

감성분석을 이용한 뉴스정보와 딥러닝 기반의 암호화폐 수익률 변동 예측을 위한 통합모형¹⁾

An Integrated Model for Predicting Changes in Cryptocurrency Return Based on News Sentiment Analysis and Deep Learning

김은미 (Eunmi Kim) 경희대학교 스마트관광연구소 연구교수²⁾

〈 국문초록 〉

암호화폐 중 대표적인 비트코인은 전 세계적으로 많은 관심을 받고 있으며 비트코인의 가격은 등·하락을 거듭하며 높은 변동성을 보이고 있다. 높은 변동성은 투자자들에게 위험 요인으로 작용하며 무분별한 투자로 인한 사회적 문제를 야기시킨다. 비트코인의 가격은 세계의 환경변화에 영향을 받으며 신속하게 반응하기 때문에 실시간으로 다양한 정보를 제공하는 뉴스 정보는 비트코인 가격의 변동성 예측에 유용한 정보를 제공한다. 즉, 긍정적인 뉴스는 투자심리를 자극할 것이며 반대로 부정적인 뉴스는 투자심리를 위축시킬 것이다. 따라서 본 연구에서는 비트코인의 수익률 변동을 예측하기 위해 뉴스의 감성정보와 딥러닝을 적용하였다. 로짓, 인공신경망, SVM, LSTM을 적용하여 단일 예측모형을 구축하였으며 예측성고를 향상시키기 위한 방법으로 통합모형을 제안하였다. 과거의 가격정보를 기반으로 구축한 예측모형과 뉴스의 감성정보를 반영한 예측모형의 성과를 비교한 결과 뉴스의 감성정보를 반영한 예측모형의 성과가 우수하게 나타났으며 통합모형의 성과가 가장 우수한 것으로 나타났다. 본 연구는 비트코인 수익률 변동에 대한 예측모형을 통해 무분별한 투자를 예방하고 투자자들의 현명한 투자가 이루어질 수 있도록 유용한 정보를 제공할 수 있을 것이다.

주제어: 암호화폐, 감성분석, 딥러닝, LSTM, 통합모형

1) 이 논문 또는 저서는 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019S1A3A2098438)

2) keunmi100@gmail.com

1. 서론

암호화폐 시장이 급성장하면서 대표적인 암호화폐 중 하나인 비트코인이 전 세계적으로 높은 관심을 받고 있다. 국내에서도 짧은 기간 동안 가격이 폭등하며 많은 투자자들이 관심을 보이고 있으나 높은 변동성으로 인해 일부 전문가들은 우려의 목소리를 내고 있다. 2017년 말 비트코인의 가격은 비정상적으로 급등하여 당해 연초와 비교하면 약 20배로 가격이 폭등하였으며 과열된 비트코인의 가격을 정부가 규제하며 2018년부터 가격이 하락하기 시작하였다. 2021년 초부터 다시 급등세를 보이며 특히 최근에는 전기자동차 회사 테슬라에서 비트코인 결제를 허용하며 가상화폐 시장이 폭등하여 4월 초에는 1비트코인 당 8,000만 원을 웃돌았다. 그러나 2달여 만에 테슬라에서 비트코인 결제 허용에 대해 번복하면서 비트코인을 통한 결제 중단을 발표하고 비트코인의 가격은 급락하기 시작했으며 5월에는 5,000만 원 이하로 떨어지고 있다. 비트코인의 가격은 급등과 급락을 반복하며 사회적 이슈가 되고 있으며 비트코인 가격의 높은 변동성은 높은 수익률로 투자심리를 자극하고 있다. 그러나 비트코인 가격의 비정상적인 급등으로 무리하게 투자한 투자자들은 가격의 급락으로 파산을 맞이하기도 하며 다양한 사회적 문제를 야기시킨다.

