하는 것이 아닌 주제에 대한 저자의 태도나 생각을 판단

하는 기술이다[김진옥, 2011]. 이러한 오피니언 마이닝은 일반적으로 다음 3단계를 거쳐게 된다. 첫째, 이용자들이 사용한 문장에서 긍/부정과 관련된 표현들을 추출하고, 둘째, 세부 평가요소와 그것이 가리키는 오피니언의 연결 관계를 포함한 문장을 인식하고, 셋째, 긍/부정 표현 및 유용한 문장들을 추출하여 요약문을 생성한다. 각 단계 대한 설명은 아래와 같다. 첫째로, 오피니언 마이닝 기법 중에서는 긍정/부정 표현에 해당하는 어휘 정보를 명확하게 추출하는 것이 중요한데 이는 기존에 구축된 오피니언 사전을 이용하거나 연구자가 직접 수작업을 통해서 해당 콘텐츠에서 빈도율이 높은 긍정/부정을 표현하는 단어들을 추출하는 방법이 있다. 또한, 기계적인 학습 데이터를 활용한 통계 프로그램을 이용하여 자동으로 어휘 정보를 얻을 수도 있다. 한편, 통계적인 방법을 이용할 때는 Mutual information과 같은 평가척도를 사용할 수 있다[Zhuang et al., 2006, Blair-Goldensohn et al., 2008]. 둘째, 세부 평가요소와 오피니언으로 구성된 문장을 인식하는 것이 중요하다. 이 단계에서는 첫 번째에서 구축된 어휘 정보를 사용하여 추출한 오피니언에 대해 세부 평가요소와 긍정/부정 표현을 찾게 된다. 또한, 긍정적 혹은 부정적 오피니언인지 판단하기 위해서는 여러 가지 방법을 적용할 수 있다. 예를 들면, 형용사를 오피니언 단어로 간주하고 긍정/부정을 결정하기 위하여 기존의 데이터 사전 등을 활용한다. 셋째로, 긍/부정어 표현의 개수나 중요 문장을 추출한 후 이를 통해 데이터 사전을 만들고 또한 세부 평가요소와 매치되는 오피니언을 포함하는 문장들 중 유의미한 문장들을 긍정/부정 별로 분류하여 추출하여야 한다. 이렇게 생성된 오피니언 마이닝의 결과는 긍정어와 부정어 평가의 정도를 나타내거나 오피니언 사전 형태로 제시될 수 있다[박경미, 2011].

본 연구에서 이러한 오피니언 마이닝 연구를 하기 위해 국내 SNS 분석 서비스 전문 기업인 다음소프트를 이용하였는데, 다음소프트사 홈페이지에 게재된 내용을 인용하면 "소셜네트워크의 이슈와 관심

키워드의 실시간 모니터링으로 상황에 맞는 대응 전략 및 마케팅 전략의 수립을 지원하는 '소셜 매트릭스 서비스'를 제공한다."고 되어 있다. 또한 이 서비스는 자연어 처리, 데이터마이닝 등의 분석틀을 활용하여 블로그와 트위터에서 형성되는 트렌드와 여론을 일반과 공공으로 구분하여 제공하고 있다. 2013년 다음소프트사 관계자와의 인터뷰에 의하면, "이 서비스는 주로 트위터를 기반으로 실시간으로 유포되고 있는 글들을 모아 이를 단어 단위로 쪼개 뒤 빅데이터 분석을 통해 키워드를 추출하는 방식을 쓰고 있다. 소셜매트릭스는 국내에 유통되고 있는 대부분의 트윗 등을 분석해 결과를 제시한다. 여기에 연관어 분석이 추가된다. 예를 들어 분석을 통해 '갤럭시', '아이폰' 등과 같은 키워드를 추출한 뒤 이 단어의 앞뒤에 붙어있는 단어들을 추가로 분석해 긍정적인지, 부정적인지

등에 대한 정보를 추가로 파악한다. 갤럭시라는 단어가 속한 문장의 앞뒤로 '좋다' '안좋다'라는 등의 연관어가 얼마나 붙어 있는지를 함께 분석하는 것이다."라고 밝혔다.2)

이러한 기준으로 본 논문에서는 기업에 대한 이용자의 태도를 긍정/부정 두 카테고리로 분류를 하였고 긍정어는 주로, 이익, 호조, 혁신적, 신개념, 상승세 등 기업 발전에 관여하는 단어들로 구성되어 있었으며, 긍정어는 총 200여 개의 단어로 구성되어 있었다. 반면 부정어의 경우는 경기침체, 불법, 불량, 하락세, 손실, 급락, 재정적자, 정보유출, 불매운동, 사고발생 등 기업 발전을 저해하는 단어들로 구성되어 있었으며, 부정어는 총 184개 정도의 단어로 구성되어 있었다. 수집된 긍정어와 부정어의 내용은 아래의 표와 같다.

<표 1> 긍정어와 부정어 리스트

	긍정어	부정어
단어형	최고, 이익, 기대, 호조, 프리미엄, 혁신적, 성공적, 기대감, 기부, 긍정적, 강제, 안정적, 상승세, 국내 최초, 응원, 호평, 기능 추가, 우승, 최저, 세계 최고, 합리적, 독자적, 공정, 우위, 특별, 대박행진, 축하, 소망, 신개념, 스페셜, 자신감, 적극적, 시장선점, 매진, 우승자, 창의적, 윤리적, 비용절감, 비려, 존경, 도움, 추천, 정상적, 저가, 조직적, 진심, 아름다움, 공감, 보람, 베스트, 상위권, 무료, 침착, 만장일치, 정성, 신중, 환상적, 파이팅, 투혼, 웃음, 원조, 흥행, 권장, 어메이징, 기능강화, 폭소, 최고 서비스, 최선, 자유, 탄생, 행운, 당첨, 가압비 면제, 낭만적, 국제적, 실용적, 희망, 반값합인, 즐거움, 고민해결, 감사인사, 적극활용, 시장화대, 진리, 냄새제거, 완벽, 함박웃음, 빵터진, 친환경적, 기분전환, 사랑, 완승, 기적, 노력, 정상, 감탄, 열광, 효과적, 정밀, 활약, 포옹, 득템, 친절, 안전, 가격인하, 건강, 체계적, 트렌디, 럭셔리, 만족, 가치창출, 성공사례, 보상, 존중, 결합할인, 알뜰, 믿음, 지지, 클래식, 1등급, 감각적, 무상지원, 개방적, 해피, 추모, 신선, 정복, 대만족, 쫄다, 강추, 눈호강, 매력적, 유명세, 애정, 청력, 성실, 고무적, 시청률 상승, 위생적, 호감, 단독선두, 퍼펙트, 환호, 칭찬, 역동적, 울킬, 유감적, 최고 인기, 열애, 과학적, 파격적, 무머해명, 극찬, 획기적, 흥행돌풍, 기립박수, 매력발산, 좋은책, 으뜸, 시간단축, 소신발언, 열풍	경기침체, 피해, 불법, 몰락, 공격적, 부정적, 불량, 충격, 경제, 탄식, 루머, 우려, 비상, 하락세, 손실, 급락, 불만, 반발, 위기, 실수, 재정적자, 출시연기, 피해발생, 정보유출, 부진, 설상가상, 불매운동, 유료, 협상결렬, 악영향, 사고발생, 특허권침해, 경쟁심화, 폭락, 논란, 과실, 고가, 소극적, 불편, 소송제기, 유가상승, 고민, 충격적, 역차별, 문제발생, 능욕, 탈락, 패배, 이혼, 폭행, 비난, 의심, 심려, 비판, 멘붕, 분노, 무개념, 음주운전, 좌절, 악의적, 만취상태, 욕설, 막말, 악취, 귀차니즘, 불편한 진실, 명예훼손, 역관광, 발암물질, 욕, 먹튀, 악평, 오해, 불륜, 종말, 낭비, 황포, 편법, 성형의혹, 하극상, 저작권 침해, 엽기적, 부정행위, 멘탈붕괴, 짜증, 병맛, 공포, 난감, 망신, 불가능, 허위사실, 마비, 이기적, 고장, 부담, 연패, 방송사고, 부작용, 굴욕, 최악, 손해, 경제위기, 청년실업, 하자, 아쉬운점, 압수수색, 이물질, 피부건조, 고질적, 충동, 과도포장, 아픔, 접속장애, 경악, 과부하, 갈등, 그다, 미달사태, 고압적, 사용불가, 비극, 의도적, 정전, 요금폭탄, 유해, 장애, 후유증, 회의적, 미련, 액정파손, 괴물, 혼란, 후회, 반대, 배신, 폭등, 짝퉁, 참사, 통곡, 불안감, 가치하락, 생활고, 파국, 불행, 피로, 불법적, 절망적, 만취상태, 마보, 급지, 상처, 충격고백, 체중증가, 가짜, 질책, 외면, 정체불명, 당황, 이별, 전쟁, 질투, 고층, 결별설, 비현실적, 뇌물수수, 긴장, 패닉, 치명적, 위협, 민폐, 체포, 서버다운
서술형	저렴하다, 대단하다, 행복하다, 소중하다, 자랑스럽다, 적당하다, 인기 많다, 쓰고 싶다, 먹고 싶다, 다양하다, 시원하다, 아름답다, 깨끗하다, 유리하다, 재미있다.	예민하다, 죄송하다, 답답하다, 지나치다, 억울하다, 심심하다, 이상하다, 안타깝다.

