

금융감성사전의 구축과 금융감성지수의 유효성*

임 소 연** · 김 인 배***

논문 초록

본 연구는 행동경제학적 시각에서 투자자들의 심리변수가 주가의 움직임에 대해서 설명력을 갖는지 검증하였다. 본 연구의 기여는 다음과 같다. 첫째, 기존 연구에서 주로 사용되었던 Loughran and McDonald(2011)의 금융감성사전을 Word2Vec 모형으로 분석하여 국내 금융 텍스트에 적합한 금융감성사전을 개발하였다. 둘째, 주로 월별 금융감성지수를 구축한 선행연구와는 달리 최초로 일별 금융감성지수를 만들어서 고빈도 주가수익률에 대한 영향력을 살펴보았다. 마지막으로 한국 주식시장의 대표지수로서 코스피200지수에 대한 '코스피 금융감성지수'와 개별주식으로서 대표주인 삼성전자 주가에 대한 '삼성전자 금융감성지수'를 만들어서, 이들의 주가수익률에 대한 설명력을 검증하였다. 그 결과 첫째, 표준화된 일별 코스피 금융감성지수가 1-표준편차만큼 증가하였을 때 일별 코스피 수익률은 약 0.60% 증가함을 보였다. 둘째, 표준화된 일별 삼성전자 금융감성지수가 1-표준편차만큼 증가하였을 때 일별 삼성전자 주가의 초과수익률은 약 0.75% 증가하였다. 셋째, 코스피 금융감성지수는 삼성전자 주가의 초과수익률에 영향력이 없었고, 삼성전자 금융감성지수는 코스피 수익률에 영향력이 없음을 발견하였다. 이는 각 주가에 대해서 그 고유의 금융감성지수가 설명력이 있음을 보인 것으로서, 본 연구 금융감성지수의 정교함을 보인 것이다.

핵심 주제어: 행동경제학, 금융감성사전, 금융감성지수

경제학문헌목록 주제분류: G1

투고 일자: 2022. 12. 19. 심사 및 수정 일자: 2023. 3. 17. 게재 확정 일자: 2023. 4. 24.

* 본 논문에 대해 건설적인 논평을 해주신 익명의 심사자분들께 깊이 감사드립니다.

** 제1저자, 이화여자대학교 대학원 경제학과 박사과정, e-mail: limsoyun2000@naver.com

*** 교신저자, 이화여자대학교 경제학과 교수, e-mail: ibkim@ewha.ac.kr

I. 서 론

2021년은 Covid-19로 인한 세계적인 불황 속에서도 한국의 주식시장이 대체로 상승장을 보였던 시기였다. 실물경제의 불황과는 대조적으로 ‘삼천피’(3000의 코스피지수), ‘천스닥’(1000의 코스닥지수)이란 용어가 대중화될 정도로 주식시장은 호황기를 맞았으나 2022년 2월 러시아-우크라이나 전쟁발발로 인한 주식시장의 하락세로 현재까지 2021년의 수준을 회복하지 못하고 있다. 이처럼 주식시장에서는 주가에 거품이 형성되고 파열되는 일이 빈번하다.

행동경제학은 심리학, 사회학 등 다양한 학문의 관점에서 인간의 행동을 해석하는 경제학의 한 분야로서 Simon(1955) 으로부터 시작되었다.¹⁾ 전통적 경제학은 인간이 합리적으로 행동하고 모든 정보에 대해 접근이 가능하며 반복된 상황 속에서도 같은 선택을 한다고 가정한다. 하지만 Simon은 “제한된 합리성”의 개념을 통해 인간은 제한된 정보 내에서 충분히 좋은 대안을 선택한다고 주장하였다. 이는 현실적으로 모든 경우를 고려해 의사 결정을 내리는 것이 불가능하다는 것과 비합리적 선택이 존재한다는 것을 인정한 것이다(KDI, 2021).²⁾ 이러한 행동경제학의 전제는 주식 가격의 결정에도 적용된다. 즉 주가가 실물경제 펀더멘털의 영향을 받을 뿐 아니라 투자자들의 투자심리에 의해서도 상당 부분 영향을 받는다는 것인데, 실제로 이를 지지하는 실증연구들은 쉽게 찾아볼 수 있다.³⁾

본 연구는 행동경제학적 관점에서 신문기사의 텍스트가 갖는 긍정과 부정의 뉘앙스가 투자자들의 심리에 영향을 미쳐 투자 자체에 영향을 준다고 가정하고, 이러한 심리요인을 지표화시키기 위한 기본자료로서 우선 금융감성사전을 구축하였다. 구체적으로 기존 연구에서 주로 사용되었던 Loughran and McDonald 금융감성사전(이후 ‘L&M 사전’이라고 함)의 한국어로 번역된 단어들과 저자들이 수집한 신문기사 단어들과의 유사도를 측정하는 Word2Vec 머신러닝 기법을 사용하여 국내 금융단어들의 감성을 새롭게 분류하였다. 본 연구의 목적은 한국의 금융 텍스트⁴⁾ 데이터를 분석할

1) Simon으로부터 시작된 행동경제학의 분야는 Kahneman(1979)의 전망이론으로 확장되며 Thaler의 베스트셀러 저서인 「넛지」(2008)를 통해서 대중화되었다.

2) 행동경제학에 대한 보다 상세한 정보는 ‘<https://eiec.kdi.re.kr/material/pageoneView.do?idx=1499>’에서 확인할 수 있다. 행동경제학을 응용한 논문들도 황재홍(2019), 양아정·유세경(2022) 등을 포함하여 다수 존재한다.

3) 대표적으로 Baker and Wurgler(2006), Jiang, Lee, Martin, and Zhou(2019) 등 많은 관련 연구들이 존재한다.

수 있는 기초 자료인 금융감성사전을 구축하며, 이를 이용하여 한국의 주식시장을 대표하는 ‘코스피 금융감성지수’와 개별주식을 대표하는 ‘삼성전자 금융감성지수’를 만들어서 이로써 한국 주식시장의 움직임에 대한 설명력을 실험적으로 보이는 것이다.

본 연구의 의의는 다음과 같다. 첫째, 관련 해외연구에서는 L&M 사전이 많이 이용되어왔으나 본 연구에서는 한국어 텍스트 분석에 적합한 금융감성사전을 국내 연구 최초로 구축하였다. 둘째, 기존의 선행연구에서 제시된 금융감성지수는 모두 월별 데이터인 반면 본 연구에서 구축된 금융감성지수는 일별 데이터이기에, 금융감성지수가 주가의 변화에 미치는 단기적 영향을 포착할 수 있었다. 셋째, 금융감성지수로서 주식시장 전체에 대한 코스피 금융감성지수와 개별 주식에 대한 금융감성지수(삼성전자 금융감성지수)를 만들어서, 이 두 가지 금융감성지수의 주가에 대한 설명력을 검증하였다. 먼저, 일별 코스피 금융감성지수가 1-표준편차만큼 증가할 때 일별 코스피 수익률은 0.60% 증가한 반면, 삼성전자 금융감성지수는 코스피 수익률에 대한 유의한 설명력이 없었다. 또한, 일별 삼성전자 금융감성지수가 1-표준편차만큼 증가할 때 일별 삼성전자 주가의 초과수익률은 0.75% 증가하였으나 코스피 금융감성지수는 삼성전자 주가의 초과수익률에 영향력이 없었다. 이는 각 주가에 대해 그 고유의 금융감성지수가 설명력이 있음을 보인 것으로서, 본 연구의 금융감성지수가 정교하게 구축되었음을 시사한다.

본 연구는 다음과 같이 구성된다. 제Ⅱ절에서는 금융감성사전의 구축 및 감성지수와 주가수익률 간의 관계를 분석한 기존 연구를 살펴보았다. 제Ⅲ절에서는 본 연구에 사용한 데이터 및 방법론을 제시한다. 제Ⅳ절에서는 금융감성지수가 주가에 미치는 영향을 분석한다. 마지막 제Ⅴ절에서는 연구결과의 요약 및 한계점, 그리고 향후 연구방향에 대해 서술한다.

Ⅱ. 선행연구

1. 금융감성사전 관련 연구

지금까지 많은 연구자들이 금융 관련 텍스트를 객관적으로 분석하기 위해 필수적인

4) ‘텍스트’는 텍스트 분석에서 분석대상이 되는 것을 이르며, 신문기사, 보고서, Social Media에서 언급된 글 등을 이른다.