비트코인은 다른 금융상품보다 높은 변동성을 보이며 높은 변동성 만큼 높은 수익률로 인해 많은 개인 투자자들에게 관심을 받아왔다. 비트코인 거래시장에서도 투자의 고위험 고수익 원칙이 적용되고 있으며 (Thies and Molnár, 2018), 무분별한 투자를 예방하기 위해서도 수익률 변동에 대한 정보제공의 필요성이 대두된다. 비트코인 가격에 영향을 미치는 요인에 대해 많은 연구가 이루어지고 있으며(Hayes, 2018, Kristoufek, 2015), Kristoufek(2015)는 비트코인 가격과

구글 트렌드 지수 사이의 상관성을 확인하였다. 이기광 등(2019)은 한국시장에서 비트코인 가격에 영향을 미치는 요인을 검증하였으며 경제적 변수보다 네이버 트렌드 지수나 신문기사 등과 같이 대중의 관심과 연관된 심리적 요인에 더 강한 영향을 받는다고 하였다. 비트코인과 관련된 검색 트래픽은 대중의 관심정도는 파악할 수 있으나 네이버나 구글 트렌드의 경우 긍정과 부정으로 분류할 수 없다는 한계점을 가지고 있다 (이기광 등, 2019). 또한 비트코인 가격은 세계의 경제 환경변화나 투자환경변화에 영향을 받으며 이러한 변화에 대한 반응속도가 매우 빠르게 나타난다(임병진, 2019). 뉴스는 실시간으로 다양한 정보를 제공하며 긍정적인 뉴스는 가격의 상승을 기대할 수 있으며 투자 심리를 자극할 것이고 반대로 부정적인 뉴스는 수익률에 대한 불확실성으로 투자심리를 위축시킬 것이다. 뉴스의 감성정보는 비트코인의 가격변동에 유용한 정보를 제공하고 있으나 기존연구에서는 비트코인에 대한 검색 트래픽만을 적용하여 가격과의 관련성을 확인하였다.

따라서 본 연구에서는 비트코인의 수익률을 예측하기 위해 뉴스의 감성정보를 활용하여 가격의 상승 또는 하락에 대한 예측모형을 구축한다. 투자자들에게 비트코인의 가격지수보다 가격의 상승과 하락에 따른 수익률 정보가 투자결정에 중요한 요인으로 작용할 것이며 수익률 변동에 대한 정보는 무분별한 투기를 예방할 수 있을 것이다. 비트코인의 가격정보는 시계열 데이터를 기반으로 하지만 계절성이 적어 기계학습을 적용하는 것이 적합하다(Chen et al., 2020). 그러나 시계열 데이터의 특성만 고려하여 통계적인 기법을 적용한 비트코인 가격 예측모형을 구축하여 왔으며(Roy et al., 2018; Garg, 2018; Wirawan et al., 2019) 기계학습을 적용한 연구는 부족하다. 본 연구는 금융 시장 플랫폼인 Investing.com에서 비트코인의 가격정

보와 비트코인 관련 뉴스 정보를 수집하여 로짓, 인공 신경망, SVM(Support Vector Macines), LSTM(Long Short-Term Memory)을 적용하여 예측모형을 구축하며 과거의 가격정보를 기반으로 한 예측모형과 뉴스의 감성정보를 반영한 예측모형의 예측성과를 비교한다. 또한 예측모형은 데이터의 특성에 따라 예측성과가 상이하게 나타나기 때문에 다양한 예측모형을 적용하여 예측모형별 예측성과를 단순히 비교하고 있다. 이는 적합한 모형선정에 대한 문제가 존재하며 본 연구에서는 모형선정에 대한 문제를 해결하고 예측성과를 향상시키기 위해 사례기반추론을 이용한 통합모형을 제안하고자 한다.

