



Information Systems Review

제21권 제1호

ISSN : 1229-5078(Print) 2713-8143(Online)

소셜 감성과 암호화폐 가격 간의 관계 분석 : 빅데이터를 활용한 계량경제적 분석

유상이, 현지연, 이상용

To cite this article : 유상이, 현지연, 이상용 (2019) 소셜 감성과 암호화폐 가격 간의 관계 분석 : 빅데이터를 활용한 계량경제적 분석 , Information Systems Review, 21:1, 91-111

① earticle에서 제공하는 모든 저작물의 저작권은 원저작자에게 있으며, 학술교육원은 각 저작물의 내용을 보증하거나 책임을 지지 않습니다.

② earticle에서 제공하는 콘텐츠를 무단 복제, 전송, 배포, 기타 저작권법에 위반되는 방법으로 이용할 경우, 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

www.earticle.net

소셜 감성과 암호화폐 가격 간의 관계 분석: 빅데이터를 활용한 계량경제적 분석

An Analysis of Relationship between Social Sentiments and Cryptocurrency Price: An Econometric Analysis with Big Data

유 상 이 (Sangyi Ryu)

한양대학교 경영대학 석사

현 지 연 (Jiyeon Hyun)

한양대학교 경영대학 석사

이 상 용 (Sang-Yong Tom Lee)

한양대학교 경영대학 교수, 교신저자

요 약

2017년 말, 전 세계적으로 비트코인을 필두로 암호화폐에 대한 투자 열풍이 시작되었으며, 특히 한국은 그 중심에 서 있는 상황이었다. 한국의 투자자들이 그간 수익성이 있는 투자 기회를 찾기가 어려웠던 만큼 새로운 투자처에 투자심리가 몰린 것으로 보인다. 하지만 암호화폐에 대한 이러한 한국의 열기는 자산의 본질적인 가치에 기초한 투자가 아니라 단기적 차익 실현 기대 및 사회적 분위기에 따른 것이기 때문에 심리적 현상에 좌우되는 바가 크다고 할 수 있다. 따라서 본 연구에서는 이를 살펴보기 위해 트위터와 비트코인을 대표로 선정하여 사람들의 소셜 감성이 암호화폐에 미치는 영향을 분석해보고자 하였다. 데이터는 2017년 11월 1일부터 2018년 4월 30일까지 총 181일간 트위터상에 노출된 비트코인 관련 게시물과 빗썸/업비트의 비트코인 가격을 대상으로 수집하였다. 수집된 트위터 데이터는 감성 분석을 통해 긍정어 및 긍·부정어로 정제해주었고, 정제된 긍정어, 긍정어, 부정어는 비트코인 가격에 어떤 영향을 미치는지 확인하기 위해 회귀분석 모형에 투입하였다. 회귀분석을 통해 관계를 살펴본 후에는 Granger Causality test를 통해 인과관계의 존재 여부를 확인하였다. 그 결과, 긍정어는 비트코인 가격과 정의 관계로 나타났고, 부정어는 부의 관계로 나타났다. 또한 소셜감성과 비트코인 가격간에는 양방향의 인과관계가 있음을 확인하였다. 즉, 비트코인 가격 변동이 소셜감성에 영향을 미치기도 하지만, 동시에 소셜감성의 변화도 암호화폐 투자자들의 행동에 영향을 미칠 수 있음을 확인할 수 있었다.

키워드 : 비트코인, 암호화폐, 소셜 감성, 감성 사전, 감성 분석, 오피니언 마이닝, Granger 인과관계

† 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017S1A3A2066740).

I. 서 론

2017년 말, 전 세계적으로 비트코인(Bitcoin)을 필두로 암호화폐(cryptocurrency)에 대한 투자 열풍이 시작되었다. 특히 한국에서는 남녀노소를 가리지 않고 많은 사람들이 투자에 나섰고, 대다수의 암호화폐 가격이 급격히 상승하였다. 각종 미디어와 인터넷 커뮤니티, 소셜 미디어(social media) 상에서도 연일 암호화폐 이슈와 성공기들이 쏟아져 나왔고, 이에 따라 더 많은 투자자들이 유입되었다. 이러한 한국의 암호화폐 열풍을 대변하듯 각종 신조어들도 등장했다. 그 중에서 국내 암호화폐 가격이 해외 암호화폐 가격보다 높게 형성되어 있음을 의미하는 ‘김치프리미엄’이라는 신조어는 한국의 암호화폐 투자 열기가 얼마나 뜨거운지 짐작할 수 있게 한다. 암호화폐 시장은 주식시장과 마찬가지로 수요와 공급에 따라 가격이 결정되는데, 각 거래소마다 가격이 독립적이므로 한국에서의 높은 가격은 그만큼 수요가 많다는 것을 의미하기 때문이다.

통계분석업체인 닐슨코리아클릭에 따르면 암호화폐 열풍이 정점에 달한 2018년 1월의 암호화폐 서비스 총 이용자 수는 약 509만 명에 달했으며, 일 거래액은 10조 원까지 치솟았다. 이는 월 방문자가 100만 명이 채 안되던 2017년 초 대비 수 배 증가한 수치이며, 같은 기간 증권 이용자 수(776만 명)의 약 65.6% 수준이다. 규모면에서 보더라도 미국, 일본 다음으로 큰 규모이며, GDP에 대비해 봤을 때는 전 세계에서 가장 거대한 규모의 시장이라 할 수 있다. 심지어 암호화폐 거래소 순위 제공 업체인 코인힐스를 보면, 한국의 신생 거래소 ‘업비트(Upbit)’가 오픈 3개월 만에 전체 거래액 기준 세계 1위를 차지하였으며, ‘빗썸(Bithumb)’은 그 뒤를 이어 3위에 오르기도 했다.

한편, 이러한 투자 열기에 따라 암호화폐를 이용한 다양한 사업 모델도 등장하였다. 하지만 원래의 목적과는 달리 범죄 수단으로 악용되는 경우가 종종 발생하고 있어(이희중, 조재영, 2018) 이

로 인한 피해가 속출할 것으로 우려된다. 특히 각종 미디어와 인터넷, 소셜 미디어 상에서 연일 암호화폐 이슈와 성공기가 쏟아져 나오면서 이를 본 사람들이 아무런 배경지식없이 투자에 뛰어들고 있어 더 큰 피해가 확산되고 있다. 다단계 폰지사기 및 개인정보 해킹 등은 물론 심지어 신종 환치기 수법 수단으로 동원되기도 했다. 여기에는 학생, 직장인은 물론 노년층까지 가세해 피해를 보고 있는데, 이러한 분위기에 휩쓸린 암호화폐 투자자들의 행동은 소셜 감성의 영향력을 통하여 분석 및 해석이 가능할 것으로 보인다.

다시 말하여, 암호화폐에 대한 한국의 열기는 자산의 본질적인 가치(intrinsic value)에 기초한 투자가 아니라 단기적 차익 실현 기대 및 사회적 분위기에 따른 것이기 때문에 심리적 현상에 좌우되는 바가 크다. 최근 들어 사람들은 이러한 사회적인 이슈 및 분위기를 인터넷과 소셜 미디어로 가장 신속하게 접하므로, 결과적으로 소셜 미디어와 같은 도구를 통하여 형성되는 소셜 감성(social sentiment)은 암호화폐에 대한 투자 열기와 상호작용한다고 볼 수 있다. 소셜 감성이 암호화폐 투자에 영향을 미친다면 이는 사람들의 투자가 사회적 분위기 및 다른 사람들의 감정에 영향을 받는다는 것을 의미하기 때문에 본 연구에서는 소셜 감성과 암호화폐 간의 관계를 살펴보고자 하는 것이다.

경영정보학계에서도 2010년대 이후 정보시스템과 공학적인 학문들을 융합한 기술적인 연구 동향이 진행되고 있다(안정국 등, 2016). 특히 많은 연구자들이 인터넷 및 소셜 미디어의 영향으로 인해 소셜 미디어를 사회에 적용하는 것에 관심을 기울이기 시작했다(Kim, 2015). 그중에서도 소셜 감성을 통해 자산시장의 변화를 예측하고자 하는 연구가 활발히 진행되어 왔으며, 많은 연구자들이 소셜 감성과 주가 사이에 높은 상관관계가 있음을 밝혀냈다(Bollen *et al.*, 2011; Gilbert *et al.*, 2010; Oh and Sheng, 2011; Rao and Srivastava, 2012; Zhang *et al.*, 2009). 이러한 대부분의 연구에

서는 소셜 미디어 중에서도 트위터의 데이터를 주로 활용하고 있는데, 트위터는 전 세계적으로 가장 많이 사용되는 소셜 미디어 중 하나이며 수백만 명의 사람들이 주로 다양한 주제에 대한 의견을 제시하고 공유하는데 사용하기 때문이다. 트위터에서는 특정 주제나 이슈에 대한 대량의 심리적인 데이터를 비교적 쉽게 구할 수 있으며, 이 데이터에 대한 분석을 수행하여 가치 있는 인사이트를 얻을 수 있다(Chheda *et al.*, 2018). 뿐만 아니라 트위터는 투자 의사 결정에 대한 아이디어와 생각을 공유하는 수단으로도 급속히 성장하였다(Matta *et al.*, 2015). 따라서 소셜 감성이 암호화폐 가격에 미치는 영향을 분석하는 것은 최근의 이러한 연구와 흐름을 같이 하는 것이라 볼 수 있다. 이에 본 연구에서는 소셜 감성과 암호화폐 가격간의 관계를 살펴보기 위해 대표로 트위터와 비트코인을 선정하여 실증분석을 수행하였다.

II. 연구배경

2.1 암호화폐

암호화폐란 P2P(Peer to Peer) 네트워크 기반의 전자 금융 거래 시스템으로서, 디지털 화폐 또는 가상화폐라 부르기도 한다. 2009년 사토시 나카모토(Satoshi Nakamoto)에 의해 개발된 비트코인을 필두로 현재까지 라이트코인(Litecoin), 리플(Ripple), 이더리움(Ethereum) 등 약 이천여 개 이상의 암호화폐가 시중에 거래되고 있다.

사회과학 분야에서의 암호화폐에 대한 초기 분석은 주로 비트코인과 관련하여 이루어졌다. Hencic and Gourieroux(2014)는 비트코인과 USD 통화 환율 사이에 버블이 있는지 감지하기 위해 Non-Causal Autoregressive model을 적용하였으며, Sapuric and Kokkinaki(2014)의 연구에서는 6개 주요 통화에 대한 비트코인의 환율 변동성을 측정했다. Chu *et al.*(2015)은 USD 통화 대비 비트코인 환율의 log-returns에 대한 통계적 분석을 제공하였다.