금융감성사전의 구축을 위하여 노력해왔다. 기존에 이 분야의 대표적인 연구로서 Loughran and McDonald (2011)은 1994년부터 2008까지 Electronic Data Gathering, Analysis, and Retrieval system(EDGAR) 사이트에 공시된 기업보고서 10-Ks의 텍스트 데이터를 기반으로, 단어를 긍정적(positive), 부정적(negative), 불확실한(uncertainty), 법적인(litigious), 강한 어조의(strong modal), 약한 어조의(weak modal) 표현 등 총 6개 그룹으로 분류하여 금융감성사전(L&M 사전)을 구축하였다. 특히 이 금융감성사전은 기존의 텍스트 분석에서 사용되던 하버드 대학 심리사전(Harvard IV 사전)의 부정어로 정의된 단어가 금융적 문맥에서는 긍정어나 중립어로 정의될 수 있는 약점을 극복하였다. 이후 L&M 사전은 추후 연구들에서 활발히 사용되었다(Nguyen and Huynh, 2022; Frankel, Jennings, and Lee, 2022).

텍스트 분석을 위한 도구로서 금융감성사전의 필요성을 인지한 연구자들은 주로 머신러닝 및 딥러닝 기법을 통해 금융감성사전을 개발해왔다. Chen, Huang, and Chen(2018)은 기존의 L&M 사전이 금융 SNS 텍스트 데이터의 특징을 잘 반영하지 못한다는 점을 지적하며 이를 보완하였다. 구체적인 방법으로서 딥러닝 기법의 하나인 합성곱 신경망(Covolutional Neural Network)을 이용하여 금융 SNS 텍스트의 고유한 특성이 담긴 새로운 금융감성사전(NTUSD-Fin)을 만들었다. Palmer, Roeder, and Muntermann(2021)은 다우존스 산업평균지수에 대한 애널리스트 공시보고서를 대상으로 불용어와 중복값을 제거하며 단어의 감성을 긍정, 부정 그리고 중립의 감성으로 분류한 뒤, 전문가가 사전을 평가하는 방식을 통해 ‘금융애널리스트사전’을 구축하였다. Katayama, Kino, and Tsuda(2019)은 지역 경제 조사인 ‘Economy Watchers Survey’의 텍스트 데이터를 일본 교토대학의 정보학과 대학원에서 개발한 토큰나이저(MeCab)을 이용하여 형태소 단위로 쪼개고, 이 단어들을 머신러닝 기법인 Word2Vec을 통해 벡터화시켰다. 나아가 딥러닝 기법인 장-단기 메모리(Long Short-Term Memory)와 합성곱 신경망으로써 단어를 긍정어와 부정어로 분류하여 극성 사전(Polarity Dictionary)을 구축하였다. Bian, Jia, Li, and Yan(2021)은 중국에서 개발된 감성사전인 HOWNET, DLUTSD, NTUSD에 포함된 단어들 중 기업 공시보고서 등에 나오는 금융 관련 텍스트만을 선택하여 최종적으로 1,107개의 긍정어와 1,488개의 부정어로 구성된 중국금융감성사전(Chinese Financial Sentiment Dictionary)을 만들었다. Du, Huang, Wermers, and Wu(2022)은 기업 관련 신문 기사를 무작위로 추출하여 4명의 대학원생이 독립적으로 기사를 읽게 한 후, 기사에 포함된 단어가 만장일치로 긍정어, 부정어, 정치어⁵⁾로 구분될 때만 그 단어를 선택

하는 방법을 반복시행하여 2,235개의 긍정어와 2,986개의 부정어, 1,439개의 정치어로 된 금융감성사전을 구축하였다.

금융감성사전 구축에 대한 국내 연구는 매우 찾아보기 힘든데, 그 중 유일한 연구인 조수지·김홍규·양철원(2021)은 16,457개의 애널리스트 보고서를 분석대상으로 하여 48개의 긍정어와 47개의 부정어로 재무분석에 쓰일 수 있는 감성사전(Korean SEntiment Lexicon for Finance ;KOSELF)을 개발하였다. KOSELF는 기존 해외의 LM 사전 및 Harvard IV 사전, 그리고 국내 사전으로서 군산대학교 KNU 감성사전과 서울대학교의 KOSAC보다 우수하게 목표주가의 변화를 예측하였으나 전체 사전의 규모가 작다는 한계점이 있다.

2. 감성지수와 주가수익률 간의 관계를 분석한 연구

Baker and Wurgler(2006)는 전통적 경제학에서 설명할 수 없었던 비이성적 투자를 설명하기 위하여 투자자 심리지수를 이용하였고, 투자자 심리지수는 기업의 크기와 기업연령, 위험, 수익성, 배당정책, 성장 가능성 등의 변수를 이용하여 회귀분석을 통해 구축되었다. Baker and Wurgler 논문은 학계에 큰 반향을 일으켰고 이후 투자자 심리를 이용한 분석이 활발하게 이루어졌다(Chan, Durand, Khuu, and Smales, 2017; Zhou, 2018; Piccoli and Chaudhury, 2018). Aissia(2016)은 Baker and Wurgler의 투자자 심리지수를 응용하여 자국의 투자자 심리지수와 외국의 투자자 심리지수로 구분하였고, 프랑스 CAC지수의 초과수익률에 대한 이 두 투자자 심리지수의 설명력을 검증하였다. 저자는 Fama-French 3 factor CAPM에 자국의 투자자 심리지수와 외국의 투자자 심리지수를 설명변수로 추가하여 분석하였고 1달 전의 투자자 심리지수가 기업의 초과수익률에 유의한 음(-)의 영향이 있음을 보였다. Chen, Yu, Jin, and Bao(2019)는 선행연구에서 투자자 심리지수를 구축하기 위해 주로 사용되었던 주식 거래량, 펀드의 순 유입, 새롭게 추가된 플랫폼, 자산간 대체율 등의 변수에 인터넷 금융시장을 반영한 투자자의 수, 인터넷 금융상품의 비율 등 새로운 변수를 추가하여 인터넷 금융 투자자 심리지수(Internet Fiance Investor Sentiment; IFIS)를 만들었다. 이들은 IFIS가 인터넷 금융부문의 수익률에 음(-)의

5) 중국의 정책적 지침과 일관된 내용만을 게재하도록 하는 “Mind Politics”과 연관된 단어들을 말한다. 더 자세한 내용은 Du, Huang, Wermers, and Wu(2022) p.3을 참조한다.

영향력을 가질 뿐만 아니라 중국 인터넷 관련 기업의 주가에 대해 강한 설명력을 갖는다는 것을 발견하였다. 김가람·양희진·류두진(2020)은 국내시장을 대상으로 애널리스트의 투자 의견 변경이 투자자 심리와 금융시장에 영향을 미치는지 살펴보았다.

한편 감성지수로 투자자 심리지수를 대체하여 이로써 주가 흐름에 대한 영향력을 분석한 연구들이 존재하였다. Da, Engelberg, and Gao(2014)은 ‘구글 트렌드(Google Trend)’에서 Harvard IV 사전과 Lasswell Value 사전의 단어를 검색한 후, 그 결과값들에 대한 사람들의 관심도 추이와 연관 단어 등을 이용하여 ‘FEARS’(Financial and Economic Attitudes Revealed by Search) 지수를 구축하였다. 그리고 이 지수를 가지고 미국 S&P500 지수 수익률과 Russell1000 지수 ETF 등 여러 상장지수펀드 수익률에 대한 설명력 유무를 분석하였다. 그 결과 FEARS가 주가수익률과 동기간 음(-)의 관계를 보이고 1달 전과 2달 전의 FEARS가 현재 주가수익률에 양(+)의 영향을 미친다는 것을 보였다. Li, Xie, Chen, Wang, and Deng(2014)은 Harvard IV 사전과 L&M 사전을 이용하여 홍콩 FINET의 금융기사로부터 만든 감성지수가 홍콩 항셱지수의 수익률에 영향을 준다는 것을 머신러닝 기법(Support Vector Machines)을 통해 보였다. Heston and Sinha(2014)는 기존의 Harvard IV 사전, L&M 사전, 그리고 Thomson Reuters Neural Network 자연어 처리 기법을 이용하여 900,000개 이상의 신문기사로부터 감성지수를 만들었고 이것이 실제로 주가수익률의 1일 또는 2일 후의 값에 대한 예측력이 있다는 것을 발견하였다. 특히 이 연구는 기사를 긍정적 기사와 부정적 기사로 구분하였고, 긍정적인 기사가 1주일 이후의 주가수익률을 빠르게 상승시키는 반면 부정적인 기사는 1분기 이후의 주가수익률에 영향을 미치는 것을 보임으로써 뉴스효과(News Effect)를 실증적으로 밝혔다.