2. 이론적 배경

2.1. 암호화폐 관련연구

암호화폐(Cryptocurrency)는 블록체인을 기반으로 분산환경에서 암호화 기술을 사용하여 만든 디지털 자산(digital asset)의 일종으로, 암호화 방법을 통해 거

래의 안전성을 높이고, 화폐 발행(unit)을 통제하며, 자산의 거래나 이동을 인증할 수 있다(Bakar et al., 2017). 대표적인 암호화폐인 비트코인(Bitcoin)은 2009년 사토시 나카모토에 의해 개발되었으며 세계최초의 온라인 암호화폐이다(Nakamoto, 2008). 비트코인은 암호화폐 중 시가총액이 가장 높으며 2017년 비정상적인 가격급등과 최근 테슬라가 비트코인을 결제수단으로 받아들일것다는 발표를 하면서 급등과 급락을 반복하며 전세계적으로 관심으로 받고 있다. 비트코인의 가격 변동은 <그림 1>과 같이 짧은 기간에 발생하고 있다. 2017년 비정상적인 급등은 무분별한 투자와 시장의 혼란을 야기하였으며 비트코인 가격에 대한 정부의 규제가 시작되면서 급락하였다. 2021년 초 다시 급등세를 보이며 2021년 1월 4만 달러를 기록하였으며 불과 3달 후인 4월에는 6만 달러 이상이 되었고 한 달여 만에 4만 달러 이하로 급락하며 변동성이 다른 금융 상품보다도 크게 나타나고 있다.

비트코인에 대한 관심이 고조되면서 비트코인에 대한 연구도 활발하게 이루어지고 있다. 비트코인에 대한 소비자의 사용의도 및 수용에 대한 연구(Shin and Kim, 2016), 비트코인의 가격변화에 영향을 미치는 요

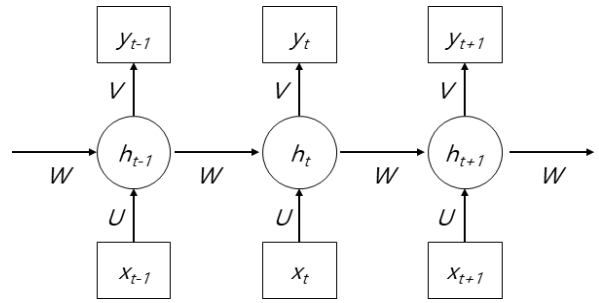


<그림 1> 비트코인 가격(종가) 데이터

인에 대한 연구(이준식 등, 2018; 이기광 등, 2019), 비트코인의 가격예측을 위한 연구(Ji et al., 2018; Chen et al., 2020) 등이 이루어지고 있다. 비트코인 가격예측 연구에는 예측성과를 향상시키기 위해 다양한 모형을 적용하여 예측성과를 비교하거나 뉴스나 검색 트래픽을 반영하여 예측성과를 향상시키고자 하였다. Jang and Lee(2018)은 인공지능망이 비트코인 가격을 우수하게 예측하는 것을 확인하였으며, Mallqui and Fernandes(2019)는 비트코인 가격 예측에서 SVM의 우수성을 확인하였다. Mittal et al.(2019)은 비트코인의 가격 예측을 위해 구글 트렌드, 트윗의 감성과 볼륨을 사용하였으며 트윗의 감성값은 예측성과를 향상시키지 못하는 것으로 나타났다. Patel et al.(2020)은 암호화폐의 가격 예측을 위해 딥러닝 기법인 GRU와 LSTM의 하이브리드 모형을 제안하였으며 우수한 성과를 확인하였다. 따라서 본 연구에서는 뉴스의 감성정보와 통합모형을 통해 예측성과를 향상시키고자한다.