2) 지디넷코리아, 실시간 검색어 + SNS분석 최신 트렌드 보인다, 2013.09.10.

Ⅲ. 연구 모형과 가설

본 연구의 기본적인 연구진행방법은 Bollen *et al.* [2011]의 연구 프로세스 모형을 바탕으로 <그림 1>과 같이 수정 제시하였다. 연구의 진행 과정은 크게 두 과정으로 나눌 수 있다. ①과정에서 오픈이언 마이닝 과정을 거쳐 데이터를 정제하고 ②과정에서 정제된 데이터를 이용해 통계 분석을 실시하였다. 제시된 <그림 1>에서 보듯이, ①과정에 추출된 모든 데이터는 일별 데이터로 구성되어 있고, ②과정에서는 이러한 데이터를 이용해 회귀분석과, Granger Causality 분석을 실시하였다.

먼저, 회귀분석은 앞서 언급한 바와 같이, SNS 효과와 기업의 주식 수익률과의 상관관계를 알아보고자 하는 것이며, SNS 효과는 빅데이터를 활용한 오픈이언 마이닝 기법에 따라 전체 빈도수, 긍정어수, 부정어수로 측정한다. 이렇게 SNS 효과가 주식 수익률에 미치는 일반적 관계를 살펴본 이후에는 SNS 효과가 산업군, 규모, 주식시장에 따라 차별적으로 발생하는지의 여부를 살펴보고자 한다. 이에 입각한 상관관계의 가설은 다음과 같다.

H1: SNS에서의 노출 빈도수의 증가율은 주식 수익률에 영향을 미칠 것이다.

H1-1: SNS에서 기업의 노출 빈도수의 증가율

과 주식 수익률과의 영향 관계는 산업군(IT/비 IT기업)에 따라 차이가 있을 것이다.

H1-2: SNS에서 기업의 노출 빈도수의 증가율과 주식 수익률과의 영향 관계는 규모(상위기업/하위기업)에 따라 차이가 있을 것이다.

H1-3: SNS에서 기업의 노출 빈도수의 증가율과 주식 수익률과의 영향 관계는 주식시장(코스피/코스닥)에 따라 차이가 있을 것이다.

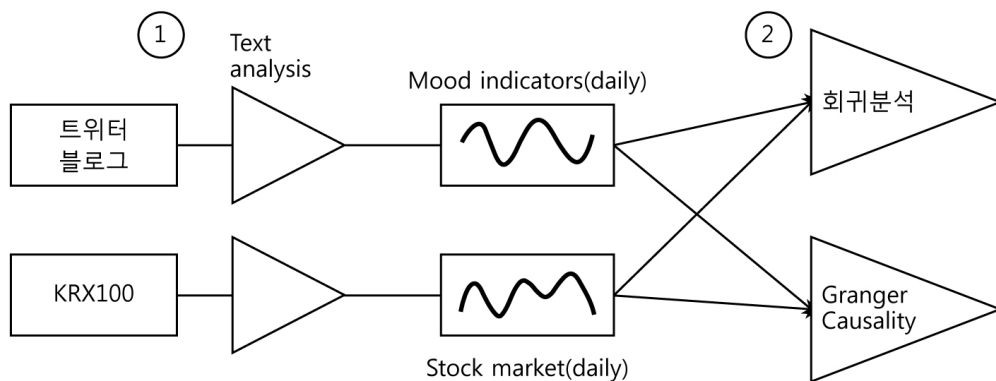
H2: SNS에서의 긍정적 노출의 증가율은 주식 수익률에 영향을 미칠 것이다.

H2-1: SNS에서 기업의 긍정적 노출의 증가율과 주식 수익률과의 영향 관계는 산업군(IT/비 IT기업)에 따라 차이가 있을 것이다.

H2-2: SNS에서 기업의 긍정적 노출의 증가율과 주식 수익률과의 영향 관계는 규모(상위기업/하위기업)에 따라 차이가 있을 것이다.

H2-3: SNS에서 기업의 긍정적 노출의 증가율과 주식 수익률과의 영향 관계는 주식시장(코스피/코스닥)에 따라 차이가 있을 것이다.

H3: SNS에서의 부정적 노출의 증가율은 주식 수익률에 영향을 미칠 것이다.



<그림 1> 연구 프로세스 모형

H3-1: SNS에서 기업의 부정적 노출의 증가율과 주식 수익률과의 영향 관계는 산업군(IT/비 IT기업)에 따라 차이가 있을 것이다.

H3-2: SNS에서 기업의 부정적 노출의 증가율과 주식 수익률과의 영향 관계는 규모(상위기업/하위기업)에 따라 차이가 있을 것이다.

H3-3: SNS에서 기업의 부정적 노출의 증가율과 주식 수익률과의 영향 관계는 주식시장(코스피/코스닥)에 따라 차이가 있을 것이다.

연구가설을 살펴보면 기본 가설인 H1, H2, H3는 독립변수와 종속변수의 관계를 나타내는 것으로 각각의 독립변수인 특정 기업의 노출 빈도수의 증가율, 긍정어수의 증가율 그리고 부정어수의 증가율은 해당 기업의 주식 수익률에 영향을 미칠 것이라는 가설을 설정하였다. 이러한 특정 기업의 노출 빈도수, 긍정어수 그리고 부정어수는 기업의 특성에 따라 해당 기업의 주가에 영향을 미치는 정도의 차이가 나타날 수도 있는데, 이러한 차이를 검증해 보기 위해 추출한 기업 데이터를 산업별(IT/비 IT), 규모별(상위기업/하위기업), 주식 시장별(코스닥/코스피)로 구분을 하여 특정 기업의 노출 빈도수의 증가율, 긍정어수의 증가율, 부정어수의 증가율과 해당 기업의 주식 수익률과의 영향관계를 검증할 위한 가설을 설정하였다.

상관관계의 분석을 마친 이후에는 독립변수인 노출 빈도수의 증가율, 긍정어수의 증가율, 부정어수의 증가율과 주식 수익률사이의 인과관계 존재 여부를 분석 할 것이다. SNS의 노출과 주가간에 상관관계가 있다고 하더라도, 이것만으로는 SNS의 변화가 주식 수익률에 인과관계적 영향을 미친다고 단정할 수는 없다. 왜냐하면, 주식 수익률에 변동이 생김으로 인해, 많은 사람들이 SNS 상에서 해당 기업을 논하게 되는 숫자가 증가할 수도 있는 역의 인과관계가 존재할 수 있기 때문

이다. 따라서 주식 수익률이 SNS의 노출의 변화에 영향을 미치는지, 아니면 SNS의 노출의 변화가 주식 수익률의 변동에 영향을 미치는지를 살펴보기 위하여 Granger 인과관계 분석을 시도하였다.