최근에는 Bianchetti *et al.*(2018)이 암호화폐가 주로 투자 심리에 의해 주도되는 가격으로 거래되기 때문에 버블과 불안정성의 잠재적 원천이 된다고 주장했다. 그리고 이를 확인하기 위해 가장 유명한 암호화폐인 비트코인과 이더리움을 선정하여 Johansen, Ledoit and Sornette(JLS)가 작성한 LPPL(Log Periodic Power Law) 모델과 Phillips, Shi 및 Yu(PSY)가 개발한 통계 모델을 결합하여 사용하였다. 비트코인은 LPPL과 PSY 모델 모두에서 강한 버블 신호가 나타났는데, PSY 모델을 보면 그 버블이 2017년 5월에 시작되어 2018년 1월까지 지속되고 있음을 알 수 있다. LPPL 모델은 2017년 12월 중순과 2018년 1월 중순에 나타나는 버블 신호로 그 이후의 폭락을 예측하기도 했다. 이더리움도 LPPL과 PSY 모델 모두에서 버블 신호를 보여주었다. 특히 PSY 모델에서는 두 번의 뚜렷한 버블 기간이 나타났다. 하나는 2017년 3월부터 9월까지, 다른 하나는 2017년 12월부터 2018년 1월까지 지속되었다. 반면 LPPL 모델은 2017년 6월에 강한 버블 신호를 감지했는데, 이는 2017년 6월 12일의 폭락에 해당한다. 즉 LPPL과 PSY에서 측정된 버블 신호로 암호화폐 시장에서 실제로 발생하는 폭락을 일부 예측할 수 있다는 의미이다. Catania and Grassi(2017)는 암호화폐의 비대칭적이고 비선형적인 움직임을 설명하기 위하여 long memory, 레버리지 효과 및 시간에 따라 변하는 고차원 모멘트를 통합할 수 있는 모델을 개발하였다. 그리고 비트코인, 이더리움, 리플, 라이트코인을 중심으로 연구를 수행하였다. 그 결과, 암호화폐 시장은 외환 시장과는 달리 레버리지 효과가 변동성에 상당한 기여를 한다는 것을 발견했다. 평균적으로 암호화폐 시장은 주식시장과 마찬가지로 부정적 영향을 받으면 변동성이 증가한다는 것이다. Lee *et al.*(2018)은 한국 시장에서만 암호화폐 가격이 높게 형성되는 현상을 설명하기 위해 Dynamic Conditional Correlation(DCC)과 GARCH 모델을 적용하여 한국과 미국 및 일본 암호화폐 시장의 가격 변동을 조사해 한국의

암호화폐 가격 결정은 대부분 국제적인 요인보다 국내 요인을 기반으로 한다는 것을 밝혀냈다.

암호화폐의 가격은 대부분 시장 감성에 의해 좌우되기 때문에(Cheah and Fry, 2015; Cheung *et al.*, 2015; Weber, 2014; Yelowitz and Wilson, 2015) 이를 웹(Web)이나 소셜 미디어 데이터로 검증하려는 연구들도 시도되었다. Matta *et al.*(2015)은 소셜 미디어 활동이나 웹 검색 매체에 의해 추출된 정보가 비트코인 가격 변동을 예측할 수 있는지 조사하였다. 이 작업은 트위터와 구글 트렌드 데이터를 사용해 측정되었는데, 특히 비트코인 가격과 구글 트렌드 데이터 간에 상당한 상호관계 값이 발견되었다. 나아가 Colianni *et al.*(2015)은 암호화폐와 관련된 트위터 데이터가 유용한 암호화폐 거래 전략을 개발하는데 활용될 수 있다고 입증하였다. 이를 위해 오픈소스 감성 분석 API(Application Program Interface)를 사용하여 각 트윗 내의 단어를 긍정과 부정으로 평가하여 사용했다. 그 후 SVM(Support Vector Machine), Naive Bayes, Logistic Regression과 같은 기계학습 알고리즘을 적용하여 특정 암호화폐의 가격이 미리 결정된 시간 간격 내에서 증가 또는 감소하는지 여부를 확인하였는데, 90%가 넘는 예측 정확도를 보여주었다. 근래에는 Mai *et al.*(2018)이 다양한 소셜 미디어 플랫폼(인터넷 포럼 vs 마이크로 블로깅)의 상대적인 영향과 VAR(Vector Autoregressive) 및 VECM(Vector Error Correction Model)을 사용하여 소셜 미디어와 비트코인 수익간의 예측 관계를 조사, 비트코인의 인기가 증가함에 따라 검색량이 증가하고 비트코인에 대한 소셜 미디어 활동이 증가한다는 증거를 발견하였다. 즉 더 많은 관심이 사용자의 비트코인 구매에 영향을 불어넣고 가격을 상승시킨다는 것이다.

이와 같은 암호화폐에 대한 연구들은 가격변동성을 설명하고자 하는 통계학적 분석 및 모형을 제공하고자 하는 연구들로서, 암호화폐의 가격 및 거래량 데이터만을 분석 대상으로 삼고 있다. 혹은 소셜 감성과의 관계를 통한 분석도 데이터

마이닝을 기반으로 가격 변동을 예측하는 것에서 그치고 있다. 따라서 본 연구가 추구하고자 하는 소셜 감성과 암호화폐 가격 사이의 상관관계 및 인과관계를 살펴보는 것은 이와 같은 연구보다 한 걸음 더 나아간 것이라 볼 수 있다.

2.2 소셜 감성

소셜 감성이란 웹이나 소셜 미디어상에 게시된 글에 나타난 사람들의 감성을 의미한다. 감성(sentiment)은 개인의 행동과 의사결정에 크게 영향을 줄 수 있기 때문에(Akerlof and Shiller, 2010 Risk and Loewenstein, 2008) 이를 파악하는 것은 사람들의 행동을 설명하거나 예측하는데 있어 매우 중요한 일이다. 과거에는 이러한 개인의 감성을 수집하기 위해서는 FGI(Focus Group Interview)나 설문조사 등을 실시하는 수밖에 없었다. 하지만 현재는 사용자 참여 중심의 인터넷 환경이 조성되면서 많은 사람들이 블로그(Blog)나 소셜 미디어와 같은 플랫폼 서비스를 통해 의견을 제시하거나 정보를 공유하게 되었고, 여기에는 개인의 감성도 포함되어 있으므로 감성 분석(sentiment analysis)과 같은 기법을 사용하면 소셜 감성이 개인의 행동에 어떤 영향을 미치는지 확인할 수 있다. 특히 이러한 소셜 감성은 개인이 경제주체로서 제품과 서비스를 이용할 때 많은 영향을 끼친다. 대다수의 사람들이 제품 구매나 서비스 이용을 결정하기 전에 타인의 의견을 참고하기 위해 웹이나 소셜 미디어에 다른 사람들이 공유한 내용을 확인하고 도움을 얻기 때문이다.

현재까지 시장에서 나타난 반응을 금융경제학 측면에서 살펴보면 비트코인에는 자산으로서의 성격이 존재한다는 해석이 가능한데(전주용, 여은정, 2014), 마침 소셜 감성이 경제주체의 행동에 미치는 영향에 대한 연구들은 금융시장에서 가장 많이 나타났다. Zhang *et al.*(2009)의 연구에서는 트위터 게시물을 분석하여 희망과 두려움 등의 감성을 측정된 후 이들 지수와 주식시장 지표

간의 관계를 분석하였다. 결론적으로 감성적 트윗 비율과 다우존스(Dow Jones), 나스닥(NASDAQ) 및 S&P 500 사이에는 음의 상관관계, 뉴욕증시 공포지수인 변동성지수(VIX)와는 양의 상관관계를 보인다는 것을 발견했다. Gilbert *et al.*(2010)은 웹 사이트 ‘라이브저널(LiveJournal)’의 게시물을 대상으로 불안, 걱정, 공포 등의 감성을 추출하여 S&P 500 지수의 하락을 예측하고, 몬테카를로 시뮬레이션(Monte Carlo simulation)을 통해 이를 입증하였다. Oh and Sheng(2011)도 향후 주가 동향에 대한 주식 마이크로 블로그(microblog) ‘스톡트윗(StockTwits)’의 예측력을 평가하고자 감성 분석 및 데이터마이닝 알고리즘을 기반으로 모델을 구축하였다. 먼저 약 10%의 게시물을 bullish, bearish, neutral로 분류하여 수동으로 라벨링하고 나머지 90%는 lexical score(Kim and Hovy, 2006)와 Bag of Words(Schumaker and Chen, 2009)를 사용하여 자동으로 라벨을 매겨 데이터 세트를 만들었다. 그 후 데이터 세트를 테스트하여 예측 정확도에 대한 감성 효과를 조사했는데, 그 결과 ‘스톡트윗(StockTwits)’에 표현된 투자 심리는 향후 시장 방향에 대한 강력한 예측 가치가 있는 것으로 나타났다. Bollen *et al.*(2011)은 대규모 트위터 게시물에서 파생된 소셜 감성을 통해 DJIA(Dow Jones Industrial Average)를 예측하고자 하였으며, 이를 위해 Google에서 제공하는 OpinionFinder와 GPOMS(Google-Profile of Mood States)를 사용해 두 가지 방법으로 감성 데이터를 분류하였다. OpinionFinder는 감성을 긍정(positive)과 부정(negative) 두 범주로 나누고, GPOMS는 보다 세분화시켜 Calm, Alert, Sure, Vital, Kind, Happy와 같이 여섯 가지 감성으로 데이터를 구분한다. 그 다음 Granger Causality test와 Self-Organizing Fuzzy Neural Network를 통해 분류된 데이터들이 시간 경과에 따라 DJIA의 변화를 예측할 수 있는지 확인하였는데, 여기서 그들은 GPOMS의 Clam이라는 감정이 DJIA의 변화보다 선행하는 현상을 발견하였다.

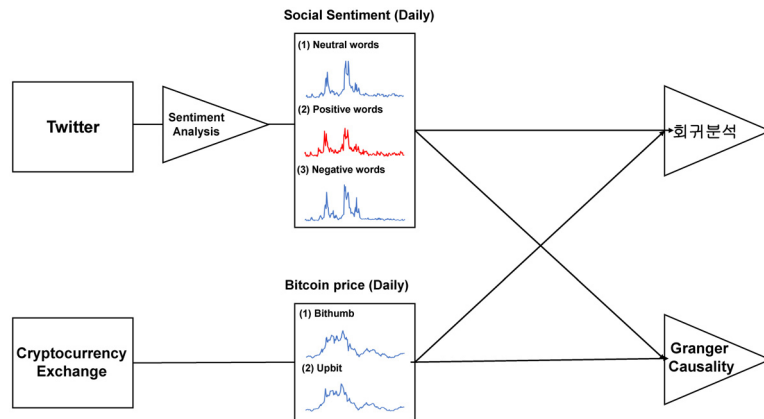
한편, 금융시장에서의 연구들 외에도 소셜 감

성이 경제주체의 행동에 영향을 줄 수 있음을 확인한 다양한 사례들이 있다. Gu *et al.*(2012)에서는 온라인 리뷰가 소매업체 판매에 중요한 영향을 미친다고 하였고, Aggarwal *et al.*(2012a, 2012b)에서도 온라인 리뷰에서의 다른 사용자의 감성이 소비자의 구매패턴에 영향을 주고 있음을 밝혔다. Luo *et al.*(2013)은 기업 지분 가치를 예측할 때 기존의 온라인 행동 메트릭(google researchs and web traffic)보다 소셜 미디어 메트릭(blog post)의 영향이 크다고 보여준 바 있다.