또한 다양한 금융 텍스트에 대한 접근으로 금융감성지수를 구축하려는 시도도 있었다. Farina, Parisi, and Pomante(2017)은 경제 블로그의 텍스트를 수집하여 Harvard IV 사전을 기반으로 감성지수를 만들었고 이 지수와 경제 정책 불확실성 지수(Economic Policy Uncertainty Index)의 유효성을 비교하였다. 결과적으로 연구자가 개발한 경제블로그 감성지수가 S&P500 지수에 대해서 유의한 설명력을 지님을 밝혔다. Jiang, Lee, Martin, and Zhou(2019)은 EDGAR 사이트에서 공시된 기업 보고서 등을 이용하여 매니저 감성지수(Manager Sentiment Index)를 구축하였다. 이들은 Baker and Wurgler의 투자자 심리지수와 Da, Engelberg, and Gao(2014)의 FEARS 지수 등과의 비교를 통해 매니저 감성지수가 시장의 초과수익률에 대해 가

장 큰 설명력을 가짐을 보였고, 여러 거시 변수들만을 설명변수로 회귀분석할 때보다 매니저 감성지수를 추가하여 분석할 때 결정계수(R^2)의 값이 커짐을 보였다.

III. 데이터 및 방법론

1. 토큰화와 Word2Vec

우선 신문기사에 대해 텍스트 분석을 하기 위해서는 텍스트를 형태소(morpheme)⁶⁾ 단위로 쪼개는 토큰화 과정을 진행해야 한다. 이를 위해 한국어 형태소 분석기인 KoNLPy를 사용하였다. KoNLPy는 직관적인 함수명⁷⁾을 통해 쉽고 간단하게 사용할 수 있는 형태소 분석기로서 친절하고 상세한 문서⁸⁾를 제공하는 장점을 지닌다(박은정·조성준, 2014). KoNLPy에는 품사 태깅⁹⁾ 클래스로서 꼬꼬마, Komoran, 한나눔, Twitter, MeCab-Ko가 있는데 본 연구에서는 이들 중 MeCab-Ko를 이용하였다. 한국어는 조사와 어미가 발달해서 띄어쓰기로 의미 단위가 나뉘는 어절에 대해서 이에 대응하는 품사를 태깅하는 데 어려움이 있을 뿐 아니라, 불규칙 활용이 발달하였기에 원형을 복원하는 작업이 필요하다(곽진아·이원재, 2016). 따라서 영어 기반의 MeCab을 한국어 특성에 맞게 개발한 MeCab-Ko는 앞에서 언급한 작업에 최적화되었으며 빠른 처리속도와 정확도를 갖고 세분화된 품사 태깅이 가능하므로, 본 연구에서는 MeCab-Ko를 토큰라이저(형태소 분석기)로 이용하였다.

토큰화된 텍스트 데이터를 수치화하기 위해서는 먼저 단어가 순서와 의미를 내포하는 벡터로 표현되어야 하는데 이때 쓰이는 모형이 Word2Vec이다. Word2Vec 모형은 많은 양의 문서 데이터를 학습하며, 문장의 맥락을 통해 단어를 수치화할 수 있어서(김정미·이주홍, 2017) 이를 이용하면 입력된 한 단어와 문맥상 가장 유사한 단어와 유사도¹⁰⁾를 얻을 수 있다. 이 방법을 통해 본 연구에서는 L&M 사전의 번역본을 입력

6) Oxford Languages에 따르면 ‘형태소’는 의미를 가지는 요소로서 더 이상 분석할 수 없는 가장 작은 단어 단위를 이른다.

7) 예시로 MeCab.nouns() 라는 함수는 텍스트에서 명사만 추출하는 함수이며, MeCab.morphs()는 텍스트를 형태소 단위로 구분하는 함수이다.

8) 이에 대해서는 대표적으로 ‘https://konlpy.org/ko/v0.6.0/morph/’ 사이트를 참조할 수 있다.

9) 품사 태깅은 형태소의 뜻과 문맥을 고려하여 형태소에 품사를 매칭시키는 작업이다. 예시로, “가방에 들어가신다”에서 ‘가방’은 NNG(보통명사)와 매치되고, ‘에’는 JKM(부사격 조사), ‘들어가’는 VV(동사), ‘시’는 EPH(존칭 전어말 어미), ‘나다’는 EFN(평서형 종결 어미)와 각각 매치된다.

단어로 설정하여 토큰화된 단어 중에서 가장 유사한 단어와 유사도를 추출하였다.

2. 금융감성사전의 구축

〈Figure 1〉은 금융감성사전의 구축방법을 요약하여 도식화한 것이다. 우선 〈Figure 1〉의 단계①에서 보듯, 일간지인 동아일보와 한겨레, 방송사인 KBS와 SBS, 경제지인 매일경제와 서울경제, 총 6개 사이트에서¹¹⁾ ‘코스피’(KOSPI)로 검색했을 때 나오는 기사들을 데이터 크롤링 기법을 이용하여 수집하였다. 수집된 기사는 2021년 1월 5일부터 2022년 6월 30일까지 총 10,798개의 기사로 구성되었다.¹²⁾ 신문기사는 앞서 설명한 대로 문장을 형태소 단위로 나눌 때 쓰이는 토큰나이저인 MeCab-Ko를 이용하여 일반명사와 동사, 형용사 단위로 구분하였다. 특히 MeCab-Ko는 고유명사와 일반명사를 구분할 수 있기에 문장 내에서 단어 자체의 긍정과 부정을 비교할 때, 중요도가 떨어지는 고유명사는 제거할 수 있는 장점도 있다. 이렇게 신문 기사를 토큰화하여 구성된 단어집은 18,835개의 단어로 구성되었다. 이후 신문의 이름을 반영한 단어들과 MeCab-Ko를 이용해서도 제거되지 않은 고유명사, 한자, 낱자, 숫자, 직위 등 6,510개의 단어를 불용어로 처리한 후, 최종적으로 12,325개의 단어를 얻게 되었다(단계②).

본 연구에서 핵심적으로 다루는 텍스트 분석과 금융감성사전은 이미 해외에서는 활발히 연구된 분야이다. 특히 감성사전은 텍스트가 긍정적인지 부정적인지를 판단하기 위해서 필수적이며, 그 객관성이 핵심이다. 현재 국내에서 이용 가능한 감성사전은 군산대학교의 KNU 감성사전과 서울대학교 언어학과에서 발전시킨 KOSAC이 있다. 하지만 두 감성사전 모두 금융 관련 텍스트를 분석하기엔 한계가 존재한다. 먼저

10) 유사도는 코사인 유사도를 의미하며, 내적공간의 두 벡터 간 각도의 코사인값을 이용하여 벡터 간 유사한 정도를 구한다. 두 벡터의 방향이 완전히 같은 경우 1의 값을 갖는다.

11) 2022년 3월 기자회견보의 ‘2021년 네이버 언론사 랭킹뉴스 매체별 점유율 현황’에 따르면, 본 연구에서 선별된 언론사인 동아일보(16위), 한겨레(19위), KBS(18위), SBS(15위), 매일경제(4위), 서울경제(13위)가 모두 20위권에 속한다. 아쉽게도 유력 일간지인 중앙일보와 조선일보, 유력 경제지인 한국경제의 경우, 현재로부터 1년 전까지의 데이터만 제공하는 등 데이터수집의 한계로 인하여 본 연구에 포함하지 못하였다(<http://www.journalist.or.kr/news/article.html?no=51101>).