4. 연구방법 및 설계

4.1 시소러스(Thesaurus) 방법

특정한 문장에서 단어를 추출하여 연구하는 분야에서 일반적으로 주제어 색인 목록을 사용하지 않고 키워드를 추출한 경우 조정 작업을 거치는 것은 필수이다. 예를 들어 반도체라는 단어는 사용자 또는 문맥에 따라 'semiconductor', 'semiconductors', 'semi-conductor', 'semi-conductors' 등은 이용자가 사용하는 방법에 따라 다양하게 쓰일 수 있으나 그 본질적인 모두 같다. 따라서 의미 있는 문헌 분석을 위해서는 위의 단어들을 모두 같은 의미를 지칭하고 있는 것으로 받아드려져야 한다. 이러한 변환 작업을 시소러스 방법이라 한다. 일반적인 시소러스의 사전적 정의는 "단어의 의미에 따라 분류, 배열한 유의어, 동의어, 혹은 반의어 사전이나 정보 검색을 위해 컴퓨터에 내제된 용어사전"을 의미한다. 시소러스의 방법론의 목적은 사람들이 일상적으로 사용하는 유사한 말을 대표적인 하나의 의미로 규격화함으로써, 검색에 있어서의 개념적 통일을 추구하고 누구나 같은 순서와 내용의 검색이 가능하도록 하는 것이다[Ding et al., 2001].

본 연구에서는 시소러스 방법을 이용해 개별 기업의 데이터를 수집하였다. 예를 들면, 현대자동차 기업의 노출빈도와 긍/부정의 수를 조회하기 위해 SNS 이용자들이 '현대자동차'라고 지칭을 할 수 있고, 혹은 '현대차'라고 할 수도 있기 때문에 두 검색어 모두 조회 후 마지막에 취합하는 방법으로 데이터를 수집하였다. 시소러스 방법을 이용해 추출한 기업 데이터는 아래의 표와 같다.

<표 2> 시소러스 방법을 통해 추출한 기업리스트

현대차	현대차	케이티	KT
현대자동차		KT	
Posco	Posco	엔씨소프트	엔씨소프트
포스코		NC소프트	
기아차	기아차	LGU	LG유플러스
기아자동차		LG유플러스	
신한지주	신한지주	BS금융지주	BS금융지주
신한은행		부산은행	
신한카드		한국금융지주	한국금융지주
신한금융투자		한국증권	
KB금융	KB금융	DGB금융지주	DGB금융지주
국민은행		대구은행	
국민카드		LIG손해보험	LIG손해보험
KB카드		LIG	
NHN	NHN	아시아나항공	아시아나항공
네이버		아시아나	
S-oil	S-oil	미래에셋증권	미래에셋증권
에스오일		미래에셋	
에쓰오일		안랩	안랩
SK텔레콤	SK텔레콤	안철수연구소	CJ요쇼핑
SK telecom		CJ요쇼핑	
SKT		CJ물	
우리금융	우리금융	GS홈쇼핑	GS홈쇼핑
우리은행		GS샵	
우리카드		GS숍	
하나금융지주	하나금융지주	CJ CGV	CJ CGV
하나금융		CGV	
하나은행		Yes24	에스24
하나SK		에스24	
하나카드	롯데쇼핑		
롯데쇼핑			
롯데백화점			
롯데물			
롯데홈쇼핑			
롯데닷컴			
롯데미도파			

4.2 회귀분석

본 연구에서 종속변수인 주식 수익률과 독립 변수인 노출빈도수의 증가율, 긍정어수의 증가율, 부정어수의 증가율과의 상관관계를 분석하기 위해 회귀분석을 실행하였다. Base Model의 회귀분석식은 아래와 같다.

$$Y = \beta_0 + \sum_{i=1}^4 \beta_i X_i + \epsilon \quad (1)$$

여기서 β_0 는 y절편을 나타내고 $\beta(\beta_1, \beta_2, \beta_3, \beta_4)$ 는 각 X_i 변수의 회귀계수들을 나타내는데, X_1 은 KRX100, X_2 는 노출빈도수, X_3 은 긍정어수, X_4 는 부정어수를 나타낸다. ϵ 는 잔차(residual)를 의미한다. 이러한 변수들은 시계열 데이터에서 발생할 수 있는 가성회귀분석의 가능성을 최소화하기 위하여, level 데이터가 아니라 차분이 된 rate 데이터이다. 즉, 모두 증가율인 것이다.

여기에서 한 가지 언급해야 할 것은 식 (1)이 이론에 기본적인 주가 변동 회귀분석 모형이 아니라, 시장평균치와의 일상적 관계에서 벗어나는 기업 단위의 비일상적 수익률(abnormal return)에 SNS상의 노출과 관련한 변수들이 어떠한 영향을 미치는지를 살펴보는 실험적으로 살펴보는 모형이라는 것이다. 재무이론에서의 자산가격결정모형(Asset Pricing Model) 이론에 의하면 개별 주가는 시장 포트폴리오에 의해 설명되는 영역과 이로부터 벗어나는 개별 기업의 특수한 움직임으로 나누어 볼 수 있다. 독립변수의 KRX100이 시장포트폴리오와 관련된 부분을 설명하는 것이고, SNS에서의 노출 관련 변수들이 나머지 개별 기업의 비일상적 수익률에 영향을 미치는지를 분석하는 것이다.

다음으로는, SNS가 기업의 주식 수익률에 영향을 미치는 것이 기업 특성별로 어떻게 다른지에 관한 조절효과를 살펴보기 위한 분석을 실시하였으며, 이와 관련된 회귀식은 다음과 같다.

$$Y = \beta_0 + \sum_{i=1}^4 \beta_i X_i + \sum_{i=2}^4 \beta(3+i) X_i M_k + M_k + \epsilon \quad (2)$$

마찬가지로, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$ 는 회귀 계수를 나타내는데, $\beta_2, \beta_3, \beta_4$ 는 각각 노출빈도수, 긍정어수, 부정어수의 증가율을 나타내는 회귀계수이고, $\beta_5, \beta_6, \beta_7$ 은 각각의 조절변수에 대한 회귀계수를 의미한다. M_k 는 조절효과를 알아보기 위한 더미변수로서, IT기업 회귀분석에서는 산업변수, 상위기업회귀 분석에서는 기업규모변수, 코스피 회귀분석에는 시장변수를 의미한다(<표 3> 참조). 이와 같은 조절효과를 설정한 것은 일반적으로 이벤트연구와 같이 주가의 변화를 통해 MIS 주제들을 다룬 연구에서 흔히 설정하는 변수들과 맥락을 같이하기 위함이다.

4.3 Granger Causality Test

인과관계에서 어느 변수가 원인변수이고 어떤 변수가 결과변수가 되는지 그 방향에 대한 검정은 Granger[1969]에 의해 제안된 것으로, Granger 인과관계의 검정은 두 변수 사이의 인과적 관계의 존재를 확인하는 분석 방법이다. 이러한 Granger 인과관계 검정은 확률변수 사이의 인과관계를 결정하는 데 보편적으로 사용되며 전통적인 F통계량을 이용한 방법이다.

가장 먼저 Granger 인과관계 검정은 한 변수가 다른 변수를 예측하는데 도움이 되지 않는다는 귀무가설에 대한 검정으로서, 변수 Y를 예측할 때, Y의 과거 값과 함께 변수 X의 과거 값도 함께 사용하는 것이 Y의 과거 값만으로 예측하는 것보다 더욱 정확하다는 것이다. 만약 이러한 관계가 성립한 다면 X에서 Y로의 인과방향이 존재한다고 말한다. 반대로 X의 예측 값이 과거 X값에 의존하는 것보다 Y의 과거 값이 포함함으로써 인과관계의 확률이 좋아진다면 Y에서 X로의 인과방향이 존재한다고 본다. 만일 두 방향 모두 유의미하게 성립된다면 X와 Y는 상호 의존적인 관계로 쌍방의 인과관계가 존재한다고 본다. 일반적으로 X와 Y간에 어떤 변수가 원인이 되었는지를 확인하기 위해 Granger 검정법을 사용하는데 다음과 같은 두 회귀방정식으로 나타낸다.

$$Y_t = \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{j=1}^p \beta_j Y_{t-j} + \epsilon_{1t} \quad (3)$$

$$X_t = \sum_{i=1}^n \gamma_i X_{t-i} + \sum_{j=1}^n \delta_j Y_{t-j} + \epsilon_{2t} \quad (4)$$

여기서 y는 주식 수익률을 나타내고 x는 빈도수, 긍정어수, 부정어수의 증가율을 각각 대입하였고 시차는 총 4시차까지 검증을 시도해 보았다.