이와 같이 경영정보학계에서 소셜 감성이 시장 및 경제주체에 미치는 영향을 분석하는 것이 점차 확대되고 있으나, 아직까지 역사가 짧은 암호화폐는 분석 대상으로 본격화되지 못하고 있다. 혹은 소셜 감성과 암호화폐간의 관계를 단순히 데이터마이닝적 차원에서 현상을 파악하려고 하는 수준이다. 따라서 본 연구에서는 암호화폐를 소셜 감성과의 연관하에서 분석하면서도 단순히 데이터마이닝적 성격에 그치지 않고, 상관관계와 인과관계를 파악해 소셜 감성으로 암호화폐 가격과의 관계를 설명하려고 한다. 특히 한국의 암호화폐 시장을 대상으로 한 연구는 아직 미비한 실정이기 때문에, 본 연구에서는 한국의 암호화폐 거래소를 대상으로 실증 분석을 실시하고자 한다.

III. 연구 분석 프로세스 및 가설

본 연구의 분석 프로세스는 Bollen *et al.*(2011)의 모형을 수정하여 <그림 1>과 같이 제시하였다. 연구는 제시된 <그림 1>에 보이는 바와 같이 총 세 단계에 걸쳐서 진행된다. 첫 번째 단계에서는 트위터 게시물과 비트코인 가격에 대한 일일 데이터를 수집하고, 수집된 트위터 게시물에 대해서는 감성 분석을 시행해 중립어와 긍·부정어로 분류한다. 다음으로 트위터의 중립어 수 및 긍·부정어 수가 비트코인 가격에 영향을 미치는지 알아보기 위해 회귀분석을 실시한다. 회귀분석을 통해 상관관계를 살펴본 후에는 마지막으로 인과



〈그림 1〉 Research Analysis Process

관계 존재여부를 Granger Causality test를 실시하여 알아본다. 이와 같은 모든 과정은 프로그래밍 언어인 파이썬(Python)과 R을 사용하여 이루어지는데, 주로 데이터 수집 단계에서는 파이썬을, 나머지 데이터 처리 및 분석 과정에서는 R을 사용하였다.

다음은 이를 바탕으로 본 연구에서 다루고자 하는 핵심 주제들을 가설로 설정해 본 것이다.

H1: 트위터에서의 소셜 감성은 비트코인 가격에 영향을 미칠 것이다.

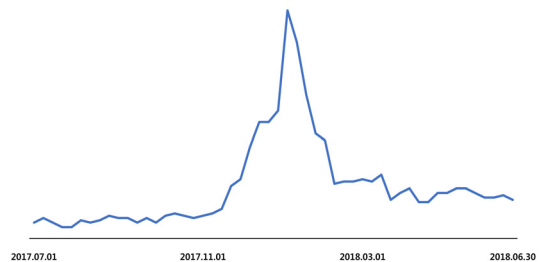
H2: 트위터에서의 소셜 감성과 비트코인 가격 사이에 인과관계가 존재할 것이다.

수립된 가설을 보면 H1은 독립변수인 소셜 감성과 종속변수인 비트코인 가격의 상관관계를 검증하고자 하는 것이다. 이는 통계적으로 두 변수간의 관계를 우선 살펴보고자 하는 것이다. 그런데, 소셜 감성이 암호화폐 투자에 영향을 미친다면 이는 사람들의 상관관계가 아닌 인과관계로 설명하여야 암호화폐 투자가 사회적 분위기 및 다른 사람들의 감정에 영향을 받는다고 결론을 내릴 수 있을 것이다. 따라서 H2에서는 소셜감성의 변화가 암호화폐 투자자의 행동에 인과적 영향을 미친다는 인과관계를 검정하기 위한 가설을 설정하였다.

IV. 연구방법

4.1 데이터

본 연구를 위한 모든 데이터는 일일 데이터로 수집하였으며, 수집 기간은 <그림 2>를 참고하여 사람들의 관심이 증가하는 2017년 11월부터 점차 사그라드는 2018년 4월까지 총 6개월(181일)로 설정하였다.



〈그림 2〉 Google Trends에서의 암호화폐에 대한 관심도 변화¹⁾

먼저 소셜 감성 데이터는 소셜 미디어 사이트 ‘트위터’를 대상으로 비트코인과 관련해 게시된

1) Google Trends, 키워드 “암호화폐” 검색 결과, 2018년 12월 기준, https://trends.google.co.kr/trends/explore?date=2017-07-01%202018-06-30&geo=KR&q=%2Fm%2F0vpj4_b.

글들을 크롤링하여 수집하였다. 트위터는 API를 제공하고 있어 보다 쉽게 게시물들을 크롤링 할 수 있다는 장점이 있지만, 해당 방법으로는 대량의 데이터나 일정 기간 이전의 과거 데이터에는 접근이 불가능하다는 문제점도 있다. 따라서 본 연구에서는 파이썬의 Selenium 라이브러리를 사용하여 직접 웹 크롤러(web crawler)를 구축해 트위터 데이터를 수집하였다.

처음에는 <그림 3>과 같이 총 154,783건의 트윗(tweet)이 수집되었다. 다만 해당 데이터에는 광고, 도배 등 분석에 부적합한 노이즈 데이터가 포

함되어 있으므로 이를 필터링(filtering)하여 <그림 4>와 같이 총 100,120건의 데이터를 확보했다. 100,120건의 데이터는 다시 전처리 과정을 거치고 <그림 5>와 같이 분석에 활용 가능한 형태로 변환시켰다. 전처리는 감성 분석 과정 중 가장 많은 시간이 소요되는 단계이지만 분석 정확도를 제고하기 위해서는 필수적으로 선행되어야 한다. 전처리 작업을 거치지 않은 데이터에는 불완전(incomplete) 문제, 노이즈(noise) 문제, 비일관성(inconsistent) 문제가 나타날 수 있기 때문이다 (Zhang *et al.*, 2003).

Date	Link	Text
2017-11-01	https://twi	비트코인...
2017-11-01	https://twi	자기 연준 의장 파월, '비트코인 싫어하진 않지만...' #Bitcoin \$BTC #비트코인
2017-11-01	https://twi	[비트코인] 곧 상품거래소에서 거래예상 http://slownote.kr/221131047527
2017-11-01	https://twi	일년만에 10배가 났 비트코인 당신은 도덕책...
2017-11-01	https://twi	비트코인 800만원 찍었네.
2017-11-01	https://twi	[다나와] 로지텍 MK270r 무선 콤보 (정품) 최저가 27,710원 http://prod.danawa.com/info/?pcode=2541329 ... #엘지 #삼성 #
2017-11-01	https://twi	비트코인 800만원 돌파.."주류 금융시장 편입 기대감" 다음뉴스
2017-11-01	https://twi	비트코인 지금이라도 사야하는걸까
2017-11-01	https://twi	비트코인가격,전망 암호화폐관련 오늘의코인 http://blog.naver.com/sealove0721/221131040433 ...
2017-11-01	https://twi	하루1번 비트코인 이야기 보는거같은 비트코인 트위터에서 떠서 곧 망할듯
2017-11-01	https://twi	어 비트코인 계속 오르네 진짜 곧 이민가겠네
2017-11-01	https://twi	#비트코인 삽니다 카톡:didid8989#비트코인 팝니다#비트코인 구입#비트코인 판매#상품권판매#상품권구매#각종상품권삽니다
2017-11-01	https://twi	#비트코인월리 = r49q0m5f1 【코인에버】#비트코인국제송금 코인에버 #비트코인국제송금 #비트코인 #비트코인가거래소 #비트
2017-11-01	https://twi	#비트코인보안 1a3j9q2pm18 ?비트코인??코인에버 #비트코인가거래소 코인에버 #비트코인시세 코인에버 #비트코인 #비트코인
2017-11-01	https://twi	#&카드한도대출 #&비트코인 대출 #&카드사무실가맹문의환영 #&신용카드 로 비트코인 구매 #&비트코인카드구매 #한 #&
2017-11-01	https://twi	통과했네 통과했어 800만원이네 이제...비트코인이 채고시다
2017-11-01	https://twi	비트코인을 사났어야해...
2017-11-01	https://twi	비트코인이 대박인가? 사촌동생 얼마 전에 적금 다 깨고 전부 비트코인 샀다던데.
2017-11-01	https://twi	비트코인 내가 돈만 있으면 투자하는데...

<그림 3> Raw Data Collected by Twitter

Date	Link	Text
2017-11-01	https://twi	비트코인...
2017-11-01	https://twi	자기 연준 의장 파월, '비트코인 싫어하진 않지만...'
2017-11-01	https://twi	[비트코인] 곧 상품거래소에서 거래예상 http://slownote.kr/221131047527
2017-11-01	https://twi	일년만에 10배가 났 비트코인 당신은 도덕책...
2017-11-01	https://twi	비트코인 800만원 찍었네.
2017-11-01	https://twi	비트코인 800만원 돌파.."주류 금융시장 편입 기대감"
2017-11-01	https://twi	비트코인 지금이라도 사야하는걸까
2017-11-01	https://twi	비트코인가격,전망 암호화폐관련 오늘의코인 http://blog.naver.com/sealove0721/221131040433 ...
2017-11-01	https://twi	하루 1번 비트코인 이야기 보는거같은 비트코인 트위터에서 떠서 곧 망할듯
2017-11-01	https://twi	어 비트코인 계속 오르네 진짜 곧 이민가겠네
2017-11-01	https://twi	통과했네 통과했어 800만원이네 이제...비트코인이 채고시다
2017-11-01	https://twi	비트코인을 사났어야해...
2017-11-01	https://twi	비트코인이 대박인가? 사촌동생 얼마 전에 적금 다 깨고 전부 비트코인 샀다던데.
2017-11-01	https://twi	비트코인 내가 돈만 있으면 투자하는데...
2017-11-01	https://twi	나도 비트코인으로 돈벌었는데 투자액이 작아서 퇴사도 뭣도 못함. 돈이 돈을 버는거다 πππ
2017-11-01	https://twi	소프트뱅크, 포트리스 인수로 비트코인 1580억원 '횡재'
2017-11-01	https://twi	바로 24시간전에 수중에있던 10만원 비트코인개시에 넣어놔으면 오늘 저녁으로 소고기 무한리필을 먹을수있었던...#비트코인
2017-11-01	https://twi	비트코인 4시간봉 분석 2 \$BTCUSD http://kr.tradingview.com/chart/BTCUSD/4M3hpYFT/ ...
2017-11-01	https://twi	비트코인 살까 진지하게 고민했던 때 인생에 두번 있었고 두번 다 날려버림ㅋㅋㅋㅋ