12) 분석기간의 시작을 1월 5일로 정한 이유는 연초의 주식시장 휴장일과 로그 ‘차분’으로 계산하는 추가수익률(이하의 회귀분석에서 사용되는 변수)의 계산방법 등을 고려하여 데이터 시점의 알 관성을 유지하기 위함이다.

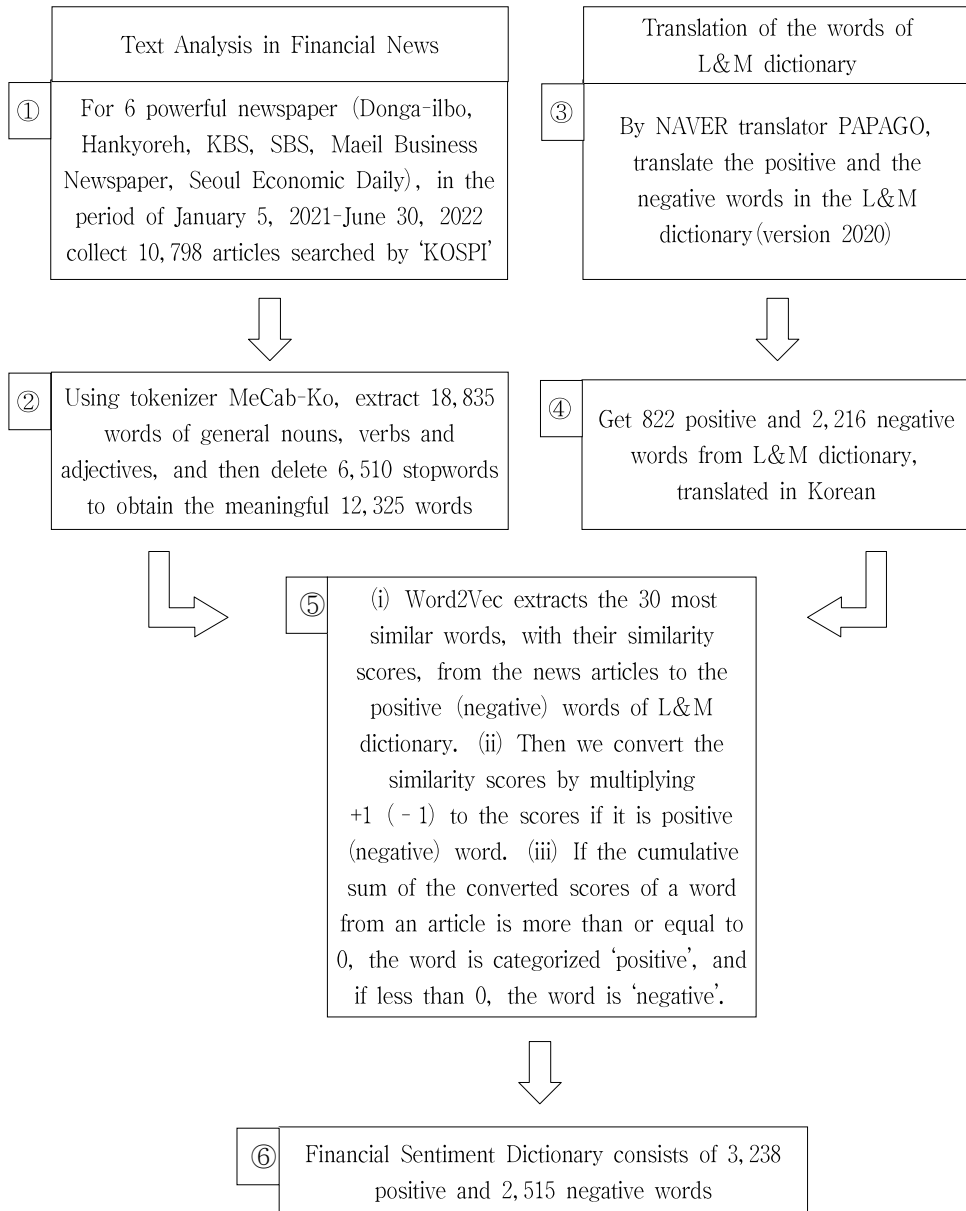
KNU 감성사전은 단어들을 인간의 보편적인 기본 감정 표현을 나타내는 긍정어와 부정어로 구분하며, 이모티콘과 일상용어들로만 구성되었을 뿐 금융용어들은 다루지 않는다. 또한 객관화된 방법을 이용하는 것이 아니라 3명 이상의 평가자들이 각 단어의 긍정과 부정, 중립을 판별하여 단어에 감성을 부여하므로 자의적 판단에 의한 것이라는 비판을 받을 수 있다. KOSAC의 경우는 말뭉치로 구성되어 있어서 단어별 긍정과 부정을 판단하기가 어렵다는 한계가 있다. 뿐만 아니라 이들 말뭉치들도 금융과 직접적인 관련이 없는 일상적 문장들이기에 금융 관련 텍스트를 분석하기엔 크게 부족하다. 추가로 선행연구에서 언급한 조수지 외(2021)의 KOSELF는 KNU 감성사전과 KOSAC의 한계를 지적하며 기업 재무 보고서를 대상으로 감성사전을 구축하였다. 그러나 48개의 긍정어와 47개의 부정어로만 구성되어 있어서 규모면에서 범용 사전의 역할을 하기에는 역부족이다.

본 연구에서 구축한 금융감성사전은 L&M 사전(정확히는 L&M 금융감성사전)의 단어들을 Word2Vec 모형의 입력값으로 이용한다. L&M 사전(2020년 버전)의 단어들은 긍정적(positive), 부정적(negative), 불확실한(uncertainty), 법적인(litigious), 제한하는(constraining) 용어 등 총 5개의 그룹으로 분류되지만, 본 연구에서는 그 중 분명하게 긍정어와 부정어로 분류된 단어들만을 이용하였다. 이후 이 단어들은 네이버 번역기 PAPAGO를 통해 한국어로 번역되어(긍정어 822개, 부정어 2,216개) 앞서 설명한 Word2Vec 모형의 입력값이 되었다(단계③과 ④).

〈Figure 1〉의 단계⑤는 단계①과 ②에서 처리된 단어들과 단계③과 ④를 통해 번역된 단어를 연결하여 분석하는 단계이다. Word2Vec 모형은 입력 단어와 가장 유사한 상위 n개의 단어와 그 각 단어에 해당하는 유사도가 출력된다. 본 연구에서는 상위 30개를 출력대상으로 하였다.¹³⁾ 본 연구의 핵심인 신문기사의 토큰화된 특정 단어가 긍정의 의미를 갖는지 혹은 부정의 의미를 갖는지 판별하려면, 추가적인 작업이 필요하다. 즉 Word2Vec 모형의 입력값으로 L&M 사전의 ‘긍정어’ 1개를 넣으면 신문기사의 토큰화된 단어 중 가장 유사한 30개의 단어가 각각의 유사도와 함께 도출되는데, 이들 단어가 L&M 사전 긍정어의 유사 단어들이므로 해당 유사도에 (+1)을 곱하여 변환한다. 마찬가지로 L&M 사전의 ‘부정어’ 1개를 Word2Vec 모형의 입력값으로 넣으면 역시 신문기사의 토큰화된 단어 중 가장 유사한 30개의 단어가 각각의 유사도와 함께 도출되고, 이들은 L&M 사전 부정어의 유사 단어들이므로 유사도에

13) 상위 30개를 출력할 때 상위 10개, 20개, 40개, 50개일 때보다 긍정어, 부정어 분류가 상식에서 더 부합된 결과가 나왔다.

〈Figure 1〉 Construction of Financial Sentiment Dictionary



(-1)을 곱하여 변환한다. 이 과정을 L&M 사전에 포함된 긍정어 822개와 부정어 2,216개 모두에 대해서 각각 반복 시행한다. 이렇게 구한 모든 유사 단어들과 각 단어의 해당 변환값들에 대해, 만일 신문기사의 토큰화된 한 단어가 이들 유사어와 일치하면 그 해당 변환값들을 모두 더한 누적합을 구하고, 이를 토대로 그 토큰화된 단

어가 긍정어인지 혹은 부정어인지 판단한다.¹⁴⁾ 이 과정을 통해 한 단어에 해당하는 누적 합이 0 이상이면 긍정어, 0 미만이면 부정어로 라벨링하였다. 위와 같은 과정을 통하여 L&M 사전의 번역본의 규모보다 훨씬 확장된 규모의 금융감성사전을 구축할 수 있었다. 따라서 금융과의 관련성이 떨어지는 KNU 감성사전과 KOSAC, 그리고 규모의 한계가 있는 KOSELF와 비교해서, 금융에 특화되어 있을 뿐 아니라 규모면에서도 명실상부한 금융감성사전을 국내 최초로 구축하였다. 〈Figure 1〉의 단계⑥에서 보듯이 본 연구의 금융감성사전은 3,238개의 긍정어와 2,515개의 부정어로 구성된다.