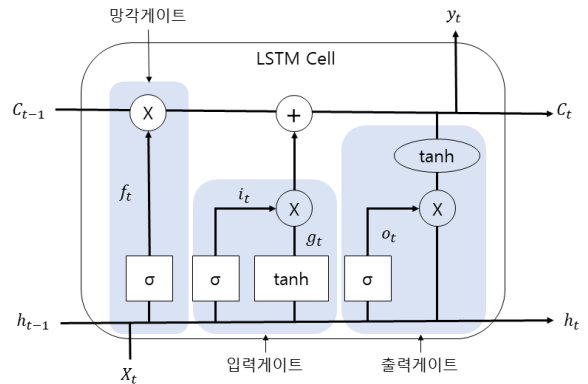
2.2. LSTM

LSTM 네트워크는 딥러닝 기법인 RNN(Recurrent Neural Network)의 한 형태로 Hochreiter and Schmidhuber (1997)에 의해 소개되었다. LSTM은 전통적인 RNN 구조에서 은닉계층의 유닛들을 LSTM 블록(block)으로 대체시킨 형태와 같다. RNN은 과거와 현재의 입력정보를 연결하는 구조로 이루어져 있으며 <그림 2>과 같이 입력층(x_{t-1}, x_t, x_{t+1}), 은닉층(h_{t-1}, h_t, h_{t+1}), 출력층(y_{t-1}, y_t, y_{t+1})으로 구성된다. 입력층과 은닉층을 연결하는 가중치 U, 은닉층과 출력층을 연결하는 가중치 V, 은닉층과 각 노드들을 연결하는 가중치 W로 구성되어 있다.



<그림 2> RNN의 구조

LSTM은 과거 학습결과를 전달하는 구조로 RNN의 장기존속에 대한 문제를 해결한다. <그림 3>과 같이 기존의 은닉 유닛들과 동일하게 재귀적 구조를 가지며 메모리 셀(memory cell)과 입력 게이트(input gate), 망각 게이트(forget gate), 출력 게이트(output gate)의 3종류의 게이트 유닛으로 이루어져 있다. LSTM은 출력 신호의 주요 정보를 유지하거나, 중요하지 않은 정보를 무시하는 능력을 가지며 메모리 셀은 장기상태에 대한 의존상태를 유지할 것인지 결정한다.



<그림 3> LSTM의 구조

LSTM은 과거정보에 대한 이해를 바탕으로 학습이 이루어지기 때문에 시계열 데이터에 적합한 것으로 알려져 있으며 독립적인 입력변수를 사용하는 기존의 신경망보다 정확한 예측이 가능하여 수요예측을 위해 다양한 분야에 활용되고 있다. 신동하 등(2017)은 주

가예측을 위해 딥러닝 기법을 제안하였으며 시계열 딥러닝 네트워크인 RNN과 LSTM이 다른 기법보다 안정적인 학습을 하며 약 15% 정도의 향상된 예측성과를 보였다. 송한진 등(2019)는 증권시장의 변동성을 예측하기 위해 LSTM을 적용하였으며 변동성 예측에 LSTM 기법이 유용하다는 것을 검증하였다. McNally et al.(2018)은 비트코인의 가격예측을 위해 LSTM을 적용하였으며 우수한 예측성과를 보였다.