<그림 4> Filtering Data Set

Date	Link	Text
2017-11-01	https://twi	비트코인
2017-11-01	https://twi	차기 연준 의장 파월 비트코인 싫어하진 않지만
2017-11-01	https://twi	비트코인 곧 상품거래소에서 거래예상
2017-11-01	https://twi	일년만에 배가 뽀뽀 비트코인 당신은 도덕책
2017-11-01	https://twi	비트코인 만원 찍었네
2017-11-01	https://twi	비트코인 만원 돌파 주류 금융시장 편입 기대감
2017-11-01	https://twi	비트코인 지금이라도 사야하는걸까
2017-11-01	https://twi	비트코인가격 전장 암호화폐관련 오늘의코인
2017-11-01	https://twi	하루 번 비트코인 이야기 보는거같음 비트코인 트위터에서 떠서 곧 망할듯
2017-11-01	https://twi	어 비트코인 계속 오르네 진짜 곧 이민가겠네
2017-11-01	https://twi	통과했네 통과했어 만이네 이제 비트코인이 채고시다
2017-11-01	https://twi	비트코인을 사봤어야해
2017-11-01	https://twi	비트코인이 대박인가 사촌동생 얼마 전에 적금 다 깨고 전부 비트코인 샀다던데
2017-11-01	https://twi	비트코인 내가 돈만 있으면 투자하는데
2017-11-01	https://twi	나도 비트코인으로 돈벌었었는데 투자액이 작아서 퇴사도 뭣도 못함 돈이 돈을 버는거다
2017-11-01	https://twi	소프트뱅크 포트리스 인수로 비트코인 억원 횡재
2017-11-01	https://twi	바로 시간전에 수중에있던 만원 비트코인캐시에 넣어놨으면 오늘 저녁으로 소고기 무한리필을 먹을수있었던 비트코인에 미친
2017-11-01	https://twi	비트코인 시간봉 분석
2017-11-01	https://twi	비트코인 살까 진지하게 고민했던 때 인생에 두번 있었고 두번 다 날려버림

〈그림 5〉 Cleaning Data Set

본 연구에서는 먼저 트위터상에서 추출한 데이터를 CSV(Comma Separated Values) 형식의 반정형 구조로 변경시키기 위해 하나의 트윗 내용은 한 줄로 정리하였다(전희국 등, 2014). 엑셀 형식에서는 개행 문자가 들어간 트윗도 하나의 데이터로 인식되지만 CSV 형식에서는 줄 단위로 데이터를 인식하기 때문에 이와 같이 전처리 과정에서 줄 바뀔 현상을 없애는 작업이 필요하다. 이 작업은 R에서 replace 함수와 정규표현식을 사용하면 간단하게 처리할 수 있다. 그 다음엔 같은 의미를 가졌지만 축약되어 사용됐거나 다른 형태로 사용된 단어들을 하나의 형태로 통일하였다. 예를 들어, ‘김프’, ‘비캐’, ‘비꿀’과 같이 축약된 단어들이 있으면 ‘김치프리미엄’, ‘비트코인캐시’, ‘비트코인골드’와 같은 원형으로, ‘스테이터스 네트워크 토큰’, ‘슨트’, ‘SNT’ 등 다양한 형태로 사용된 단어는 ‘스테이터스네트워크토큰’의 원형으로 통일시키는 것이다. 마찬가지로 영문과 한자 어휘도 한글로 형태를 통일해주었다. 마지막으로 URL과 특수문자, 숫자 등을 제거하여 분석에 필요한 데이터만 남겨주었다. 그리고 이렇게 전처리가 끝난 데이터는 형태소 분석 및 감성 분석을 시행하여 중립과 긍·부정으로 분류하였다.

다음으로 비트코인 가격에 대한 데이터를 수집

하기 위해 한국에서 가장 많은 암호화폐 투자자들이 이용하는 거래소 사이트인 ‘빗썸’과 ‘업비트’를 선정하였다. 암호화폐 시장은 주식 시장처럼 시간이 정해져 있지 않고 24시간 개방되어 있으므로 비트코인 가격에 대한 데이터는 암호화폐 거래소 빗썸과 업비트가 제공하는 종가를 기준으로 한다. 이때, 빗썸은 그 기준이 자정, 업비트는 오전 9시이다. 추가로 금융시장의 주요 자산인 주식 시장과 암호화폐 시장은 동일한 움직임을 보일 수 있으므로 KOSPI 지수를 통제변수로 포함하였다.

따라서 총 변수는 <표 1>과 같이 빗썸 비트코인 가격(BBP), 업비트 비트코인 가격(UBP), 중립어 수(NEU), 긍정어 수(POS), 부정어 수(NEG), 코스피 지수(KOSPI)가 된다.

〈표 1〉 Variable and Definition

Variable	Definition
i	Individual cryptocurrency market (Bithumb, Upbit)
t	Calendar time in days(1, 181)
BBP	Bithumb Bitcoin Price
UBP	Upbit Bitcoin Price
NEU	Neutral Words
POS	Positive Words
NEG	Negative Words
KOSPI	KOSPI Index

4.2 형태소 분석

감성 분석을 수행하기 위해서는 먼저 형태소 분석을 통해 필요한 형태소들만 추려주어야 한다. 형태소란 뜻을 가지고 있는 최소한의 말의 단위이며, 형태소 분석은 문서 혹은 문장을 이런 형태소 단위로 추출하는 방법이다. 본 연구에서는 맞춤법 등이 정확하게 지켜지지 않는 트위터 데이터의 특성을 고려하여 키워드가 되는 명사만 분리하는 방식을 택하였다. 이를 위해 R과 파이썬에 내장되어 있는 대표적인 형태소 분석 패키지인 꼬꼬마(Kkma), 코모란(Komoran), 한나눔(Hannanum), 그리고 가장 최근에 개발된 RHINO까지 전부 실행하여 비교해보았다. 그 결과, 명사는 RHINO에서 가장 명확하게 분리되었으므로 이를 채택해 전처리를 마친 데이터 약 10만 건의 형태소 분석을 수행하였다. 경희대학교 빅데이터 연구센터에서 개발한 RHINO 분석기는 국내 1,200만 어절을 수작업으로 분석한 말뭉치를 활용하였으

므로 분석 결과의 정확도가 높으며, <표 2>와 같이 다른 형태소 분석기들과 달리 띄어쓰기 단위가 아닌 문맥에 맞는 결과를 제시하므로 보다 명확한 결과를 얻을 수 있다.

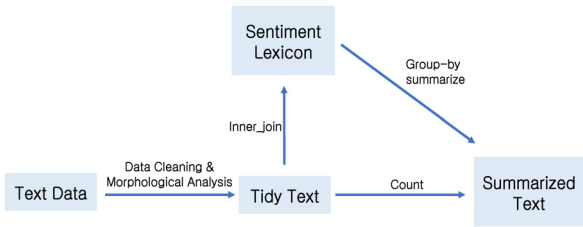
4.3 감성 분석

오피니언 마이닝(opinion mining)이라고도 불리는 감성 분석은 텍스트에 나타난 사람들의 느낌이나 감정, 주관성 등을 분석하는 자연어 처리(Natural Language Processing, NLP) 기술이다. 따라서 텍스트의 주제보다 텍스트가 어떠한 감성을 나타내고 있는가에 초점을 맞춰 분석해야 한다. 감성 분석을 활용하면 기본적으로 문서 또는 문장이 어떤 견해를 갖고 있는지 판별할 수 있으므로 감성 분석은 소셜 미디어에서 발생하는 데이터, 즉 소셜 감성을 분석하는데 유용한 대안이 된다(김영민 등, 2014).

감성 분석과 관련한 상당수의 연구에서 감성은

<표 2> ‘아버지가방에들어가신다’에 대한 형태소 분석 예시

Kkma	Komoran	Hannanum	RHINO
아버지 / NNG	아버지가방에들어가신다 / NNP	아버지가방에들어가 / N	아버지 / NNG
가방 / NNG		이 / J	가 / JKS
에 / JKM		시니다 / E	방 / NNG
들어가 / VV			에 / JKB
시 / EPH			들 / VV
니다 / EFN			어 / EC
			가 / JKS
			신 / VV
			다 / MAG



<그림 6> Sentiment Analysis Process

일반적으로 극성(polarity)에 따라 ‘긍정’과 ‘부정’ 양 극으로 구분되는 이분법적 분류를 하거나 혹은 ‘중립’을 포함한 삼분법적 분류를 한 뒤 사용한다. 이때 극성은 주로 각 문서 최소 단위인 어휘의 감성 극성(sentiment polarity)을 기반으로 이루어지므로 감성을 분류하기 위해선 감성 극성이 미리 정의된 감성 어휘 사전(sentiment lexicon)이 필요하다. 하지만 특정 어휘의 극성은 한 가지로 고유하게 정해져 있지 않으며 분석의 목적에 따라 그 극성이 상이하게 나타날 수 있다(유은지 등, 2013). 즉 도메인(domain)에 따라 동일한 어휘여도 긍정/부정/중립이 다르게 판별될 수 있다는 뜻이다. 따라서 범용 감성사전을 사용해 감성 분석을 실시할 경우에는 결과에 대한 신뢰성이 떨어질 수도 있다. 예를 들어, ‘사표’, ‘퇴사’라는 어휘는 일반적으로 부정성을 의미하지만 암호화폐 시장에서는 긍정적으로 인식되기 때문에 이러한 경우 범용 감성사전

을 사용하면 전혀 다른 결과가 나타날 수도 있다. 이에 본 연구에서는 암호화폐 시장에 맞는 감성사전을 직접 구축하여 수집된 데이터들의 극성을 긍정/부정/중립으로 분류하였다.