본 연구의 전반적인 연구과정은 먼저 앞서 제시한 가설을 검증하기 위해 종속변수를 주식 수익률로 독립변수를 KRX100, 노출 빈도수의 증가율, 긍정어 수의 증가율, 부정어 수의 증가율로 정의하고 회귀분석을 실시하여 독립변수와 종속 변수 사이의 상관관계를 검증하였고, 이 후에 Granger Causality 분석을 통하여, 종속변수와 독립변수 간의 인과 관계가 존재하는지 검증하였다.

4.4 데이터

본 연구의 목적인 SNS에서 특정기업의 노출 빈도수의 증가율, 긍정어수의 증가율, 부정어수의 증가율과 해당 기업의 주식 수익률과의 관계를 알아보기 위해 2012년 5월 25일부터 9월 1일까지 총 100일에 걸쳐 데이터를 수집하였다. 기업 데이터는 SNS에 회자되고 있는 기업을 대상으로 데이터 수집 당시 시가 총액을 기준으로 상위기업과 하위기업의 비율을 맞추어 전체 252개 기업으로 구성하였고 해당 기업들의 주식 수익률(증가)과 KRX100 지수를 수집하였다. 또한 동 기간에 각 개별 기업들의 SNS에 노출된 빈도수와 긍정어와 부정어를 수집하였다. 본 연구에서는 시계열 데이터의 특성 중 하나인 데이터의 추세를 제거하기 위해 모든 변수를 증감률로 변형하여 분석하였고, 또한 이러한 효과로 인해 대기업 일수록 빈도수가 많고 중소기업 일수록 빈도수가 적어짐으로써 대기업에 편중되는 문제를 해결할 수 있었다.

한편, 각 기업별 SNS 노출 빈도수와 긍정, 부정어 수는 다음소프트에서 제공하는 소셜메트릭

스 프로그램을 통해 수집하였고, 개별 기업의 주식 수익률과 KRX100 지수는 한국거래소에서 추출하였다. 모든 수집된 데이터는 일별데이터로 구성되어 있고, 회귀 분석을 통해 어떠한 요인이 주식 수익률의 변화에 영향을 미치는지 확인하였다. KRX100 데이터를 독립변수로 사용하는 이유는 개별주식 수익률이 시장 전체의 움직임에서 벗어나는 일종의 비정상적 움직임(abnormal return)이 실제의 분석 대상이 되어야 하기 때문이다. 이를 통하여 SNS 관련 변수들의 변화가 개별주식 수익률의 비정상적 움직임에 미치는 영향을 올바르게 분석할 수 있는 것이다.

또한, 기업 데이터를 상위기업과 하위기업, IT기업과 비 IT기업, 코스피 기업과 코스닥 기업으로 분류하여 어떠한 기업 그룹이 노출 빈도수, 긍정어 수, 부정어 수에 더 많은 영향을 받는지 분석하였다. 상위기업과 하위기업의 분류방법은 분류된 기업의 숫자를 균등하게 맞추기 위하여 분류 당시의 시가 총액을 사용하여 200위 상위를 상위기업 하위를 하위기업으로 나누었다. 이를 통하여 상위 기업은 1, 하위기업은 0 이라는 기업 규모 더미변수를 생성하였다.

IT/비 IT기업 분류 방법은 한국거래소 홈페이지에서 업종 구분을 KRX IT, (유)코스피200 정보통신, (코)IT H/W, (코)IT S/W and SVC, (코)IT 부품, (코)코스닥 IT 등으로 체크하고 여기에 등재된 기업을 IT기업으로 그 이외의 기업을 비 IT기업으로 분류하였고 '산업변수'라는 더미변수를 생성하여 IT기업은 1, 비 IT기업은 0으로 설정하였다. 마지막으로, 코스피/코스닥 기업으로 분류하여 '시장변수'라는 더미변수를 생성하였고 코스닥 기업은 1, 코스피 기업은 0값으로 설정하였다. 아래 표는 대상기업의 분류표이다.

<표 3> 기업 분류표

기업규모		산업변수		시장변수		총
상위기업(1)	120	IT(1)	121	코스피(1)	141	252
하위기업(0)	132	비 IT(0)	131	코스닥(0)	111	

V. 분석 및 결과

본 연구에서 사용된 데이터는 총 252개의 기업 데이터를 100일간 추출 하였으며, 총 데이터 구성은 25,200개이다. 각 변수의 평균값은 주식 수익률, KRX100, 노출 빈도수, 긍정어 수, 부정어, 기업 규모, 산업변수, 시장변수 순으로 각각 51,989, 4,047, 182, 46, 17, 0.48, 0.48, 0.56이다. <표 4>에 기초 통계량이 나타나 있다. <표 4>의 SNS 관련 변수들은 증감율로 변화시키기 이전의 수준변수들에 대한 값이다. 여기서 주목해서 보아야 할 것 중의 하나는 긍정어 수와 노출 빈도수와의 상관관계가 0.447로 높은 편이라는 것이다. 이는 이 후의 회귀 분석에 다중공선성이 있을 수도 있음을 의미하기도 한다. 다중공선성이 있으면 일반적으로 회귀분석에서 계수 유의성이 떨어지게 되어, 실제로는 연관성이 있는 변수도 설명력이 없게 나올 가능성이 있다. 그런데, 회귀분석시 VIF 분석에서 모두 1 정도의 낮은 값을 보였기 때문에, 본 연구에서는 다중공선성을 염려할 필요가 없음을 확인할 수 있었다.

또 한 가지 살펴보아야 할 것은 각 더미변수간의 상관관계이다. 상위기업들이 KOSPI, 하위기업들이 KOSDAQ에 상장되어 있을 가능성으로 인해, 두 개의 더미변수가 사실상 같을 수도 있다는 우려가 있었으나, 실제 상관계수는 0.622로 어느 정도 차별성은 존재하고 있었다. 오히려 이 두 개의 더미변수보다는 시장변수와 산업변수의 상관관계가 더 높게 나타났다. 그런데, 뒤의 분석에서 보면 시장변수의 조절효과가 산업변수와는 다소 다르게 나타나고 있음을 확인하였기 때문에, 더미변수간의 차별성은 어느 정도 확보된 것으로 파악되었다. 또한 각 더미변수를 각각 대입한 것이기 때문에 다중공선성을 우려할 필요는 없다고 하겠다.

<표 4> 기초 통계량 분석

	주식 수익률	KRX100	빈도수	긍정어	부정어	기업규모	산업변수	시장변수
평균	51,989	4,047	182	46	17	0.48	0.48	0.56
표준편차	112,856	102	751	173	177	0.499	0.5	0.496
주식 수익률	1							
KRX100	0.369**	1						
빈도수	0.077**	0.014*	1					
긍정어	0.055**	0.005	0.447**	1				
부정어	-0.003	-0.002	0.239**	0.251**	1			
기업규모	-0.005	0.000	-0.038**	0.014*	0.078**	1		
산업변수	0.004	0.000	0.031**	0.002	-0.064**	-0.503**	1	
시장변수	-0.010	0.000	-0.029**	0.001	0.065**	0.622**	-0.731**	1

<표 5> Base Model 회귀분석

종속변수	독립변수	Co eff.	SE	T	p	VIF
주식 수익률 증감률	상수	0.029	0.013	2.294	0.0218**	
	KRX100	79.99	1.270	62.969	0.0000***	1.000
	빈도수	0.049	0.005	9.604	0.0000***	1.278
	긍정어	0.012	0.003	4.737	0.0000***	1.285
	부정어	-0.012	0.003	-4.117	0.0000***	1.091
	통계량	R ² = 0.142111, F = 1043.404, p = 0.0000				

5.1 Base Model 회귀분석

<표 5>는 주식 수익률과 노출 빈도수, 긍정어 수, 부정어 수의 증가율간의 상관관계를 분석한 결과이다. 독립변수인 SNS 관련 변수들의 증가율이 기본적으로 주식 수익률과의 상관관계가 모두 유의미하게 나온 것을 알 수 있다. 앞서 언급한대로 긍정어 수와 노출 빈도수의 상관관계가 높아 다중 공성선의 가능성을 보였지만, 이 두 변수 모두 유의미하게 나온 것을 알 수 있다. 따라서 만약, 다중공성선의 가능성을 배제 한다면 두 변수는 지금보다도 더 유의미한 결과 값을 얻을 수 있을 것이다. 즉, 긍정어 수와 빈도수는 주식 수익률에 영향 관계가 있다는 것이 확실하다는 것이다. 위 회귀분석의 R²가 낮게 나오는 것은 level 데이터가 아니라 차분을 거친 rate 데이터

이기 때문이라 볼 수 있으며, 시계열 데이터 분석에서 흔히 일어날 수 있는 일이다.