3. 코스피 금융감성지수의 구축

본 항에서는 앞에서 기술한 금융감성사전을 이용하여 코스피 금융감성지수를 만드는 방법을 설명하고자 한다. 구체적으로 Katayama, Kino, and Tsuda (2019)의 방법을 이용하여 금융감성사전 구축할 때와 마찬가지로 앞서 언급한 6개 신문기사 사이트(동아일보, 한겨레, KBS, SBS, 매일경제, 서울경제)로부터 추출한 10,798개 ‘코스피’ 관련 기사들의 감성을 판별하여 이를 지수화하였다.

$$S_{it}^{KOSPINews} = \frac{(+1) * N_{it}^{pos words} + (-1) * N_{it}^{neg words}}{N_{it}^{pos words} + N_{it}^{neg words}} \quad (1)$$

($i = 1 \dots M_t$; M_t = the number of news articles at time t ; $t = 2021.01.05. \sim 2022.06.30.$)

$$S_t^{KOSPI} = \frac{1}{M_t} \sum_{i=1}^{M_t} S_{it}^{KOSPINews} \quad (2)$$

$$Std_S_t^{KOSPI} = \frac{S_t^{KOSPI} - \overline{S_t^{KOSPI}}}{\sigma(S_t^{KOSPI})} \quad (3)$$

14) 예시로 ‘청약’이라는 신문기사의 토큰화된 한 단어는 L&M 사전의 긍정어 ‘아파트’와의 유사도가 0.9, L&M 사전의 부정어 ‘포기’와의 유사도가 0.3, 마지막으로 L&M 사전의 부정어 ‘이사’와의 유사도가 0.2로 나타날 수 있다. 이 경우 토큰화된 단어 ‘청약’의 유사도 변환값들의 누적합은 $0.9 - 0.3 - 0.2 = 0.4$ 이고 이는 양수(+)이므로 본 연구의 금융감성사전에서 ‘청약’은 긍정어로 분류가 된다. 물론 그 누적합이 음수(-)이면 부정어로 분류된다.

식 (1)의 $S_{it}^{KOSPINews}$ 는 t 시점에 게시된 M_t 개의 신문기사 중 i 번째 기사의 감성을 지수화하는 식으로서, 그 기사에 포함된 긍정어의 개수($N_{it}^{pos words}$)에 1의 값을 곱한 것과 부정어의 개수($N_{it}^{neg words}$)에 (-1)의 값을 곱한 것의 합을 총 긍정어와 부정어 개수의 합으로 나눈 것이다. 식 (2)의 S_t^{KOSPI} 는 일별 코스피 금융감성지수로서 t 시점의 코스피 관련 신문기사 감성지수 전체의 평균을 나타낸다. 식 (3)의 $Std_S_t^{KOSPI}$ 는 S_t^{KOSPI} 를 표준화한(standardized) 것으로서 우변의 $\overline{S_t^{KOSPI}}$ 는 S_t^{KOSPI} 의 평균값을, $\sigma(\cdot)$ 은 표준편차를 의미한다. 따라서 $Std_S_t^{KOSPI}$ 는 평균 0과 분산 1의 분포를 갖는다. 이하에서 코스피 금융감성지수는 $Std_S_t^{KOSPI}$ 를 이른다.

4. 코스피 수익률과 여타변수

제IV절의 분석에 필요한 변수들의 구축방법을 설명하기로 한다. 일별 코스피 수익률(R_t^{KOSPI})은 한국 주식시장의 대표 지수인 코스피200지수의 종가($KOSPI200_Index_Close_t$)를 이용하여 식 (4)와 같이 로그 차분하여 구하였다.

$$R_t^{KOSPI} = \ln\left(\frac{KOSPI200_Index_Close_t}{KOSPI200_Index_Close_{t-1}}\right) \quad (4)$$

$$S_{it}^{SamsungElectron News} = \frac{(+1)*N_{it}^{pos words} + (-1)*N_{it}^{neg words}}{N_{it}^{pos words} + N_{it}^{neg words}} \quad (5)$$

($i = 1 \dots L_t$; L_t = the number of news articles at time t ; $t = 2021.01.05. \sim 2022.06.30.$)

$$S_t^{SamsungElectron} = \frac{1}{L_t} \sum_{i=1}^{L_t} S_{it}^{SamsungElectron News} \quad (6)$$

$$Std_S_t^{SamsungElectron} = \frac{S_t^{SamsungElectron} - \overline{S_t^{SamsungElectron}}}{\sigma(S_t^{SamsungElectron})} \quad (7)$$

삼성전자 금융감성지수의 구축방법은 코스피 금융감성지수의 구축방법과 동일하

며, 6개 신문기사 사이트(동아일보, 한겨레, KBS, SBS, 매일경제, 서울경제)에서 ‘삼성전자’ 관련 12,952개의 기사를 수집한 후, 앞서 구축한 금융감성사전을 이용하여 기사를 지수화하였다. 즉 ‘코스피’라는 주식 관련 포괄적인 검색어로 수집한 금융 기사를 토대로 만든 금융감성사전을 범용 금융감성사전으로 사용하며, 삼성전자 금융감성지수를 구축하는 데도 이용하였다. 그 이유는 금융감성사전의 주된 의의가 금융시장을 대변하는 단어들을 추출하여 이를 긍정어와 부정어로 구분하는데 있을 뿐 아니라, ‘코스피’라는 검색어로 수집된 금융기사가 대부분의 주요 주식시장 관련 긍정어와 부정어를 포함할 것이라는 가정에 기반한 것이다. 식 (5)는 삼성전자 관련 신문기사의 감성지수($S_{it}^{Samsung\ Electron\ News}$)를 구축하는 방법이다. 구체적으로 t 시점에 게시된 L_t 개의 신문기사 중 i 번째 기사의 감성을 지수화하는 식으로서, 그 기사에 포함된 긍정어의 개수($N_{it}^{pos\ words}$)에 1의 값을 곱한 것과 부정어의 개수($N_{it}^{neg\ words}$)에 (-1)의 값을 곱한 것의 합을 총 긍정어와 부정어 개수의 합으로 나눈 것이다. 식 (6)은 $S_{it}^{Samsung\ Electron\ News}$ 을 이용하여 t 시점의 일별 삼성전자 금융감성지수($S_t^{Samsung\ Electron}$)를 만드는 방법을 나타낸다. 식 (7)의 $Std_S_t^{Samsung\ Electron}$ 은 식 (3)과 마찬가지로 표준화된 삼성전자 금융감성지수를 이르면, 우변의 $\overline{S_t^{Samsung\ Electron}}$ 는 $S_t^{Samsung\ Electron}$ 의 평균값을, $\sigma(\cdot)$ 는 $S_t^{Samsung\ Electron}$ 의 표준편차를 의미한다. 이하에서 삼성전자 금융감성지수는 $Std_S_t^{Samsung\ Electron}$ 을 지칭한다.