감성사전은 주로 각 어휘와 해당 어휘의 극성 혹은 극성과 감성 점수가 함께 이루어져 있다. 따라서 감성사전을 구축하기 위해서는 가장 먼저 사전에 포함할 어휘들을 결정해야 한다. 만약 사전에 어휘가 미리 등록되어 있지 않으면 트윗에 해당 어휘가 존재하더라도 감성 분석 시에는 제외될 수 있다. 본 연구에서는 수집된 트위터 데이터에 형태소 분석을 실시하고, 형태소 분석을 통해 나온 명사들을 기반으로 감성사전을 구축했기 때문에 이와 같은 사례는 발생하지 않을 것이다. 다만 암호화폐 시장에는 ‘존버’, ‘가즈아’와 같이 새롭게 등장한 신조어가 많기 때문에 이러한 어휘들은 따로 감성사전에 추가해주었다. 그 다음

〈표 3〉 감성 어휘 사전 예시

중립어	긍정어	부정어
가격, 가상통화, 가상화폐, 개미, 거래량, 거래소, 결제, 경제, 국가, 국내, 국민, 금값, 금리, 기자, 나스닥, 단타, 대학생, 라이트코인, 랜섬웨어, 리플, 리플코인, 마이닝, 미국, 미래, 바이낸스, 발표, 방법, 방송, 변동성, 분석, 블록체인, 비트렉스, 비트코인, 비트코인골드, 비트코인다이아몬드, 비트코인캐시, 비트코인플래티넘, 빗썸, 사토시 나카모토, 상품, 상황, 선물, 세그윗, 세금, 수수료, 시세, 시장, 시황, 실물, 아시아, 알고리즘, 알트코인, 암호화폐, 양자컴퓨터, 언론, 업비트, 에이다, 여론, 예상, 유행어, 은행, 이더리움, 일본, 장타, 재산, 전망, 전문가, 정보, 정부, 정책, 중국, 창시자, 채굴, 청년, 추천, 친구, 카드, 코리아, 코빗, 코스닥, 코인네스트, 코인원, 타임머신, 테더, 통화, 투자, 투자금, 투자자, 트레이딩, 트렌드, 트론, 패러다임, 플랫폼, 하드포크, 한국, 해외, 현금, 현실, 현황, 환율, 환전	가즈아, 각광, 강세, 강추, 갱신, 경신, 경축, 고성장, 고수익자, 골든크로스, 관심, 급등, 급등세, 급반등, 급부상, 급상승, 급성장, 긍정, 기대, 기대감, 기원, 기적, 기폭제, 기회, 낙관, 달성, 대박, 대세, 도달, 도약, 돈방석, 돈벼락, 돌파, 동아줄, 떡상, 떼돈, 떼부자, 랠리, 만세, 만족, 믿음, 반등, 반등세, 벼락부자, 비전, 상승, 상승률, 상승세, 상승장, 상향, 성공, 성공기, 성공담, 성장, 성장세, 성황, 수익, 수익금, 수익률, 수혜, 순기능, 승리, 승승장구, 승자, 신고가, 역전, 역헤드앤숄더, 열정, 오름, 오름세, 우상향, 이득, 이익, 인기, 잠재력, 장밋빛, 장점, 짝꿍, 전성기, 정점, 증가, 질주, 차익, 찬스, 천정부지, 최고, 최대, 출범, 퇴사, 퇴사자, 퇴직, 투더문, 폭등, 폭증, 플러스, 한몫, 해피, 행복, 행복회로, 행운, 행진, 호재, 호조, 호황, 환영, 환호, 활황, 회복, 회복세, 획득, 훈풍, 흑자, 흡족, 흥미, 희망, 히트	거품, 경고, 고통, 공매도, 공포, 과열, 광기, 급락, 김치프리미엄, 끝물, 낙폭, 난리, 널뛰기, 논란, 논쟁, 눈물, 단단게, 단속, 대출, 도박, 돈세탁, 떡락, 롤러코스터, 마이너스, 몰락, 문제점, 바닥, 박살, 버블, 부작용, 부정, 분노, 불법, 불안, 붕괴, 비명, 사기, 사행성, 세력, 손실, 손해, 수사, 스캠, 신기루, 실패, 아우성, 악재, 압박, 엄단, 욕심, 우울증, 위기, 위험, 위협, 유출, 음모, 의심, 의혹, 자살, 작전, 잘못, 재앙, 적발, 제로, 제재, 제한, 조작, 존버, 좀비, 종말, 주의, 중단, 중독, 중지, 중발, 짜증, 쪽박, 차단, 처벌, 철회, 최악, 추락, 충격, 큰손, 큰일, 탈세, 탐욕, 토막, 통제, 투기, 투기자, 틀림, 파산, 패닉, 패닉셀, 펌핑, 폐쇄, 폐지, 포기, 폭락, 폭망, 폭탄, 폭탄돌리기, 폰지사기, 프리미엄, 하락, 하락세, 하락장, 한강, 한강 가즈아, 한계, 한탕, 한탕주의, 해커, 해킹, 허상, 호구, 혼란, 환치기, 횡령, 후회

에는 각 어휘에 극성 및 범주 값을 부여해야 하는데, 이때 주로 Manning *et al.*(2015)이 제시한 TF-IDF (Term Frequency, Inverse Document Frequency) 방법론이 사용된다. TF-IDF는 특정 단어가 문서 내에서 얼마나 자주 등장했는지 단어의 빈도수 및 중요성을 정량화 시킬 때 쓰는 방법이다. 하지만 본 연구는 명사만 추출하여 카테고리화했기 때문에 TF-IDF를 적용하지 않고(최기철, 이상용, 2018) 개별적으로 극성만 부여하였다. 중립 및 긍정·부정에 관한 극성 값은 관련 연구자 총 20인의 의견을 반영하여 결정되었으며, 최종적으로 긍정어 282개, 부정어 794개, 중립어 21,549개가 도출되었다. 도출된 중립어와 긍·부정어의 예시는 <표 3>에 나타나있다.

4.4 단위근 검정

소셜 감성과 비트코인 가격간의 동학적 관계를 살펴보기 전에 먼저 단위근 검정(unit root test)을 통해 각 시계열 데이터의 정상성(stationary) 여부를 확인하였다. 만약 단위근이 존재하는 non-stationary한 시계열 데이터를 그대로 사용하면 상관관계가 없는 변수간에도 매우 강한 상관관계가 있는 것처럼 보이는 가성회귀(spurious regression) 문제가 발생할 수 있기 때문이다.

대표적인 단위근 검정 방법에는 DF(Dickey-Fuller), ADF(Augmented Dickey-Fuller), PP(Phillips-Perron) 검정법이 있는데, Dickey and Fuller(1979)가 제시한 ADF(Augmented Dickey-Fuller) 검정이 가장 보편적이므로 본 연구에서도 이를 이용하여 단위근 검정을 실시하였다.

4.5 회귀분석

독립변수인 중립어 수, 긍정어 수, 부정어 수가 종속변수인 비트코인의 가격에 미치는 전반적인 관계를 살펴보기 위하여 회귀분석을 시행하였다. 기본 회귀식은 다음과 같다.

$$\text{BitcoinPrice}_{it} = \beta_0 + \beta_1 \text{NEU}_t + \beta_2 \text{POS}_t + \beta_3 \text{NEG}_t + \beta_4 \text{KOSPI}_t + \epsilon_{it}$$

여기서 BitcoinPrice_{it} 는 t시점에서 i거래소의 비트코인 가격을 나타내고, β_0 는 y절편을 의미한다. 나머지 $\beta_1, \beta_2, \beta_3$ 는 순서대로 t시점에 대한 중립어 수, 긍정어 수, 부정어 수이며, β_4 는 통제변수로서 t시점의 KOSPI 지수이다. ϵ_{it} 는 t시점에 i거래소의 잔차항이다.

4.6 Granger Causality Test

Granger(1969)에 의해 최초로 제안된 Granger Causality test는 두 변수 사이에 인과관계가 존재하는지를 검정하는 기법이다. 변수들이 관계가 있음은 알고 있지만 어떤 변수가 원인 변수이고 어떤 변수가 결과 변수인지 불분명할 경우 사용된다.

본 연구에서 회귀분석을 통해 나온 결과가 소셜 감성의 노출과 비트코인 가격에 상관관계가 있다고 하더라도, 이것만으로는 소셜 감성의 변화가 비트코인 가격에 인과관계적 영향을 미친다고 단정할 수는 없다. 비트코인 가격에 변동이 생김으로 인해 많은 사람들이 비트코인을 논하게 되는 숫자가 증가하는 역의 인과관계가 존재할 수 있기 때문이다. 따라서 비트코인 가격이 소셜 감성의 변화에 영향을 미치는지, 아니면 소셜 감성의 변화가 비트코인 가격 변동에 영향을 미치는지를 살펴보기 위하여 Granger Causality test를 실시하였다(Kim *et al.*, 2015). 이를 두 회귀방정식으로 나타내면 다음과 같다.

$$Y_t = \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{j=1}^p \beta_j X_{t-j} + \epsilon_{1t}$$

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

여기서 y는 비트코인 가격을 나타내고, x는 중립어 수, 긍정어 수, 부정어 수를 각각 대입하였다. 그리고 시차는 2시차부터 5시차까지 적절한 구간을 실행해보고 최종적으로 5시차로 설정하였다.

V. 분석 결과

5.1 단위근 검정 결과

각 시계열 데이터들의 단위근 존재 여부를 파악하기 위해 다음과 같이 ADF 검정 방법을 사용하여 단위근 검정을 실시하였다.

〈표 4〉 Unit Root Test

변수	원시계열		차분시계열	
	t	p	t	p
BBP	-2.05	0.56	-12.70	0.00***
UBP	-1.98	0.58	-11.54	0.00***
NEU	-3.53	0.04**	-15.27	0.00***
POS	-4.32	0.01**	-16.95	0.00***
NEG	-3.90	0.02**	-16.78	0.00***
KOSPI	-2.46	0.39	-14.52	0.00***

〈표 4〉의 단위근 검정 결과를 보면, 1% 유의수준에서 모든 시계열이 단위근을 가지는 것으로 나타나 본 시계열이 nonstationary한 것이 확인되었다. 이에 따라 시계열을 차분하여 정상 시계열로 변환하고 회귀분석과 Granger Causality test에 투입하였다. 이때 정상성을 만족시키기 위한 차수를 알아내기 위해 R의 forecast 패키지를 사용하였다. 그 결과, 종속변수인 비트코인 가격과 독립변수인 중립어 수, 긍정어 수, 부정어 수, 그리고 통제변수인 KOSPI 지수 모두 값이 1로 나와 1차 로그 차분을 취해 시계열의 안정성을 확보하였다.

5.2 회귀분석 결과

감성 분석을 통해 추출된 소셜 감성 데이터가 비트코인 가격에 영향을 주는지 확인하기 위해 회귀분석을 실시하여 그 상관관계를 살펴보았다.