개별 단위로 살펴보면 주식 수익률과 빈도수의 증가율은 긍정적인 상관관계를 나타내고 있다. 이는 주식 가격과 긍정어 수에도 같은 관계가 있음을 알 수 있다. 반대로, 주식 수익률과 부정어 수의 증가율은 부정적인 상관관계가 유의미하게 나타나고 있음을 확인 할 수 있다. Base model의 결과에서 알 수 있듯이, 기업의 SNS상에 긍정 및 부정적인 노출의 변화는 해당 기업의 주식 수익률과 상관관계가 존재한다는 것을 알 수 있었다. 즉, SNS상에서 특정 기업의 노출빈도수와 긍정어의 증가는 해당 기업의 주식 수익률과 긍정적인 상관관계가 그리고 부정어 수의 증가는 부정적인 상관관계가 존재한다는 것을 알 수 있었다.

5.2 산업군 조절효과분석

<표 6>은 IT기업에서 SNS와의 관계를 분석한 결과이다. 전체 기업에서는 빈도수와 긍정어 수의 증가율은 주식 수익률과 긍정적인 상관관계에 있으며, 부정어 수의 증가율 부정적 상관관계가 있는 것으로 나타났다. IT기업만을 따로 분류하여 알아보기 위해 생성한 산업변수를 추가하여 분석한 결과에서는 긍정어 수와 부정어 수의 증가율은 IT기업과 상관관계가 없는 것으로 나타났지만, 빈도수의 증가율과 주식 수익률은 긍정적 상관관계가 있는 것으로 나타났다.

현재, IT기기는 상대적으로 다른 기기들보다 사람들이 자주 쓰는 제품으로 IT기기의 성능에 대한 이용자들의 민감도는 상당히 높은 편이다. 따라서 이러한 제품에 대한 높은 민감도는 이용자들의 관심으로 이어지고 이러한 관심은 IT기기에 대한 이용자들의 오피니언으로 연결이 된다. 이러한 개인의 오피니언은 이용자들끼리의 오피니언 공유를 위해 온, 오프라인상에서 그 노출이 자연스럽게 빈번해 지고, 이용자들의 욕구를 충족시켜야 하는 기업의 입장에서는 이러한 오피니언을 제품에 반영하기 위해 노력할 것이다. 따라서 IT기업의 입장에서는 SNS상의 여론에 의해 형성된 요인이 다른 요인보다 상대적으로 크기 때문에 IT기업의 주식 수익률에 영향을 미친다고 판단된다.

<표 7>은 IT기업과 비 IT기업을 좀 더 자세히 비교하기 위해 실행한 회귀분석 결과이다. 이 표를 보면 IT기업과 비 IT기업 모두 빈도수, 긍정어 수, 부정어 수의 증가율과 상관관계가 존재하는 것으로 분석되었다. IT기업과 비 IT기업에서 빈도수, 긍정어 수에서는 긍정적인 상관관계가 나타났고, 부정어 수와는 부정적인 상관관계가 나타났다. IT기업은 빈도수와 주식 수익률과의 상관관계가 가장 큰 것으로 나타난 반면, 비 IT기업은 긍정어 수와 상관관계가 가장 큰 것으로 분석되었다. 전체적으로 볼 때 IT기업이나 비 IT기업 모두 SNS상에서 사람들에게 의해 회자되었을 때 해당 기업의 주식 수익률과 상관관계가 존재한다는 것을 알 수 있다. IT기업과 비 IT기업은 SNS상에서 긍정어 수가 많으면 당연히 해당 기업의 주식 수익률과 긍정적인 상관관계가 그리고 부정어 수가 많으면 부정적인 상관관계가 존재한다.

이 결과에서 중요한 것은 빈도수에 대한 IT기업과 비 IT기업의 주식 수익률과의 상관관계이다. 두 그룹에서 그 강도의 차이가 존재하는데 이는 비 IT기업과 비교했을 때, IT기업의 경우 SNS상에 빈번하게 노출되면 비 IT기업의 주식 수익률 보다 강한 상관관계가 존재하는 것으로 보여진다. 이는 앞서 언급했듯이 IT산업에서 IT기기에 대한 이용자들의 관심이 크기 때문에 SNS상에 IT기업의 노출은 비 IT기업이 노출되는 파급효과보다 크다고 해석할 수 있다.

<표 6> 산업군 조절효과 회귀분석

종속변수	독립변수	계수	SE	t	P
주식 수익률 증감률	상수	0.044	0.018	2.471	0.0135**
	KRX100	79.918	1.269	62.966	0.0000***
	빈도수	0.013	0.007	1.766	0.0774*
	긍정어	0.013	0.004	3.723	0.0002***
	부정어	-0.008	0.003	-2.480	0.0132**
	빈도수×산업변수	0.071	0.010	6.900	0.0000***
	긍정어×산업변수	-0.005	0.005	-0.869	0.3849
	부정어×산업변수	-0.008	0.008	-1.047	0.2951
	산업변수	-0.033	0.026	-1.274	0.2027
통계량	R ² = 0.143888, F = 529.2350, p = 0.0000				

<표 7> IT기업과 비 IT기업 회귀분석

종속변수	독립변수	Co eff.	SE	T	p
주식 수익률 증감률	상수	0.044	0.018	2.471	0.0135**
	KRX100	79.918	1.270	62.966	0.0000***
	빈도수×IT	0.084	0.007	11.485	0.0000***
	긍정어×IT	0.009	0.004	2.260	0.0238**
	부정어×IT	-0.016	0.007	-2.353	0.0186**
	빈도수×비 IT	0.013	0.007	1.766	0.0774*
	긍정어×비 IT	0.013	0.004	3.723	0.0002***
	부정어×비 IT	-0.008	0.003	-2.480	0.0132**
	산업변수	-0.033	0.025	-1.274	0.2027
	통계량	R ² = 0.143888, F = 529.2350, p = 0.0000			

5.3 기업규모 조절효과분석

<표 8>은 상위기업과 SNS와의 관계를 분석한 결과이다. 전체 기업에서는 빈도수, 긍정어 수의 증가율과 주식 수익률과의 관계가 긍정적인 상관관계가 있다는 것으로 나타났지만, 부정어 수의 증가율과는 관계가 없는 것으로 나타났다. 반면, 상위기업을 따로 분류하여 알아보기 위해 생성한 기업규모변수를 추가하여 확인해보면, 하위기업의 SNS 노출 빈도수와 주식 수익률은 상관관계가 존재하는 반면, 긍정어 수와 부정어 수와는 관계가 없는 것으로 나타났다.

상위기업의 경우, 주식 수익률에 영향을 미치는

다른 요인들이 SNS상의 여론에 의해 형성된 요인보다 상대적으로 강하기 때문에 유의미하지 않게 나온 반면, 하위기업의 경우에는 주식 수익률에 영향을 주는 다른 요인보다 SNS상의 여론에 의해 형성된 요인이 상위기업보다 상대적으로 강하기 때문에 하위기업이 더 유의미하게 나온 것으로 판단된다. 즉, 상위기업은 그 기업 자체 이미지와 사업 실적 등 주식 수익률에 영향을 미치는 요인들이 매우 강하기 때문에 SNS상에서 형성된 여론이 해당 상위기업의 주식 수익률에 영향을 미치지 못하는 것이다. 반면, 상대적으로 인지도가 낮은 하위기업의 경우, 기업 자체의 이미지가 확고하지 않기 때문에 SNS상에 형성된 여론이 주식 수익률에 영향을 줄 수 있을 것이다.