IV. 분 석

1. 코스피 수익률에 대한 코스피 금융감성지수의 회귀분석

$$R_t^{KOSPI} = \alpha_t + \beta_1 Std_S_t^{KOSPI} + \beta_2 R_{t-1}^{KOSPI} + \beta_3 \ln(TA_t^{KOSPI}) + \epsilon_t \quad (8)$$

코스피 금융감성지수($Std_S_t^{KOSPI}$)가 코스피 수익률(R_t^{KOSPI})에 대해 설명력이 있는지 식 (8)의 회귀분석을 통해 보이고자 한다. 통제변수로는 코스피 수익률의 1기 시차값(R_{t-1}^{KOSPI})과 코스피 거래량의 로그값($\ln(TA_t^{KOSPI})$)을 사용한다. 본 연구에서 쓰인 변수들의 기초 통계량은 <Table 1>에서 보이고 있다. 코스피 수익률

(R_t^{KOSPI}) 과 삼성전자 주가수익률($R_t^{Samsung\ Electron}$) 은 분석기간동안 평균적으로 음(-)의 값을 보인다. 본 연구의 모든 회귀분석에 사용되는 변수들은 단위근 검정을 통해 그 안정성이 확인되었다.¹⁵⁾

회귀분석에 앞서 <Figure 2>는 코스피 수익률과 코스피 금융감성지수의 동행 관계를 보여주며 이러한 관계는 코스피 금융감성지수와 코스피 수익률 간의 상관관계가 0.43이라는 사실을 통해서도 확인할 수 있다(<Table 2>). <Table 3>는 회귀분석 결과인데, 일별 코스피 금융감성지수가 1-표준편차 증가할 때 일별 코스피 수익률은 약 0.60% 증가함을 보여 코스피 금융감성지수가 코스피 수익률에 유의미한 영향력이 있음을 알 수 있다.

이러한 코스피 금융감성지수의 설명력은 코스피 금융감성지수가 포함된 식(식 (8))과 포함되지 않은 식(아래 식 (8)')을 대상으로 Mean Squared Error(MSE)¹⁶⁾를 통해 예측력을 비교함으로써도 보일 수 있다. 구체적으로 train set(2021.01.05. - 2021.12.30)에 대해 식 (8)의 회귀분석을 통해 beta 추정치를 구하고, 이를 이용하여 test set(2022.01.03-2022.06.30)에 대해 코스피 수익률 예측치와 실제치를 비교한 MSE를 구한다. 동일한 방법을 식 (8)'에 대해서도 적용하여 MSE를 구한 후 이 둘을 비교해 본 결과, 식 (8)의 MSE(0.139e-03)가 식 (8)'의 MSE(0.155e-03)보다 작은 것으로 나타나서 코스피 금융감성지수가 추가적인 설명력이 있음을 다시 한번 확인할 수 있었다.

$$R_t^{KOSPI} = \alpha_t + \beta_1 R_{t-1}^{KOSPI} + \beta_2 \ln(TA_t^{KOSPI}) + \epsilon_t \quad (8)'$$

15) 변수들의 단위근 검정의 결과는 다음과 같다.

<Augmented Dickey-Fuller Unit Root Test>

Variables	Test Statistics
R_t^{KOSPI}	-19.186***
$R_t^{Samsung\ Electron}$	-18.317***
$Std_S_t^{KOSPI}$	-10.544***
$Std_S_t^{Samsung\ Electron}$	-14.236***

Note: *** indicates 1% significance.

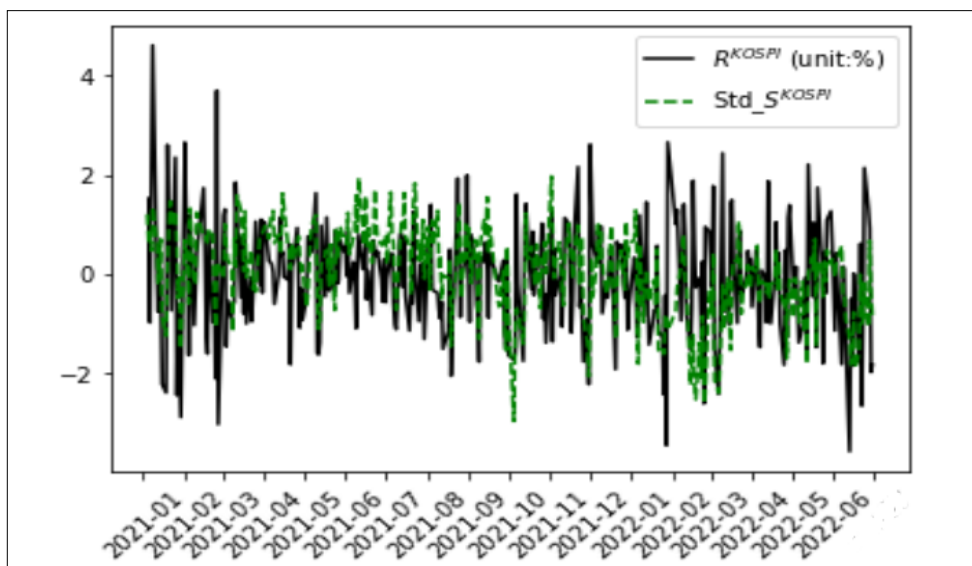
16) 평균제곱오차를 뜻하며 값이 작을수록 예측값의 정확도가 높음을 의미한다.

〈Table 1〉 Descriptive Statistics of Variables and Data Sources

Variables	Data Period	Source	Average	Standard Deviation
R_t^{KOSPI}	2021. 01. 05. -2022. 06. 30.	Korea Exchange	-0.0007	0.0113
$R_t^{Samsung\ Electron}$		Korea Exchange	-0.0100	0.0132
$\ln(TA_t^{KOSPI})$		Korea Exchange	12.0929	0.3307
$\ln(TA_t^{Samsung\ Electron})$		Korea Exchange	16.6078	0.3783
$Std_S_t^{KOSPI}$		Made by Authors	0.0674	0.8908
$Std_S_t^{Samsung\ Electron}$		Made by Authors	-0.0049	0.9906
R_t^f		Bank of Korea	0.652e-04	0.146e-04

Note: R_t^f is converted from annual to daily data to match a daily return of the stock prices.

〈Figure 2〉 R_t^{KOSPI} and $Std_S_t^{KOSPI}$



〈Table 2〉 Correlation between R_t^{KOSPI} and $Std_S_t^{KOSPI}$

	R_t^{KOSPI}	$Std_S_t^{KOSPI}$
R_t^{KOSPI}	1.0000	
$Std_S_t^{KOSPI}$	0.4356	1.0000

〈Table 3〉 Regression of R_t^{KOSPI} on $Std_S_t^{KOSPI}$
(control variables : R_{t-1}^{KOSPI} , $\ln(TA_t^{KOSPI})$)

	coefficient	adjusted R^2
$Std_S_t^{KOSPI}$	0.0060*** (0.000)	0.2022
R_{t-1}^{KOSPI}	-0.1287*** (0.009)	
$\ln(TA_t^{KOSPI})$	-0.0020 (0.210)	
constant	0.0236 (0.234)	

Note: *** indicates 1% significance. P-values are reported in parentheses.

2. 삼성전자 주식을 이용한 CAPM 응용

본 항에서는 Capital Asset Pricing Model(CAPM)을 이용하여 삼성전자 금융감성지수($Std_S_t^{Samsung\ Electron}$)가 개별 주식을 대표하는 삼성전자 주식 가격의 움직임을 설명할 수 있는지 분석한다. CAPM은 다양한 연구에서 활용되었으나 행동경제학적 관점에서 CAPM을 이용하여 추가 변화에 대한 투자자 심리 영향을 포착하려는 시도가 다수 있었다. 대표적으로 Verma and Soydemir(2006)와 Apergis and Rehman(2018) 등은 CAPM 분석을 통해 잔차를 추정하였는데, 잔차가 투자자 심리와 연관되어 있음을 보였다. 본 연구에서는 CAPM 분석을 통해 잔차를 따로 추정하기보다는 CAPM의 설명변수에 금융감성지수를 추가하는 방식으로 삼성전자 주가의 초과수익률을 설명하기로 한다.