〈표 5〉 Regression Analysis(BBP)

종속변수	독립변수	Co eff.	t	p
BBP	Constant	14331.1 (65430.0)	0.219	0.827
	NEU	448.8 (100.7)	4.455	0.000***
	POS	4833.5 (2096.5)	2.305	0.022*
	NEG	-4942 (698.0)	-7.080	0.000***
	KOSPI	-4714.3 (3924.4)	-1.201	0.231
	통계량	$R^2 = 0.227$, $F = 14.17$, $p = 0.000$		

〈표 5〉는 암호화폐 거래소 빗썸의 비트코인 가격과 중립어 수, 긍정어 수, 부정어 수의 상관관계를 분석한 결과이다. 중립어 수는 정(+)의 관계를 보였고, 긍정어 수도 정(+)의 관계로 나타났다. 반면 부정어 수는 부(-)의 관계를 보였다. 즉, 트위터상에서 비트코인에 대한 중립적 감성이나 긍정적 감성이 증가하면 빗썸의 비트코인 가격이 상승하고, 부정적인 감성이 증가하면 빗썸의 비트코인 가격은 하락한다고 볼 수 있다. 따라서 소셜 감성과 빗썸 비트코인 가격의 전반적인 관계는 가설과 일치한다고 할 수 있다.

〈표 6〉 Regression Analysis(UBP)

종속변수	독립변수	Co eff.	t	p
UBP	Constant	13834.0 (72094.7)	0.192	0.848
	NEU	61.27 (111.0)	0.552	0.582
	POS	11721.3 (2310.0)	5.074	0.000***
	NEG	-3567.8 (769.1)	-4.639	0.000***
	KOSPI	-6911.5 (4324.1)	-1.598	0.112
	통계량	$R^2 = 0.187$, $F = 11.30$, $p = 0.000$		

<표 6>은 업비트 비트코인 가격에 대한 회귀분석 결과이다. 먼저 회귀계수를 살펴보면, 빗썸 비트코인 가격 모형에서와 같이 긍정어 수, 부정어 수 모두 유의함을 알 수 있다. 즉, 사람들이 트위터에 비트코인에 대해 긍정적인 의견을 많이 올리면 업비트의 비트코인 가격은 상승한다. 반대로 사람들이 트위터에 비트코인에 대해 부정적인 의견을 공유하면 업비트 비트코인 가격은 하락한다. 따라서 트위터에서의 소셜 감성이 비트코인 가격에 영향을 미치지 못할 것이라는 귀무가설은 기각된다.

반면 중립어 수는 통계적으로 유의미한 관계가 없었는데, 이는 앞서 빗썸의 비트코인 가격 모형에서 중립어 수가 유의하게 나타났던 것과 다소 차이를 보이는 결과이다. 거의 동일한 가격 변동을 보이는 암호화폐 거래소의 특성상 이와 같은 결과는 단순히 각 거래소에서 제공하는 종가 기준에 따른 것으로 보인다. 빗썸의 종가는 자정, 업비트의 종가는 오전 9시라는 점을 고려하면, 비트코인에 대한 중립적인 의견이 비트코인 가격에 미치는 영향은 점차 시간이 지남에 따라 사라진다고 할 수 있다.

5.3 Granger Causality Test 결과

앞서 회귀분석 결과를 보면, 비트코인에 대한 중립어 수, 긍정어 수, 부정어 수의 증감이 비트코인 가격 변동과 관계가 있다고 판단된다. 하지만 이는 단순히 상관관계를 파악한 것에 지나지 않기 때문에, 추가적으로 Granger Causality test를 실시해 중립어 수, 긍정어 수, 부정어 수 각각이 실질적으로 비트코인 가격과 인과관계가 있는지 확인하였다.

<표 7>은 중립어 수와 빗썸 비트코인 가격과의 인과관계를 분석한 결과이다. 이를 보면 중립어 수는 4, 5시차에서 유의하게 나타났고, 빗썸 비트코인 가격은 모든 시차에서 유의하였다. 이러한 결과는 중립어 수와 빗썸 비트코인 가격간에 인과관계

가 미약하지만 양방향으로 존재하는 것을 의미한다. 다만 빗썸 비트코인 가격으로부터 중립어 수로의 인과관계가 보다 강한 것으로 볼 수 있다.

<표 7> Granger Causality Test(NEU & BBP)

lags	Null Hypothesis	Obs	F	p
1	NEU \nRightarrow BBP	177	2.063	0.153
	BBP \nRightarrow NEU		10.516	0.001**
2	NEU \nRightarrow BBP	175	2.929	0.056
	BBP \nRightarrow NEU		5.519	0.005**
3	NEU \nRightarrow BBP	173	2.600	0.054
	BBP \nRightarrow NEU		5.053	0.002**
4	NEU \nRightarrow BBP	171	2.850	0.026*
	BBP \nRightarrow NEU		4.665	0.001**
5	NEU \nRightarrow BBP	169	2.815	0.018*
	BBP \nRightarrow NEU		4.034	0.002**

<표 8> Granger Causality Test(POS & BBP)

lags	Null Hypothesis	Obs	F	p
1	POS \nRightarrow BBP	177	8.084	0.005**
	BBP \nRightarrow POS		11.203	0.000***
2	POS \nRightarrow BBP	175	5.426	0.005**
	BBP \nRightarrow POS		4.237	0.016*
3	POS \nRightarrow BBP	173	3.787	0.012*
	BBP \nRightarrow POS		3.172	0.026*
4	POS \nRightarrow BBP	171	5.121	0.000***
	BBP \nRightarrow POS		2.742	0.030*
5	POS \nRightarrow BBP	169	4.287	0.001**
	BBP \nRightarrow POS		1.979	0.084

긍정어 수와 빗썸 비트코인 가격 사이에 인과관계가 존재하는지 알아보기 위하여 Granger Causality test를 실시해 <표 8>에 요약하였다. 먼저 긍정어 수를 보면 1시차부터 5시차까지 전부 유의미한 영향력을 보였음을 알 수 있는데, 이는 곧 트위터상에서의 긍정적 감성이 빗썸의 비트코인 가격에 영향을 미친다는 것을 의미한다. 하지만 반대로 빗썸의 비트코인 가격도 1시차부터 4시차까지 유의미하게 나타났으므로 서로 영향을

주요받는 양방향 인과관계가 있다고 할 수 있다.

<표 9> Granger Causality Test(NEG & BBP)

lags	Null Hypothesis	Obs	F	p
1	NEG \Rightarrow BBP	177	2.740	0.100
	BBP \Rightarrow NEG		5.472	0.020*
2	NEG \Rightarrow BBP	175	4.061	0.019*
	BBP \Rightarrow NEG		3.094	0.048*
3	NEG \Rightarrow BBP	173	3.989	0.009**
	BBP \Rightarrow NEG		2.653	0.050
4	NEG \Rightarrow BBP	171	3.575	0.008**
	BBP \Rightarrow NEG		3.283	0.013*
5	NEG \Rightarrow BBP	169	3.039	0.012*
	BBP \Rightarrow NEG		2.765	0.020*

<표 9>는 부정어 수와 빗썸 비트코인 가격 사이의 인과관계를 분석한 결과표이다. 결과값을 보면 알 수 있듯이, 부정어 수는 처음 1시차에서는 빗썸 비트코인 가격에 영향을 미치지 못하다가 2시차부터 지속적으로 가격 변화에 영향을 주었다. 이는 ‘트위터에서의 소셜 감성과 비트코인 가격 사이에 인과관계가 존재할 것이다.’라는 가설을 입증하였지만, 빗썸 비트코인 가격도 3시차를 제외하고는 1시차부터 계속해서 통계적으로 유의한 결과를 보이고 있어 본 모형의 인과관계는 양쪽으로 방향성이 있다고 할 수 있다.

<표 10> Granger Causality Test(NEU & UBP)

lags	Null Hypothesis	Obs	F	p
1	NEU \Rightarrow UBP	177	1.476	0.226
	UBP \Rightarrow NEU		0.001	0.977
2	NEU \Rightarrow UBP	175	1.595	0.206
	UBP \Rightarrow NEU		6.922	0.001**
3	NEU \Rightarrow UBP	173	4.108	0.008**
	UBP \Rightarrow NEU		8.465	0.000***
4	NEU \Rightarrow UBP	171	3.092	0.017*
	UBP \Rightarrow NEU		6.253	0.000***
5	NEU \Rightarrow UBP	169	5.145	0.000***
	UBP \Rightarrow NEU		5.391	0.000***

<표 10>은 ‘Granger Causality test(NEU & UBP)’ 모형에서의 시차별 관측치, F통계량 값, 유의확률을 보여준다. 이를 보면 <표 7>의 빗썸 비트코인 가격과 중립어 수에 대한 Granger Causality test와 결과가 유사함을 알 수 있다. 본 모형에서도 중립어 수는 업비트 비트코인 가격에 3시차부터 유의한 결과를 보여주었고, 업비트 비트코인 가격 또한 2시차부터 5시차까지 통계적으로 유의하였다. 따라서 이는 가설 H2를 지지하는 결과라 할 수 있다. 다만 업비트 가격으로부터 중립어 수로의 인과관계가 중립어 수로부터 업비트 가격으로의 인과관계보다 강하기 때문에 이 점에 대해 유의할 필요가 있다.

<표 11> Granger Causality Test(POS & UBP)

lags	Null Hypothesis	Obs	F	p
1	POS \Rightarrow UBP	177	1.791	0.183
	UBP \Rightarrow POS		0.137	0.712
2	POS \Rightarrow UBP	175	1.486	0.229
	UBP \Rightarrow POS		7.453	0.000***
3	POS \Rightarrow UBP	173	6.160	0.000***
	UBP \Rightarrow POS		8.139	0.000***
4	POS \Rightarrow UBP	171	4.930	0.000***
	UBP \Rightarrow POS		5.960	0.000***
5	POS \Rightarrow UBP	169	5.805	0.000***
	UBP \Rightarrow POS		4.472	0.000***

<표 11>은 긍정어 수와 업비트 비트코인 가격의 Granger Causality test 결과를 보여준다. 먼저 긍정어 수를 보면, 1, 2시차까지는 업비트 가격에 별다른 영향을 미치지 못하다가 3시차부터는 1% 유의수준에서 인과관계가 있는 것으로 나타났다. 이는 곧 긍정적인 소셜 감성이 업비트의 비트코인 가격 변화에 영향을 미친다는 것을 뜻한다. 하지만 2시차부터 업비트의 비트코인 가격도 긍정어 수에 1% 유의수준에서 인과관계가 있는 것으로 나타났기 때문에 이는 일방향의 인과관계가 아닌 양방향의 인과관계로 봐야한다.