<표 8> 기업규모 조절효과 회귀분석

종속변수	독립변수	Co eff.	SE	t	p
주식 수익률 증감률	상수	0.008	0.018	0.455	0.6494
	KRX100	79.894	1.269	62.955	0.0000***
	빈도수	0.071	0.006	10.993	0.0000***
	긍정어	0.016	0.004	4.172	0.0000***
	부정어	-0.005	0.006	-0.917	0.3591
	빈도수×기업규모	-0.066	0.011	-6.172	0.0000***
	긍정어×기업규모	-0.005	0.005	-0.947	0.3437
	부정어×기업규모	-0.005	0.006	-0.719	0.4720
	기업규모	0.042	0.026	1.646	0.0999*
	통계량	R ² = 0.144128, F = 530.2675, p = 0.0000			

<표 9> 상위기업과 하위기업 회귀분석

종속변수	독립변수	Co eff.	SE	t	p
주식 수익률 증감률	상수	0.008	0.018	0.455	0.6494
	KRX100	79.894	1.269	62.955	0.0000***
	빈도수×상위기업	0.005	0.009	0.550	0.5825
	긍정어×상위기업	0.011	0.004	3.109	0.0019**
	부정어×상위기업	-0.010	0.004	-2.802	0.0051**
	빈도수×하위기업	0.071	0.007	10.993	0.0000***
	긍정어×하위기업	0.016	0.004	4.172	0.0000***
	부정어×하위기업	-0.005	0.006	-0.917	0.3591
	기업규모	0.042	0.026	1.646	0.0999*
통계량	R ² = 0.144128, F = 530.2675, p = 0.0000				

<표 9>는 상위기업과 하위기업들 좀 더 자세히 알아보기 위해 실행한 회귀분석 결과이다. 이 표를 보면 상위기업에서는 긍정어 수 증가율과 주식 수익률의 상관관계는 긍정적인 상관관계, 부정어 수 증가율과 주식 수익률과는 부정적인 상관관계가 존재하였고, 하위기업에서는 빈도수 증가율과 긍정어 수 증가율에서 긍정적인 상관관계가 존재 하였다.

전체적으로 상위기업 보다는 하위기업이 SNS 파급효과가 큰 것으로 나타났는데 이는 상위기업에 구성된 기업들은 비 IT기업이 많고 하위기업에 구성된 기업은 IT기업이 많은 이유에 기인한다. 상위기업의 경우 긍정어의 증가율이 높을수록 주식 수익률과 긍정적인 상관관계가, 부정어 수의 증가율이 높을수록 부정적인 상관관계가 나타났다. 이 분석에서 눈여겨볼 점은 노출 빈도수인데 앞서 언급한대로 상위기업은 이미 그 자체로 일반인들에게 인지도가 있기 때문에 SNS 상의 노출빈도수와 주식 수익률에는 상관관계가 없는 것으로 해석이 된다. 즉 이미 친숙한 기업의 이름이 SNS상에서 노출된 것 자체를 사람들은 의례 있는 일이라 생각한다. 따라서 이로 인해 노출 빈도수는 주식 수익률과의 상관관계가 존재하지 않는다는 것이다. 반면 하위기업의 경우는 상위기업과 비교해 낮은 인지도로 인해 SNS상에 빈번한 노출 빈도수는 그 자체로 사람들의 관심

을 불러일으키고 해당 기업에 대한 긍정적인 표현이 많으면 그 기업의 주식 수익률과 긍정적인 상관관계가 존재하는 것이다.

반면 하위기업에서 부정어수가 유의미하지 않게 나온 이유는 인지도가 낮은 하위기업의 SNS 상에 빈번한 노출 빈도수는 그 자체로 사람들의 관심을 불러일으키지만 부정어수가 많으면 사람들은 그 기업에 대한 주식 수익률과 관계를 지어 의식하는 것이 아니라 그 기업을 그대로 무시하기 때문에 유의미하지 않게 나온 것으로 해석할 수 있다.

5.4 주식시장별 조절효과분석

<표 10>은 코스피 기업과 SNS와의 관계를 분석한 결과이다. 전체 기업에서는 빈도수, 긍정어 수, 부정어 수 증가율 모두 주식 수익률과 상관관계가 존재하고 이 중 빈도수와 긍정어 수 증가율과는 긍정적인 상관관계가, 부정어 수 증가율과는 부정적인 상관관계가 존재한다는 것을 알 수 있다.

한편, 코스피 기업을 따로 분류하여 알아보기 위해 생성한 시장변수를 추가하여 확인해보면, 빈도수와 긍정어 수에서 코스닥 기업이 유의미하게 나타났는데, 이는 상위기업 회귀분석의 결과와 비슷하다. 그 이유는, 코스피와 코스닥 분류

<표 10> 시장별 조절효과 회귀분석

종속변수	독립변수	Co eff.	SE	t	p
주식 수익률 증감률	상수	0.016	0.019	0.843	0.3994
	KRX100	79.990	1.269	63.048	0.0000***
	빈도수	0.076	0.007	10.835	0.0000***
	긍정어	0.019	0.004	4.725	0.0000***
	부정어	-0.013	0.007	-1.996	0.0459**
	빈도수×시장변수	-0.064	0.010	-6.184	0.0000***
	긍정어×시장변수	-0.012	0.005	-2.219	0.0265**
	부정어×시장변수	0.005	0.008	0.701	0.4832
	시장변수	0.023	0.026	0.879	0.3792
통계량	R ² = 0.144458, F = 531.6866, p = 0.0000				

<표 11> 코스피와 코스닥 기업 회귀분석

종속변수	독립변수	Co eff.	SE	t	p
주식 수익률 증감률	상수	0.016	0.019	0.843	0.3994
	KRX100	79.990	1.269	63.048	0.0000***
	빈도수×코스피	0.014	0.007	1.918	0.0551*
	긍정어×코스피	0.007	0.003	2.197	0.0281**
	부정어×코스피	-0.008	0.003	-2.401	0.0164**
	빈도수×코스닥	0.078	0.007	10.835	0.0000***
	긍정어×코스닥	0.019	0.004	4.725	0.0000***
	부정어×코스닥	-0.013	0.007	-1.996	0.0459**
	시장변수	0.023	0.026	0.879	0.3792
통계량	R ² = 0.144458, F = 531.6866, p = 0.0000				

는 그 분류 특성상 상위기업/하위기업 분류 방법과 유사하기 때문이다. 마찬가지로 코스피 기업의 경우 SNS에서 형성된 여론의 요인보다 상대적으로 다른 강한 요인들이 주식 수익률에 영향을 미치기 때문에 코스피 기업이 유의미하지 않게 나온 반면, 주식 수익률에 영향을 주는 다른 요인보다 여론에 의해 형성된 요인에 의해 상대적으로 영향을 많이 받는 코스닥 기업이 유의미하게 나타났다. 또한, 부정어 보다는 긍정어가 코스닥 기업에 영향을 미치는 것으로 나타났다.

<표 11>은 코스피 기업과 코스닥 기업을 좀더 자세하게 알아보기 위해 실행한 회귀분석 결과이다. 이 표를 보면 코스닥과 코스피 기업 모두 빈도수, 긍정어 수, 부정어 수의 증가율 모두 상관

관계가 존재하는 것으로 분석되었다. 코스피 기업과 코스닥 기업 모두 빈도수, 긍정어 증가율에서는 긍정적인 상관관계가 나타났고, 부정어 증가율은 부정적인 상관관계가 나타났다. 전체적으로 코스피 기업보다 코스닥 기업의 경우가 SNS의 파급효과를 더 크게 받는 것으로 나타나는데, 이는 코스피 기업의 대부분이 비 IT기업들이고 코스닥으로 구성된 기업 대부분이 IT기업으로 구성된 것으로 기인한다.

코스피 기업은 상위기업과 마찬가지로 그 기업 자체의 이미지가 강하기 때문에 코스닥 기업과 비교했을 때 빈도수 증가율과 주식 수익률과 상관관계는 약하게 나타났다. 코스닥 기업 또한 하위기업과 마찬가지로 노출 빈도수는 그 자체

로 사람들의 관심을 불러일으키고 해당 기업에 대한 긍정적인 표현이 많으면 그 기업의 주식 수익률과 긍정적인 상관관계 존재 하는 것이다. 하지만 하위기업과는 달리 코스피 기업은 부정어 수가 유의미하게 나왔는데 이는 코스피 기업으로 분류된 기업의 대부분이 IT기업인 점을 감안하면 IT기업의 대한 긍정적 분위기와 관심으로 인해 부정어 수의 증가율도 그 기업의 주식 수익률과 상관관계가 존재하는 것이다.