$$R_t^{Samsung\ Electron} - R_t^f = \alpha_t + \beta_1 [R_t^{KOSPI} - R_t^f] + \beta_2 Std_S_t^{KOSPI} + \beta_3 \ln(TA_t^{Samsung\ Electron}) + \epsilon_t \quad (9)$$

$$R_t^{Samsung\ Electron} - R_t^f = \alpha_t + \beta_1 [R_t^{KOSPI} - R_t^f] + \beta_2 Std_S_t^{Samsung\ Electron} + \beta_3 \ln(TA_t^{Samsung\ Electron}) + \epsilon_t \quad (10)$$

$$R_t^{KOSPI} = \alpha_t + \beta_1 Std_S_t^{Samsung\ Electron} + \beta_2 R_{t-1}^{KOSPI}$$

$$+ \beta_3 \ln(TA_t^{KOSPI}) + \epsilon_t \quad (11)$$

$R_t^{SamsungElectron}$ 은 삼성전자의 주가수익률을 의미하고 R_t^{KOSPI} 는 코스피 수익률을 의미하며, R_t^f 는 무위험 이자율로서 원래 연율화된 데이터로 보고되므로 다른 변수들과의 빈도를 일치시키기 위하여 365로 나눠서 일별 데이터로 변환하였다. 식 (9)과 식 (10)의 회귀분석을 통해 삼성전자 주가의 초과수익률($R_t^{SamsungElectron} - R_t^f$)에 대해서 공통 변수인 시장 프리미엄($R_t^{KOSPI} - R_t^f$)의 영향력뿐만 아니라, 주식시장 전반의 투자자 심리를 대표하는 코스피 금융감성지수($Std_S_t^{KOSPI}$)와 삼성전자 주식에 한정된 삼성전자 금융감성지수($Std_S_t^{SamsungElectron}$)의 영향력을 비교할 수 있다. 이때 삼성전자 주식 거래량 로그값($\ln(TA_t^{SamsungElectron})$)을 통제변수로 이용한다. 식 (11)은 식 (8)에 대응하여 삼성전자 금융감성지수가 코스피 수익률에 대한 설명력이 있는지를 확인하는 식이다.

〈Table 4〉의 회귀분석 결과는 매우 흥미로운 결과를 보여준다. 우선 시장 프리미엄은 식 (9)과 식 (10)에서 모두 삼성전자 주가의 초과수익률에 대해서 유의하고 강한 양(+)의 설명력을 가진다. 나아가 중요한 것은 시장 전반의 심리를 대표하는 코스피 금융감성지수는 삼성전자 주가의 초과수익률에 유의한 영향을 미치지 못하였으나(식 (9)), 기업 고유의 심리지수인 삼성전자 금융감성지수는 삼성전자 주가의 초과수익률에 유의한 영향력을 보인다는 것이다(식 (10)). 구체적으로 일별 삼성전자 금융감성지수가 1-표준편차 증가할 때 삼성전자 주가의 초과수익률은 0.75% 증가함을 보인다. 〈Table 5〉에서 보인 식 (11)의 회귀분석 결과는 〈Table 3〉에서 보인 식 (8)의 결과와 달리 삼성전자 금융감성지수가 주식시장 전반의 코스피 수익률에 대한 유의한 설명력이 없음을 보여준다. 이는 각 주가에 대해 그 고유의 금융감성지수만이 설명력이 있다는 것을 보여주며, 본 연구의 금융감성지수가 정교하게 구축되었음을 시사한다.

〈Table 4〉 The Comparison of the impacts on $[R_t^{SamsungEledron} - R_t^f]$
of $Std_S_t^{KOSPI}$ and $Std_S_t^{SamsungEledron}$
(control variable : $\ln(TA_t^{SamsungEledron})$)

		coefficient	adjusted R^2
equation (9)	$R_t^{KOSPI} - R_t^f$	1.0082*** (0.000)	0.7377
	$Std_S_t^{KOSPI}$	-0.0025 (0.566)	
	$\ln(TA_t^{SamsungEledron})$	0.0002 (0.789)	
	Constant	-0.0044 (0.775)	
equation (10)	$R_t^{KOSPI} - R_t^f$	0.9955*** (0.000)	0.7405
	$Std_S_t^{SamsungEledron}$	0.0075** (0.038)	
	$\ln(TA_t^{SamsungEledron})$	0.0006 (0.487)	
	Constant	-0.0113 (0.475)	

Note: ***, ** indicate 1%, 5%, significance respectively. P-values are reported in parentheses.

〈Table 5〉 Regression of R_t^{KOSPI} on $Std_S_t^{SamsungEledron}$
(control variables : R_{t-1}^{KOSPI} , $\ln(TA_t^{KOSPI})$)

	coefficient	adjusted R^2
$Std_S_t^{SamsungEledron}$	0.0008 (0.179)	-0.0026
R_{t-1}^{KOSPI}	-0.0063 (0.904)	
$\ln(TA_t^{KOSPI})$	-0.0005 (0.782)	
constant	-0.0061 (0.782)	

Note: P-values are reported in parentheses.

V. 결 론

본 연구는 행동경제학적 시각에서 투자자 심리를 대변하는 금융감성지수를 구축하

여 이를 통해 주가의 움직임을 설명하고자 하였다. 특히 기존 선행연구에서 주로 이용된 L&M 사전을 단순 번역하여 쓰기보다는 L&M 사전의 긍정어 및 부정어와 유사한 국내 금융기사 단어들을 방대하게 추출하여 이들의 감성을 새롭게 분류하였다. 또한 각 연구자들의 재량적 판단으로 단어들의 감성을 분류한 선행연구들과 달리, 본 연구는 L&M 사전과 Word2Vec의 기법을 통해 객관성을 확보하면서도 국내 금융 텍스트 분석에 적절한 금융감성사전을 국내 연구 최초로 구축하였다. 뿐만 아니라 일별 금융감성지수를 이용하여 고빈도 금융감성지수가 주가수익률의 단기 변화를 설명할 수 있는지 실험적으로 살펴보았다.

분석결과, 주식시장을 대표하는 코스피200지수의 수익률에 대해 코스피 금융감성지수는 강한 설명력을 보였다. 또한 전통적 CAPM에 금융감성지수를 추가하여, 삼성전자 주가의 초과수익률이 시장 프리미엄뿐 아니라 금융감성지수에 의해서도 영향을 받는지 살펴보았다. 흥미롭게도 코스피 금융감성지수는 삼성전자 주가의 초과수익률에 유의한 영향을 주지 못하였으나, 삼성전자 금융감성지수는 삼성전자 주가의 초과수익률에 유의하게 양(+)의 영향을 미쳤다. 뿐만 아니라 삼성전자 금융감성지수는 코스피 수익률에는 영향을 미치지 못함을 발견하였다. 이러한 분석결과를 요약하면, 주식시장 전체의 가격 변화에 대해서는 시장 금융감성지수가 설명력을 가지는 반면, 개별 주식의 가격 변화에 대해서는 개별 기업 고유의 금융감성지수가 설명력을 가진다는 것으로서, 이는 본 연구의 금융감성지수가 정교하게 구축되었음을 시사한다.

본 연구는 국내 최초로 한국어 금융기사 분석에 적합한 금융감성사전과 고빈도 데이터인 일별 금융감성지수를 구축하여 이를 통해 단기 주가의 움직임을 설명하였다는 점에서 분명 관련 연구에 기여하는 바가 있다고 사료된다. 하지만 방대한 데이터수집이라는 현실적 제약으로 분석 기간이 1년 반에 그쳐, 선행연구보다 상대적으로 분석 기간이 짧다는 한계점도 존재한다. 향후 연구에서는 분석 기간을 늘려서 중장기적인 금융감성지수를 구축하여 금융감성지수의 주식시장에 대한 단기적 영향뿐만 아니라 중·장기적 영향도 포착해 볼 것이다. 또한 본 연구에서는 ‘코스피’로 검색하여 수집된 금융기사가 대부분의 주요 주식시장 관련 긍정어와 부정어를 포함할 것이라는 가정 하에 이를 토대로 만든 금융감성사전을 범용 금융감성사전으로 사용하였다. 비록 실증결과를 통해 본고의 금융감성사전이 범용 금융감성사전으로서의 가능성을 보였지만, 이에 대해 좀 더 엄밀한 검증이 필요한 것이 사실이다. 추후 연구에서는 다양한 검색어를 이용하여 본고의 금융감성사전의 한계를 보완할 것이다. 또한 국내 선행

연구에서는 한국어로 된 금융 텍스트를 영어로 번역한 뒤 이에 대해 영어 금융감성사전을 이용하여 금융감성지수를 구축하는 것이 일반적인데, 이렇게 만든 금융감성지수와 본 연구에서 구축된 금융감성사전을 이용하여 만든 금융감성지수의 차이를 분석하여 그 유효성을 비교하는 것도 의미가 있을 것이다. 마지막으로 본 연구의 금융감성사전을 이용하여 일반 신문기사가 아닌 전문가 집단의 애널리스트 보고서를 분석하여 새로운 금융감성지수를 구축하는 것도 흥미로운 연구가 될 것이다.