〈표 12〉 Granger Causality Test(NEG & UBP)

lags	Null Hypothesis	Obs	F	p
1	NEG \nRightarrow UBP	177	4.534	0.035*
	UBP \nRightarrow NEG		2.975	0.086
2	NEG \nRightarrow UBP	175	2.413	0.093
	UBP \nRightarrow NEG		1.895	0.153
3	NEG \nRightarrow UBP	173	4.271	0.006**
	UBP \nRightarrow NEG		2.534	0.059
4	NEG \nRightarrow UBP	171	3.434	0.009**
	UBP \nRightarrow NEG		3.306	0.012*
5	NEG \nRightarrow UBP	169	5.917	0.000***
	UBP \nRightarrow NEG		2.686	0.023*

〈표 12〉는 부정어 수와 업비트 비트코인 가격의 인과관계 분석 결과이다. 예상한 바와 같이 부정어 수가 업비트 가격에 미치는 영향이 1시차부터 전반적으로 유의하였다. 이와 같은 결과는 회귀분석에서 부정어 수가 업비트 가격과 높은 상관관계를 보인 것과 일치하는 결과라 할 수 있다. 반대의 경우인 업비트 비트코인 가격은 3시차까지는 부정어 수에 영향을 미치지 못하다가 4시차와 5시차에서 통계적으로 유의미한 결과를 나타냈다. 결론적으로 초기에는 부정어 수만 업비트의 가격 변동에 영향을 미치는 단방향 인과관계를 보이지만, 점차 시간이 지남에 따라 업비트 가격도 부정어 수에 영향을 주는 양방향 인과관계를 보인다.

VI. 결 론

6.1 연구결과 요약 및 논의

본 연구에서는 사람들의 감성이 암호화폐 가격에 영향을 미치는지 알아보기 위하여 트위터를 중심으로 한 소셜 감성 데이터를 활용해 실증 연구를 진행하였다. 이를 위해 암호화폐 가격에 영향을 미칠 것으로 예상되는 소셜 감성 요인으로 중립어와 긍정어, 부정어를 선정하였다.

먼저 소셜 감성 데이터는 파이썬으로 만든 웹 크롤러를 사용해 트위터에서 비트코인과 관련된 대량의 데이터를 수집하였고, 다음으로 감성 분석을 이용해 중립어 수, 긍정어 수, 부정어 수로 정제하였다. 이때, 보다 신뢰성있는 결과를 도출하기 위해 암호화폐 시장 특성을 고려한 감성 어휘 사전을 직접 제작하여 사용하였다. 그리고 이렇게 추출된 소셜 감성 데이터가 암호화폐 가격에 어떤 영향을 미치는지 확인하기 위해 종속변수로 비트코인 가격을 선정하여 그 관계를 살펴 보았다. 여기서 유념할 점은, 암호화폐 시장은 타 금융시장과는 달리 공인된 마켓이나 통합 지수가 존재하지 않기 때문에 각 거래소마다 가격에 차이를 보인다는 것이다. 이에 본 연구에서는 가장 많은 회원수를 보유하고 있는 빗썸과 업비트의 데이터를 중심으로 연구를 수행하였다. 아울러 소셜 감성과 비트코인 가격 사이에 상관관계뿐 아니라 인과관계가 존재하는지 Granger Causality test를 통해 살펴보았다.

분석 결과, 중립어 수와 긍정어 수는 비트코인 가격과 긍정적인 상관관계로 나타났고, 부정어 수는 비트코인 가격과 부정적인 상관관계로 나타났다. 또한 중립어 수, 긍정어 수, 부정어 수 모두 비트코인 가격 변동에 영향을 미치는 인과관계인 것으로 분석됐다. 다만 인과관계의 방향성이 일방향이 아닌 양방향을 띄고 있으므로, 결론적으로 트위터에서의 중립적 또는 긍·부정적 감성이 암호화폐 가격에 영향을 미치고, 반대의 경우인 암호화폐 가격도 트위터의 중립적, 긍정적, 부정적 감성에 영향을 미친다고 할 수 있겠다.

소셜 감성이 변하면 암호화폐 투자자의 패턴도 변할 수 있음을 확인한 것은 학문적으로 의미가 높은 발견이라 할 수 있다. 가격변화가 감성 변화에 영향을 미치는 것은 당연한 사실일 수 있으므로, 본 연구는 감성변화가 암호화폐 투자에 영향을 미친다는 방향성 관계를 조사하기 위하여 인과관계 분석을 시행하였다. 양방향이기는 하지만, 감성변화가 암호화폐 투자에도 영향을 미치

는 인과관계가 존재하기 때문에, 이는 암호화폐 투자자들은 본질적 가치가 아닌 다른 사람들의 감정 및 사회적 분위기에 이끌려 투자함을 확인하는 것이며, 기존의 소셜 감성이 자산가치에 미치는 영향에 대한 연구를 암호화폐로까지 확장한 것이라 할 수 있다.

6.2 기여도

본 연구를 통해 트위터의 소셜 감성이 비트코인 가격 변동과 높은 상관관계가 있고, 나아가 사람들의 감성이 모여 비트코인 가격에 영향을 미치는 인과관계도 존재하고 있다는 결론을 얻었다. 이는 암호화폐 투자자들이 정확한 정보에 의한 합리적 분석과 판단에 따라 투자하는 것이 아니라 주관적 판단이나 시장의 분위기에 반응하는 경우가 많기 때문으로 보인다. 즉 한국에서의 비트코인 열풍은 펀더멘탈한 데이터에 기반한 것이 아니라 대부분 비이성적인 버블에 의한 현상이었음을 의미한다. 따라서 본 연구는 암호화폐 가격에 대한 소셜 감성의 영향을 확인하여 사회적 현상에 대한 설명의 폭을 늘려간 데 의의가 있다. 이는 특히 한국 암호화폐 시장에 대한 소셜 감성의 영향을 인과관계와 함께 살펴본 최초의 연구로서, 사람들이 향후 비트코인 투자 시 소셜 감성의 영향력을 고려할 수 있도록 할 것이다.

아울러 본 연구는 트위터의 소셜 감성과 암호화폐 가격 간의 관계를 분석하기 위해, 암호화폐 시장에 특화된 감성사전을 구축하고 이를 활용한 데에도 그 의의가 있다. 기존에 소셜 감성을 다룬 많은 연구들은 각 어휘별로 긍정/부정의 극성을 규정해 놓은 범용 감성사전을 사용하였다. 하지만 특정 어휘의 극성은 도메인이나 분석 목적에 따라 다르게 해석될 수 있으므로 보다 정확한 결과를 이끌어 내기 위해서는 해당 도메인에 특화된 감성사전이 필요하다. 특히 암호화폐 시장과 같이 많은 신조어들이 존재하는 경우에는 그러한 사전의 필요성이 더욱 대두된다. 만약 감성사전

에 존재하지 않는 단어들이 있다면, 중요한 의미를 내포하고 있더라도 그 감성을 판별할 수 없기 때문이다. 그리고 현재까지 암호화폐 시장에 특화된 감성사전을 구축한 연구가 굉장히 미비하다는 점에서 본 연구는 이 분야의 지평을 한걸음 더 나아가게 할 것으로 보인다.

마지막으로, 본 연구는 암호화폐 시장을 수용하는 과정 및 향후의 시장 진화 모습을 예측하는 데도 중요한 역할을 할 수 있을 것이다. 현재까지 대부분의 암호화폐와 관련된 연구들은 순수 재무 측면에서의 암호화폐 가격의 변동성에 대한 연구이거나 공학적 측면에서의 기술적 부분인 블록체인(block chain)에 초점이 맞춰져 온 반면 본 연구는 사람들의 소셜 감성을 활용했기 때문이다. 따라서 향후 암호화폐 시장에 대한 정책 혹은 규제 마련 시 이러한 사람들의 소셜 감성과 그에 따른 시장 반응을 고려하면 보다 현실적인 방안을 수립할 수 있을 것이다.

6.3 한계점 및 추후 연구방향

앞서 언급한 기여도에도 불구하고 본 연구는 다음과 같은 한계점을 지니고 있다.

첫째, 사람들의 소셜 감성 데이터를 수집하는데 있어 트위터만을 고려하였다. 암호화폐가 사회적 최대 이슈였던 만큼 트위터뿐 아니라 다른 소셜 미디어와 각종 커뮤니티에서도 무수히 많은 언급이 있었는데, 본 연구에서는 이를 반영하지 못하였다. 게다가 트위터는 암호화폐에 특화된 커뮤니티가 아닌 만큼 실제 투자자가 아닌 사람들이 암호화폐에 대해 논한 비율이 상당할 것으로 보인다. 따라서 추후에는 암호화폐 관련 커뮤니티의 소셜 감성 데이터 수집을 병행하여 실제 투자자들의 언급을 변수에 포함시켜 연구를 진행하면 보다 정확한 결과가 도출될 것이다.

둘째, 트위터에서 발생하는 사람들의 풍부한 어휘를 단지 중립과 긍·부정으로 단편화시켜 반영하였다. 대부분의 사람들은 트위터에 특정 주

제나 이슈에 대해 포스팅할 때 다양한 어휘를 사용하여 본인의 심리를 표현하곤 한다. 하지만 본 연구에서는 암호화폐 도메인에 특화된 감성사전을 제작했음에도 불구하고, 사람들의 소셜 감성을 세분화하지는 못하였다. 향후 연구에서 Bollen et al.(2011)과 같이 감성 카테고리를 ‘Calm’, ‘Alert’, ‘Sure’, ‘Vital’, ‘Kind’, ‘Happy’ 등으로 보다 상세하게 분류하고 분석한다면, 좀 더 명확하게 소셜 감성과 암호화폐 가격간의 관계를 설명할 수 있을 것이다.

셋째, 각 암호화폐마다 가격 변동에 차이가 있음에도 모형에 비트코인만 투입시켰다는 한계가 존재한다. 특히 비트코인캐시의 경우 비트코인과 정반대의 움직임을 보이는 경우도 많았는데, 이러한 경우에 대한 소셜 감성의 영향을 확인하지 못하였다. 이는 데이터를 일별 데이터로 수집 시 비트코인 외 암호화폐들에 대한 트위터의 언급 표본이 충분하지 못했기 때문이다. 따라서 추후 연구에서는 시간 단위의 시계열 데이터를 활용하여 좀 더 세밀하게 분석하는 방안이 고려되어야 할 것이다. 나아가 투자자들로부터 각광을 받고 있는 이더리움, 리플, 라이트코인 등까지 연구대상을 확대하고, 동시에 해외의 데이터도 수집하여 한국의 암호화폐 가격과 어떤 차이가 있는지도 분석한다면 본 연구의 결과가 보다 더 일반화될 수 있을 것이다.