5.5 Granger 인과관계 분석

앞선 회귀분석 결과들로부터 SNS상에서 기업의 특정 분류에 따라 차이가 존재 하지만, 기업의 노출 빈도수, 긍정어 수 그리고 부정어 수의 증가율은 주식 수익률과 영향관계가 있다고 판단 된다. 이 후 검증한 부분은 독립변수와 종속변수 간에 인과관계에 대한 것이다. 이 분석을 위하여 앞서 언급한대로 Granger Causality Test를 eViews 통계 프로그램을 통해 분석하였고 아래의 표는 Granger Causality를 분석한 결과 표이다. 이 결과표를 통해 SNS상의 노출빈도수, 긍정어수 그리고 부정어수 각각이 실질적으로 주식 수익률과 인과관계가 있는지 여부를 판단할 것이다. 시계열 분석에서는 Akaike Criterion Index(AIC)의 값을 가장 작게 해주는 시차의 개수를 최적 모형으로 제시하기도 한다. 본 연구에서는 시차의 개

수가 달라짐에 따라 통계 결과가 어떻게 변화하는지를 살펴보기 위해 시차 1부터 4까지를 결과 표에 포함시켰다.

<표 12>는 특정기업의 SNS 노출 빈도수 증가율과 주식 수익률과의 인과관계를 분석한 결과이다. 위 표의 결과 값을 통해 알 수 있듯이 노출 빈도수는 1, 2, 3, 4시차 모두 유의미하게 나타났다. 이는 곧 특정 기업의 SNS 노출 빈도수는 해당 기업의 주식 수익률 변화에 영향을 미친다는 것을 의미한다. 하지만, 반대로 주식 수익률은 모든 시차에서 유의미하지 않게 나왔기 때문에 주식 수익률이 노출 빈도수에 영향을 미친다고 볼 수 없다. 따라서 앞서 언급했던 'SNS에서의 노출 빈도수의 증가율은 주식 수익률에 영향을 미칠 것이다'라는 가설을 입증하였다.

<표 13>은 긍정어 수 증가율과 주식 수익률과의 인과관계를 분석한 표이다. <표 13>에서 나타나듯이 긍정어 수는 1, 2시차에서는 영향을 미치지 못하다가 3, 4시차에서 점점 더 유의미한 결과를 나타낸다. 이는 'SNS에서의 긍정적 노출은 주식 수익률에 영향을 미칠 것이다'라는 가설을 입증하였지만, 긍정적 노출의 효과는 노출 빈도수로 인해 나타나는 효과보다 좀 더 시간이 지난 후에 나타나는 것으로 해석할 수 있다. 반면, 주식 수익률은 긍정어 수와의 인과관계 분석에서는 모든 시차에서 영향을 주지 않는 것으로 나타났다.

<표 12> 빈도수와 주식 수익률과의 Granger Causality

	lags	Null Hypothesis	Obs	F	P
노출빈도수와 주식 수익률과의 인과관계	1	빈도수 does not Granger Cause 주식 수익률	24948	10.0012	0.00157***
		주식 수익률 dose not Granger Cause 빈도수		0.84482	0.35803
	2	빈도수 does not Granger Cause 주식 수익률	24696	7.86271	0.00039***
		주식 수익률 dose not Granger Cause 빈도수		0.42357	0.65471
	3	빈도수 does not Granger Cause 주식 수익률	24444	7.58172	4.6E-05***
		주식 수익률 dose not Granger Cause 빈도수		1.92340	0.12337
	4	빈도수 does not Granger Cause 주식 수익률	24192	11.0560	5.9E-09***
		주식 수익률 dose not Granger Cause 빈도수		2.88625	0.02110**

<표 13> 긍정어수와 주식 수익률과의 Granger Causality

	lags	Null Hypothesis	Obs	F	P
긍정어수와 주식 수익률과의 인과관계	1	긍정어 dose not Granger Cause 주식 수익률	24948	2.52411	0.11213
		주식 수익률 dose not Granger Cause 긍정어		1.89481	0.16867
	2	긍정어 dose not Granger Cause 주식 수익률	24696	1.78200	0.16832
		주식 수익률 dose not Granger Cause 긍정어		0.74730	0.47366
	3	긍정어 dose not Granger Cause 주식 수익률	24444	2.16719	0.08963*
		주식 수익률 dose not Granger Cause 긍정어		0.49740	0.68407
	4	긍정어 dose not Granger Cause 주식 수익률	24192	2.99644	0.01748**
		주식 수익률 dose not Granger Cause 긍정어		0.46983	0.75794

<표 14> 부정어와 주식 수익률과의 Granger Causality

	lags	Null Hypothesis	Obs	F	P
부정어수와 주식 수익률과의 인과관계	1	부정어 dose not Granger Cause 주식 수익률	24948	1.52072	0.21752
		주식 수익률 dose not Granger Cause 부정어		3.06679	0.07992*
	2	부정어 dose not Granger Cause 주식 수익률	24696	1.15132	0.31624
		주식 수익률 dose not Granger Cause 부정어		1.28380	0.27700
	3	부정어 dose not Granger Cause 주식 수익률	24444	0.93603	0.42219
		주식 수익률 dose not Granger Cause 부정어		1.22296	0.29954
	4	부정어 dose not Granger Cause 주식 수익률	24192	1.78245	0.12922
		주식 수익률 dose not Granger Cause 부정어		0.91679	0.45294

<표 14>는 부정어 수 증가율과 주식 수익률과의 인과관계를 분석한 결과표이다. 아래 표의 결과 부정어는 주식 수익률에 어떠한 영향도 미치지 못하고 또한, 주식 수익률도 부정어에 영향을 미치지 못하는 것으로 분석된다. 이는 앞서 회귀 분석의 결과를 통해 알아본 바와 같이 부정어와 주식 수익률 사이에는 분명 상관관계가 존재하지만, 주식 수익률의 변화가 SNS에서 부정적 노출에 기인한다는 인과관계로 설명하기에는 부족하다는 것이다.

VI. 결 론

6.1 연구 결과 요약 및 논의

본 연구에서는 국내 기업들을 대상으로 트위터나 페이스북 상의 기업노출이 해당 기업의 주식 수익률에 영향을 미치는지에 대한 존재 여부를 분석 하였고 이들과 주식 수익률과의 인과관계를 분

석하였다. SNS는 비정형 텍스트 속성을 가지고 있으며, 하루에도 수많은 소통 정보가 실시간으로 양산되고 있다. 대량의 SNS 소통 정보를 처리하기 위해 빅데이터 분석기법 중 오픈피언 마이닝 분석을 이용하여 데이터를 정제하였다. 트위터나 페이스북이 주식 수익률에 영향을 미치는 요인으로는 SNS상에서 기업이 노출되는 노출 빈도수, 해당기업에 대한 긍정어 수와 부정어 수로 분류 하였다. 또한, 기업을 IT/비 IT, 상위기업/하위기업, 코스피/코스닥 등 3가지의 카테고리 별로 분류하여 각각에 대한 조절효과를 분석하여 제시하였다. 마지막으로, Granger Causality 테스트를 실시하여 주식 수익률과 요인 간의 인과관계를 분석하여 제시하였다.

분석 결과 SNS상에서 특정 기업의 노출 빈도수와 긍정어 수의 증가율은 해당 기업의 주식 수익률과 긍정적인 상관관계로 나타났고, 부정어 수의 증가율 해당 기업의 주식 수익률과 부정적인 상관관계로 나타났다. 또한, 노출 빈도수와 긍

정어 수의 증가율은 주식 수익률에 영향을 미치는 인과관계인 것으로 분석된 반면, 부정어 수는 주식 수익률과 상관관계가 존재는 하지만 부정어 수가 직접적으로 주식 수익률에 인과관계를 미치는지는 알 수 없는 것으로 분석 되었다.

결론적으로, SNS상에서 어떤 한 기업이 많이 노출된다는 것은 그 만큼 그 기업의 가치가 높아진다는 것을 의미하며 그 기업에 대한 부정적 표현보다는 긍정적인 표현이 그 기업 가치를 조금 더 적절하게 반영한다고 할 수 있겠다. 또한, 이러한 효과는 기본적으로 비 IT기업 보다 IT기업이 더 큰 것으로 나타났으며 상위기업보다는 하위기업이 코스닥 기업보다는 코스피 기업이 더 큰 것으로 나타났다. 특징적으로 SNS효과는 주식 수익률 예측에 있어서 부정적인 반응보다는 긍정적인 반응이 좀 더 큰 영향력을 발휘하는 것으로 나타났다.