■ 참 고 문 헌

1. 객진아 · 이원재, “온라인 커뮤니티에서 나타난 건강정보 전파 분석에 관한 연구: 공적/사적 의료 지식에 미치는 커뮤니티 내 연결망 효과 중심으로,” 『통계연구』, 특별호, 2016, pp. 112-128
2. 김가람 · 양희진 · 류두진, “애널리스트 투자 의견 변경이 금융시장과 투자자 심리에 미치는 영향,” 『경제학연구』, 제68집 제3호, 2020, pp. 45-81.
3. 김정미 · 이주홍, “Word2Vec을 활용한 RNN기반의 문서 분류에 관한 연구,” 『한국지능시스템학회 논문지』, 제27집 제6호, 2017, pp. 560-565.
4. 박은정 · 조성준, “KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지,” 한국정보과학회언어공학연구회 2014년도 제26회 한글 및 한국어 정보처리 학술대회, 2014, pp. 133-136.
5. 양아정 · 유세경, “행동경제학 관점에서 본 온라인 동영상 서비스(OTT)의 지속적 이용에 영향을 미치는 요인에 관한 연구,” 『한국콘텐츠학회』, 제22권 제8호, 2022, pp. 159-169.
6. 조수지 · 김홍규 · 양철원, “기업 재무분석을 위한 한국어 감성사전 구축,” 『한국증권학회지』, 제50권 제2호, 2021, pp. 131-170.
7. 황재홍, “불확실성에서의 의사결정과 확률: 케인즈와 행동경제학,” 『경제학연구』, 제67집 제3호, 2019, pp. 165-187.
8. Aissia, D. B. “Home and Foreign Investor Sentiment and the Stock Returns,” *Quarterly Review of Economics and Finance*, Vol. 59, 2016, pp. 71-77.
9. Apergis, N. and M. U. Rehman, “Is CAPM a Behavioral Model? Estimating Sentiments from Rationalism,” *Journal of Behavioral Finance*, Vol. 19, No. 4, 2018, pp. 442-449.
10. Baker, M. and J. Wurgler, “Investor Sentiment and the Cross-Section of Stock Returns,” *Journal of Finance*, Vol. 61, No. 4, 2006, pp. 1645-1680.
11. Bian, S., D. Jia, F. Li, and Z. Yan, “A New Chinese Financial Sentiment Dictionary for Textual Analysis in Accounting and Finance,” *SSRN Papers*, 2021.
12. Chan, F., R. B. Durand, J. Khuu and L. A. Smales, “The Validity of Investor Sentiment Proxies,” *International Review of Finance*, Vol. 17, No. 3, 2017, pp. 473-477.
13. Chen, C. C., H. H. Huang, and H. H. Chen, “NTUSD-Fin: A Market Sentiment Dictionary for Financial Social Media Data Applications,” *Proceedings of the First Financial Narrative Processing Workshop*, 2018.
14. Chen, R., J. Yu, C. Jin and W. Bao, “Internet Finance Investor Sentiment and Return Comovement,” *Pacific-Basin Finance Journal*, Vol. 56, 2019, pp. 151-161.

15. Da, Z., J. Engelberg and P. Gao, "The Sum of All FEARS Investor Sentiment and Asset Prices," *Review of Financial Studies*, Vol. 28, No. 1, 2015, pp.1-32.
16. Du, Z., A. G. Huang, R. Wermers and W. Wu, "Language and Domain Specificity: A Chinese Financial Sentiment Dictionary," *Review of Finance*, Vol. 26, No. 3, 2022, pp. 673-719.
17. Farina, V., A. Parisi and U. Pomante, "Economic Blogs Sentiment and Asset Prices," *International of Finance and Economics*, Vol. 22, No. 4, 2017, pp.341-351.
18. Frankel, R., J. Jennings, and J. Lee, "Disclosure Sentiment: Machine Learning VS. Dictionary Methods," *Management Science*, Vol. 68, No. 7, 2022, pp.5514-5532.
19. Heston, S. L. and N. R. Sinha, "News Versus Sentiment: Comparing Textual Processing Approaches for predicting Stock Returns," Robert H. Smith School Research Paper, 2014.
20. Jiang, F., J. Lee, X. Martin and G. Zhou, "Manager Sentiment and Stock Returns," *Journal of Financial Economics*, Vol. 132, 2019, pp.126-149.
21. Kahneman, D., "Prospect Theory: An Analysis of Decisions under Risk," *Econometrica*, Vol. 47, No. 2, 1979, pp.263-292.
22. Katayama, D., Y. Kino and K. Tsuda, "A Method of Sentiment Polarity Identification in Financial News Using Deep Learning," *Procedia Computer Science*, Vol. 159, 2019, pp.1287-1294.
23. Loughran, T. and B. McDonald, "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *Journal of Finance*, Vol. 66, No. 1, 2011, pp.35-65.
24. Li, X., H. Xie, L. Chen, J. Wang and X. Deng, "News Impact on Stock Price Return via Sentiment Analysis," *Knowledge-Based Systems*, Vol. 69, 2014, pp.14-23.
25. Lintner, J., "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, Vol. 20, No. 4, 1965, pp.587-615.
26. Nguyen, B. H. and V. N. Huynh, "Textual Analysis and Corporate Bankruptcy: A Financial Dictionary-Based Sentiment Approach," *Journal of the Operational Research Society*, Vol. 73, No. 1, 2022, pp.102-121.
27. Palmer, M., J. Roeder and J. Muntermann, "Induction of a Sentiment Dictionary for Financial Analyst Communication: A Data-Driven Approach Balancing Machine Learning and Human Intuition," *Journal of Business Analytics*, Vol. 5, No. 1, 2021, pp.1-19.
28. Piccoli, P. and M. Chaudhury, "Overreaction to Extreme Market Events and Investor Sentiment," *Applied Economics Letters*, Vol. 25, No. 2, 2018, pp.2767-2789.
29. Sharpe, W. F. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, Vol. 19, 1964, pp.425-442.
30. Simon, H. A., "A behavioral Model of Rational Choice," *Quarterly Journal of Economics*, Vol. 69, No. 1, 1955, pp.99-118.
31. Verma, R. and G. Soydemir, "The Impact of U.S. Individual and Institutional Investor Sentiment on Foreign Stock Markets," *Journal of Behavioral Finance*, Vol. 7, No. 3, 2006, pp.128-144.
32. Zhou, G. "Measuring Investor Sentiment," *Annual Review of Financial Economics*, Vol. 10, 2018, pp.239-259.

Construction of Financial Sentiment Dictionary and Effectiveness of Financial Sentiment Index

So-yeon Lim* · In-bae Kim**

Abstract

This paper investigates whether the investors' psychological factor has a predictive power for the stock market movement from the perspective of the behavioral economics. We made several contributions in this literature. First, we constructed the financial sentiment dictionary for the Korean text analysis to its name for the first time. It is based on the most popular financial sentiment dictionary of Loughran and McDonald (2011), but much expanded by Word2Vec model. Second, different from the related previous studies that mostly used the monthly data for financial sentiment index (FSI), we built its daily data to explain the high frequency stock return movement. Lastly, we built two kinds of FSI, that is, 'KOSPI FSI' for the overall Korean stock market and 'Samsung Electronics FSI' for an individual stock price, and explore their predictive powers. From this exercise we found the following results. (i) When the standardized daily KOSPI FSI increases by one standard deviation, the daily KOSPI return rises by 0.60%. (ii) When the standardized daily Samsung Electronics FSI increases by one standard deviation, the daily excess return of Samsung Electronics rises by 0.75%. (iii) However, the KOSPI FSI can not explain the excess return of Samsung Electronics. Nor can the Samsung Electronics FSI the KOSPI returns. Such results enforce the reliability and validity of the FSIs constructed in this paper.

Key Words: behavioral economics, financial sentiment dictionary, financial sentiment index

JEL Classification: G1

Received: Dec. 19, 2022. Revised: March 17, 2023. Accepted: April 24, 2023.

* First Author, Ph.D Candidate, Economics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea, e-mail: limsoyun2000@naver.com

** Corresponding Author, Professor, Economics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea, Phone: +82-2-3277-4064, e-mail: ibkim@ewha.ac.kr