참 고 문 헌

- [1] 김영민, 정석재, 이석준, “소셜 미디어 감성분석을 통한 주가 등락 예측에 관한 연구”, *Entree Journal of Information Technology*, 제13권, 제3호, 2014, pp. 59-69.
- [2] 안정국, 김소담, 김희웅, “텍스트 마이닝 기법을 이용한 정보시스템분야 연구 동향 분석”, *Information Systems Review*, 제18권, 제3호, 2016, pp. 73-96.
- [3] 유은지, 김유신, 김남규, 정승렬, “주가지수 방향성 예측을 위한 주제지향감성사전 구축 방안”, *지능정보연구*, 제19권, 제1호, 2013, pp. 95-110.
- [4] 이희중, 조재영, “암호화폐에 관한 연구”, *비교사법*, 제25권, 제2호, 2018, pp. 657-696.
- [5] 전주용, 여은정, “비트코인의 이해-금융경제학적 관점에서”, *Korea Business Review*, 제18권, 제4호, 2014, pp. 211-239.
- [6] 전희국, 현근수, 임경빈, 이우현, 김형주, “영화 흥행 실적 예측을 위한 빅데이터 전처리”, *정보과학회 컴퓨팅의 실제 논문지*, 제20권, 제12호, 2014, pp. 615-622.
- [7] 최기철, 이상용, “직원을 위한 내부마케팅이 기업의 시가 총액 변동률에 미치는 영향 분석: 잡플래닛 기업 리뷰를 중심으로”, *Information Systems Review*, 제20권, 제2호, 2018, pp. 39-62.
- [8] Aggarwal, R., R. Gopal, A. Gupta, and H. Singh, “Putting money where the mouths are: The relation between venture financing and electronic word-of-mouth”, *Information Systems Research*, Vol.23(3-part-2), 2012a, pp. 976-992.
- [9] Aggarwal, R., R. Gopal, R. Sankaranarayanan, and P. V. Singh, “Blog, blogger, and the firm: Can negative employee posts lead to positive outcomes?”, *Information Systems Research*, Vol.23, No.2, 2012b, pp. 306-322.
- [10] Akerlof, G. A. and R. J. Shiller, *Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism*, Princeton University Press, 2010.
- [11] Bianchetti, M., C. Ricci, and M. Scaringi, *Are cryptocurrencies Real Financial Bubbles? Evidence from Quantitative Analyses*, A version of this paper was published in *Risk*, Vol.26, 2018.
- [12] Bollen, J., H. Mao, and X. Zeng, “Twitter mood predicts the stock market”, *Journal of Computational Science*, Vol.2, No.1, 2011, pp. 1-8.
- [13] Catania, L. and S. Grassi, *Modelling Crypto-*

- Currencies Financial Time-Series*, CEIS Working Paper, No.417, 2017.
- [14] Cheah, E. T. and J. Fry, "Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin", *Economics Letters*, Vol.130, 2015, pp. 32-36.
- [15] Cheung, A., E. Roca, and J. J. Su, "Crypto-currency bubbles: An application of the Phillips-Shi-Yu(2013) methodology on Mt. Gox bitcoin prices", *Applied Economics*, Vol.47, No.23, 2015, pp. 2348-2358.
- [16] Chheda, S. R., A. K. Singh, P. S. Singh, and A. S. Bhole, "Automated trading of cryptocurrency using twitter sentimental analysis", *International Journal of Computer Sciences and Engineering*, Vol.6, 2018, pp. 209-214.
- [17] Chu, J., S. Nadarajah, and S. Chan, "Statistical analysis of the exchange rate of bitcoin", *PloS One*, Vol.10, No.7, 2015, e0133678.
- [18] Colianni, S., S. Rosales, and M. Signorotti, *Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis*, CS229 Project, 2015.
- [19] Dickey, D. A. and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association*, Vol.74, No.366a, 1979, pp. 427-431.
- [20] Gilbert, E. and K. Karahalios, "Widespread worry and the stock market", *ICWSM*, 2010, pp. 59-65.
- [21] Granger, C. W., "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica: Journal of the Econometric Society*, Vol.37, No.3, 1969, pp. 424-438.
- [22] Gu, B., J. Park, and P. Konana, "Research note: The impact of external word-of-mouth sources on retailer sales of high-involvement products", *Information Systems Research*, Vol.23, No.1, 2012, pp. 182-196.
- [23] Hencic, A. and C. Gouriéroux, "Noncausal autoregressive model in application to bitcoin/usd exchange rates", *Econometrics of Risk*, Springer, Cham, 2015, pp. 17-40.
- [24] Kim, H. W., H. C. Chan, and S. Gupta, "Social media for business and society", *Asia Pacific Journal of Information Systems*, Vol.25, No.2, 2015, pp. 329-336.
- [25] Kim, S. M. and E. Hovy, "Extracting opinions, opinion holders, and topics expressed in online news media text", In *Proceedings of the Workshop on Sentiment and Subjectivity in Text, Association for Computational Linguistics*, July 2006, pp. 1-8.
- [26] Kim, T., W. J. Jung, and S. Y. Lee, "The analysis on the relationship between firms' exposures to SNS and stock prices in Korea", *Asia Pacific Journal of Information Systems*, Vol.24, No.2, 2015, pp. 233-253.
- [27] Lee, G., D. Y. Joe, and J. Jeong, "An investigation of dynamic price movements of the cryptocurrency coin in Korea", *한국경영학회 통합학술발표논문집*, 2018, pp. 1476-1488.
- [28] Luo, X., J. Zhang, and W. Duan, "Social media and firm equity value", *Information Systems Research*, Vol.24, No.1, 2013, pp. 146-163.
- [29] Mai, F., Z. Shan, Q. Bai, X. Wang, and R. H. Chiang, "How does social media impact Bitcoin value? A test of the silent majority hypothesis", *Journal of Management Information Systems*, Vol.35, No.1, 2018, pp. 19-52.
- [30] Manning, C. D., P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, Vol.39, 2008.
- [31] Matta, M., I. Lunesu, and M. Marchesi, *Bitcoin Spread Prediction Using Social and Web Search Media*, UMAP Workshops, 2015.
- [32] Oh, C. and O. Sheng, "Investigating predictive

- power of stock micro blog sentiment in forecasting future stock price directional movement”, *Thirty Second International Conference on Information Systems(ICIS 2011)*, Shanghai, China, 2011, pp. 1-19.
- [33] Rao, T. and S. Srivastava, “Analyzing stock market movements using twitter sentiment analysis”, In *Proceedings of the 2012 international conference on advances in social networks analysis and mining(ASONAM 2012)*, *IEEE Computer Society*, 2012, pp. 119-123.
- [34] Rick, S. and G. Loewenstein, “The role of emotion in economic behavior”, *Handbook of Emotions*, Vol.3, 2008, pp. 138-158.
- [35] Sapuric, S. and A. Kokkinaki, “Bitcoin is volatile! isn’t that right?”, *International Conference on Business Information Systems*, Springer, Cham, 2014, pp. 255-265.
- [36] Schumaker, R. P. and H. Chen, “Textual analysis of stock market prediction using breaking financial news: The AZFin text system”, *ACM Transactions on Information Systems(TOIS)*, Vol.27, No.2, 2009, p. 12.
- [37] Weber, B., “Bitcoin and the legitimacy crisis of money”, *Cambridge Journal of Economics*, Vol.40, No.1, 2014, pp. 17-41.
- [38] Yelowitz, A. and M. Wilson, “Characteristics of Bitcoin users: An analysis of Google search data”, *Applied Economics Letters*, Vol.22, No.13, 2015, pp. 1030-1036.
- [39] Zhang, S., C. Zhang, and Q. Yang, “Data preparation for data mining”, *Applied Artificial Intelligence*, Vol.17, No.5-6, 2003, pp. 375-381.
- [40] Zhang, Z., Q. Ye, and R. Law, “Sentiment classification of online reviews to travel destinations by supervised machine learning approaches”, *Expert Systems with Applications*, Vol.36, No.3, 2009, pp. 6527-6535.

An Analysis of Relationship between Social Sentiments and Cryptocurrency Price: An Econometric Analysis with Big Data

Sangyi Ryu^{*} · Jiyeon Hyun^{**} · Sang-Yong Tom Lee^{***}

Abstract

Around the end of 2017, the investment fever for cryptocurrencies—especially Bitcoin—has started all over the world. Especially, South Korea has been at the center of this phenomenon. Since it was difficult to find the profitable investment opportunities, people have started to see the cryptocurrency markets as an alternative investment objects. However, the cryptocurrency fever in South Korea is mostly based on psychological phenomenon due to expectation of short-term profits and social atmosphere rather than intrinsic value of the assets. Therefore, this study aimed to analyze influence of people's social sentiment on price movement of cryptocurrency. The data was collected for 181 days from Nov 1st, 2017 to Apr 30th, 2018, especially focusing on Bitcoin-related post in Twitter along with price of Bitcoin in Bithumb/UPbit. After the collected data was refined into neutral, positive and negative words through sentiment analysis, the refined neutral, positive, and negative words were put into regression model in order to find out the impacts of social sentiments on Bitcoin price. After examining the relationship by the regression analyses and Granger Causality tests, we found that the positive sentiments had a positive relationship with Bitcoin price, while the negative words had a negative relation with it. Also, the causality test results show that there exist two-way causalities between social sentiment and Bitcoin price movement. Therefore, we were able to conclude that the Bitcoin investors' behaviors are affected by the changes of social sentiments.

Keywords: *Bitcoin, Cryptocurrency, Social Sentiment, Sentiment Lexicon, Sentiment Analysis, Opinion Mining, Granger Causality test*

* School of Business, Hanyang University

** School of Business, Hanyang University

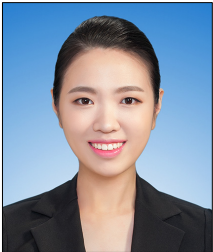
*** Corresponding Author, Professor, School of Business, Hanyang University

◎ 저 자 소 개 ◎



유 상 이 (ryussang2@hanyang.ac.kr)

한양대학교 일반대학원 경영학과에서 MIS(Management Information Systems) 전공으로 석사학위를 취득하였다. 주요 관심분야는 Big Data Analysis, Business Strategy, Digital Business, Social Media, Text Mining 등이다.



현 지 연 (wldus5578@hanyang.ac.kr)

한양대학교 일반대학원 비즈니스인포매틱스학과에서 석사학위를 취득하였다. 주요 관심분야는 Big Data Analysis, Data Mining, Recommendation Systems, Social Media, Text Mining 등이다.



이 상 웅 (tomlee@hanyang.ac.kr)

현재 한양대학교 경영대학 교수로 재직 중이다. 서울대학교 경제학과를 졸업하고, Texas A&M University에서 박사학위를 취득하였다. 주요 관심분야는 정보경제, 개인정보보호(privacy) 및 보안, 소셜미디어, 정보통신정책, 기술경영 등이다. 관련 연구들을 MIS Quarterly, Management Science, Journal of Management Information Systems를 비롯한 다수의 저널에 관련 논문을 게재하고 있다.

논문접수일 : 2018년 11월 15일

게재확정일 : 2019년 01월 29일

1차 수정일 : 2019년 01월 25일