6.2 연구의 한계점 및 연구 방향

본 연구를 통해 대중의 오피니언이 기업의 주식 수익률과 깊은 상관관계가 있고 나아가 개인들의 오피니언이 모여 기업의 주식 수익률에 영향을 미칠 수도 있다는 결론을 얻었다. 이것은 우리나라의 잘 발달된 IT 인프라의 역할을 간과할 수 없는데, 언제 어디서나 개인의 의견을 SNS상에 올리고 다른 사람들과 해당 이슈에 대해 소통을 할 수 있는 여건이 주어졌기 때문이다. 이러한 사회적 인프라로 인해 SNS는 대중과 연구자를

연결시켜주는 하나의 플랫폼 역할을 하고 있고, 연구자들은 이러한 플랫폼을 이용해 대중들의 오피니언을 보다 쉽게 이용하여 연구를 진행할 수 있게 되었다. 특히, 수많은 정보로부터 영향을 받는 주식 수익률의 경우, 기존의 주식 수익률예측을 하는 연구에 이러한 대중들의 오피니언을 추가하여 연구하는 것이 보다 높은 정확한 유추를 해 낼 수 있을 것이다.

따라서 본 연구는 주식 수익률을 예측하는 연구에 있어서 주식 수익률에 영향을 주는 또 다른 영역을 검증했다는 데에 의의를 둘 수 있고, 향후 주식 수익률 예측에 관한 연구를 하는데 있어 좀 더 높은 정확도를 이끌어 낼 수 있을 것이라 기대한다.

본 연구에서 주식 수익률 변동성(volatility)에 대한 회귀 분석도 실시하였으나 전반적으로 유의미한 결과를 얻지 못하였다. 또한, SNS에서 발생하는 풍부한 대중들의 감성 어휘를 단지 긍/부정으로 단편화하여 분석하고 이를 주식 수익률에 반영 했다는 한계가 존재한다. 또한 주가를 통하여 분석을 실시하는 다른 연구들과 마찬가지로, 주가에 영향을 미치는 무수히 많은 다른 요인들을 통제하지 못하였다는 점도 한계로 지적할 수 있다. 향후 연구에서는 먼저 주식시장에 특화된 감성 카테고리를 세분화하여 구축하고, 이후 구축된 감성 카테고리를 이용하여 분석한다면, 좀 더 정확한 SNS와 주식 수익률간의 관계를 설명할 수 있을 것이다.

〈References〉

- [1] Adrian, Merv, "Big Data Era," *Teradata Magazine*, 2011.
- [2] Ahn, S.W. and Cho, S.-B., "Stock Prediction Using News Text Mining and Time Series Analysis," *KISS Korea Computer Congress 2010*, Vol. 37, No. 1(C), 2010.
- [3] Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G.A., and Reynar, J., "Building a sentiment summarizer for local service reviews," In *WWW Workshop on NLP in the Information Explosion Era*, 2008.
- [4] Bollen, J., Mao, H., and Zeng, X., "Twitter mood predicts the stock market," *Journal of Computational Science*, Vol. 2, 2011, pp. 1-8.

- [5] ChosunBiz, "Social Media Investment and its Return," 2013. 04. 25.
- [6] Cho, S.W., *The technology of Big Data Era*, KT Advanced Institute of Technology, 2011.
- [7] Ding, Y., Gobinda G. Chowdhury, and Schubert, F., "Bibliometric cartography of information retrieval research by using co-word analysis," *Information processing and management*, Vol. 37, No. 6, 2001, pp. 817-842.
- [8] Fama, E.F. and French, K.R., "Permanent and Temporary Components of Stock Price," *Journal of Political Economy*, Vol. 96, 1988, pp. 246-273.
- [9] Granger, C.W.J., "Investigating Causal Relations by Econometric Models and Cross Spectral Methods," *Econometrica*, 1969, pp. 424-438.
- [10] HyдайFutures, "Wag-the-Dog phenomenon in bond market and swap market," 2008.
- [11] Kaplan, A.M. and Haenlein, M., "Users of the world, unite! The challenges and opportunities of Social Media," *Business horizons*, Vol. 53, 2010, pp. 59-68.
- [12] Kim, H.N., "The trend and implication of Big Data," *Communications Policy*, Vol. 24, No. 19, 2012.
- [13] Kim, J.-O., Lee, S.-S., and Yong, H.-S., "Automatic Classification Scheme of Opinions Written in Korea," *Journal of KIISE: Databases*, Vol. 38, No. 6, 2011.
- [14] Kim, J.S., "Digital Personal Connections and Human Network," *Information and Communication Policy*, Vol. 20, No. 23, 2004, pp. 1-19.
- [15] Kim, J.H., "Role of Financial Economics in Securities Fraud Cases based upon ECMH as a Measurement Method of Disgorgement and Abnormal Returns," *The Korea Commercial Law Association*, Vol. 28, No. 3, 2009.
- [16] Kim, Y.-K. and J.-B. Bae, "The study about the time series dependency of Korea stock returns," *Asian Review of Financial Research*, Vol. 8, 1994.
- [17] Manovich, L., "Trending: The Promises and the Challenges of Big Social Data," *Debates in the Digital Humanities*, 2011.
- [18] Park, K.M., Park, H.G., Park, H.G., and Go, H.D., "The Opinion-mining study in SNS," *Communications of the Korea Information Science Society*, Vol. 29, No. 11, 2011, pp. 54-60.
- [19] Ratchford, B.T., Talukdat, D., and Lee, M.S., "A model of consumer choice of the internet as an information source," *International Journal of Electronic Consumer*, Vol. 5, No. 3, 2001, pp. 7-22.
- [20] Schumaker, R.P. and Chen, H., "A discrete stock price prediction engine based on financial news," *IEEE Computer Society*, Vol. 43, No. 2, 2010, pp. 51-56.
- [21] Sun, T., Youn, S., Wu, G., and Kuntaraporn, M., "Online word-of-mouth(or mouse): An exploration of its antecedents and consequences," *Journal of Computer-Mediated Communication*, Vol. 11, No. 4, article 11, 2006.
- [22] Sung, Y.S., Park, J.Y., and Park, E.A., "The Influence of On-line Word of Mouth Information On Viewing Intention toward Motion Picture," *Advertising Research*, Vol.57, 2002, pp. 31-52.
- [23] The Federation of Korean Information Industries, "The Comprehension and the best practices of the Social Media," 2006.
- [24] ZDNetKorea, "Real-time Search Key Words +SNS Analysis Trend," 2013. 09. 10.
- [25] Zhuang, L., Jing, F., and Zhu, X.Y., "Movie review mining and summarization," In *Proceedings of the 15th ACM international conference on Information and knowledge management*, 2006, pp. 43-50.

◆ About the Authors ◆



Taehwan Kim

Taehwan Kim is in doctoral student in Management Information System (MIS) at the School of Business of Hanyang University. He received a master's degree (2014) and bachelor degree (2010) in the same school. His primary research interests include the productivity of IT, social media, big data analysis and opinion mining.



Woo-Jin Jung

Woo-Jin Jung received his Master of Science in MIS from Hanyang University, Seoul, Korea (2012). He is in a doctoral student in the same school. His research interests are in the area of Privacy and Information Security, IT value measurement and R&D performance measurement. His papers have been published in Asia Pacific Journal of Information Systems and Telecommunications Review.



Sang-Yong Tom Lee

Sang-Yong Tom Lee is a Professor in School of Business, Hanyang University, Seoul, Korea. He was on the faculty of Department of Information Systems, National University of Singapore before joining Hanyang University. He received his Ph.D degree from Texas A&M University (1999). His research interests include economics of information systems, online information privacy, and value of IT. His papers have been published in Management Science, MIS Quarterly, Journal of Management Information Systems, Communications of the ACM, IEEE Transactions on Engineering Management, Information and Management, and others.

Submitted : April 15, 2014

1st revision : May 30, 2014

Accepted : June 17, 2014