2.3. 통합모형

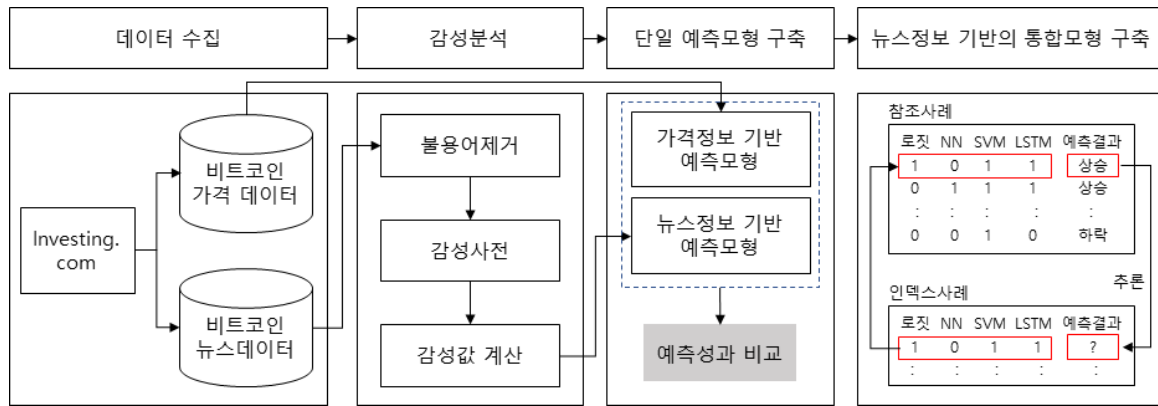
비트코인 가격의 변동성을 예측하기 위해 다양한 기법을 적용하여 예측모형을 구축하였다. 초기에는 통계적 기법을 적용한 예측모형이 많았으나 예측모형의 성과를 향상시키기위해 최근에는 인공지능 기법과 딥러닝 기법이 많이 적용된다. Roy et al.(2018)은 비트코인의 가격을 예측하기 위해 ARIMA를 적용하였으며 AR모형과 MA모형과의 예측성과를 비교하였다. Wirawan et al.(2019)는 ARIMA를 통해 1일과 7일 뒤의 가격예측을 통해 예측기간에 따른 성과를 비교하였다. Chen et al.(2020)은 통계적인 기법과 기계학습 기법으로 예측모형을 구축하고 성과를 비교한 결과 기계학습을 적용한 모형에서 우수한 성과를 확인하였으며 계절성을 보이지 않는 비트코인 가격의 변동성 예측에는 기계학습이 보다 적합하다고 하였다. Lamothe-Fernández et al.(2020)은 예측모형의 성과를 향상시키기 위해 딥러닝 기반의 DNN, DT, SVR을 적용하여 예측성과를 비교하였다. 기존의 비트코인 가격 예측모형은 다양한 기법들을 적용하여 예측성과를 비교하거나 예측모형의 성과를 향상시키기 위해 딥러닝 기법을 적용한 연구가 주를 이루고 있다. 그러나 예측모형과 데이터의 특성에 따라 예측성과가 상이하게 나타나기 때문에 모형선정에 대한 문제가 존재한다. 두 개이상의 모형을 통합한

통합모형은 모형선정에 대한 문제를 해결하는 동시에 예측성과를 향상시킬 수 있다(홍태호, 박지영, 2009).

통합모형은 예측모형의 성과향상을 위해 다양한 분야에서 적용되고 있으며 김명중(2009)는 정량추론과 정성추론을 통합하여 추가예측에 적용하였다. Versace et al.(2004)는 금융예측 모델링을 위해 인공신경망과 유전자알고리즘을 통합하였으며 선택수와 홍태호(2004)는 신용등급 예측 분야에서 인공신경망과 로짓모형을 통합하였다. 비트코인 가격 예측 연구에서도 Patel et al.(2020)이 딥러닝 기법인 GRU와 LSTM을 통합하여 예측성과의 향상을 검증하였다. 본 연구에서도 예측성과의 향상과 모형선정에 대한 문제를 해결하기 위해 통합모형을 제안하며 사례기반추론(Case-Based Reasoning; CBR)을 활용한다. 사례기반추론은 기존의 사례를 참조하여 새로운 사례에 가장 비슷한 사례를 통해 문제를 해결하기 때문에 서로 다른 의견을 가진 사용자들의 추천 결과를 통합하는데 유용하다. 홍태호와 박지영(2009)은 고객반응 예측모형을 구축하기 위해 로짓, 인공신경망, SVM을 적용하여 단일모형을 구축하였으며 사례기반추론으로 단일모형을 통합하였다. 홍태호와 김은미(2010)의 연구에서도 단일모형을 구축하고 이를 사례기반추론으로 통합하여 향상된 예측성과를 검증하였다.

3. 연구 프레임워크

본 연구에서는 비트코인 가격의 수익률 변동에 대한 예측을 위해 뉴스정보와 딥러닝을 적용한 통합모형을 제안한다. 금융시장 플랫폼인 Investing.com은 최고의 국제 금융 사이트 중 하나로 평가받고 있으며 전 세계 거래소 250곳에 대한 실시간 데이터 및 최신뉴스를 44개의 언어로 제공하고 있다. 데이터 수집을 위



〈그림 4〉 연구 프레임워크

해 Investing.com에서 비트코인의 과거 가격 데이터와 비트코인에 대한 뉴스의 헤드라인 정보를 python을 활용하여 크롤링한다. 비트코인의 가격은 세계의 경제 환경이나 투자환경의 변화에 신속하게 반응하기 때문에 외부상황에 대한 정보는 비트코인 관련 뉴스 정보를 활용하며 수집된 뉴스는 감성분석을 통해 감성값을 계산한다. 긍정적인 감정의 뉴스는 투자자들의 긍정적인 관심과 함께 가격상승으로 이어지는 반면 부정적인 뉴스는 수익에 대한 불안감으로 인해 거래를 망설이게 된다. 또한 뉴스는 다양한 매체에서 제공되고 있기 때문에 작성된 뉴스의 수, 뉴스의 감성분석 값을 일별로 계산하여 적용한다. 다음으로 과거의 가격 데이터에 의존한 가격정보 기반의 예측모형과 감성분석을 통해 뉴스의 정보를 반영한 뉴스정보 기반의 예측모형을 구축한다. 투자자들이 비트코인 투자에 대한 의사결정 시 가격보다 수익률에 보다 높은 관심을 가질 것으로 판단하여 예측모형은 상승과 하락에 대한 모형으로 구축한다. 예측모형은 로짓, 인공신경망, SVM, LSTM을 적용하여 가격기반 예측모형으로 P_Logit, P_ANN, P_SVM, P_LSTM을 구축한다. 뉴스정보 기반의 예측모형은 N_Logit, N_ANN, N_SVM, N_LSTM으로 과거의 비트코인 수익률과 일별로 계산된 뉴스의 감성값과 긍정뉴스의 비율을 입력변수로

하여 구축하며 예측모형의 성과를 향상시키기 위해 본 연구에서는 사례기반추론(cased-based reasoning, CBR)에 의한 통합모형을 제안한다. 사례기반추론에 의한 통합모형은 뉴스정보 기반으로 구축된 단일모형에 대한 예측값의 유사도를 기반으로 최종 예측값을 결정하게 된다. 단일모형에 의한 예측값은 사례집합에 저장되며 새로운 값의 특성에 대한 유사도를 측정하여 사례집합에서 가장 가까운 사례를 찾는다. 유사도 측정은 유클리드 거리에 의해 계산하며 가장 가까운 거리에 있는 사례를 추출하여 결정한다. 본 연구의 연구 프레임워크는 <그림 4>와 같다.

4. 실험 및 결과분석

4.1. 데이터

본 연구에서는 암호화폐 중 시가총액이 가장 높은 비트코인을 대상으로 하였으며 금융 플랫폼인 Investing.com에서 제공하는 비트코인 가격정보와 비트코인 관련 뉴스를 활용하였다. 비트코인에 대한 대중의 관심은 2017년 하반기 비정상적으로 가격이 급등하면서 커지기 시작했으며 비트코인에 대한 뉴스기

사도 2017년 하반기부터 많아지기 시작하였다. 비정상적인 가격급등에 따라 투기과열과 시장의 혼란이 야기되었으며 시장의 혼란을 해결하기 위한 정부의 규제가 시작되면서 가격이 급락하기 시작하였다. 본 연구에서는 비정상적으로 급등하였던 2017년을 제외하고 2018년 1월 1일부터 2021년 4월 30일까지의 일별 가격 데이터 1,216개와 2017년 12월 31일부터 2021년 4월 29일까지의 비트코인 뉴스의 헤드라인 정보 8,227개를 수집하여 일별로 적용하였다. 2018년, 2019년, 2020년의 데이터는 학습용으로 사용하고 2021년 1월 1일부터 2021년 4월 30일까지는 검증용으로 사용하였다. 가격정보는 1일 전, 2일 전, 3일 전의 정보를 적용하였으며 뉴스정보는 실시간으로 반영되는 정보로 1일 전의 정보를 반영하였다. 비트코인의 수익률의 변동은 증가를 기준으로 전날보다 상승하였으면 1, 하락하였으면 0으로 하였으며 예측모형에 사용된 입력변수는 <표 1>과 같다.

4.2. 가격정보 기반의 단일 예측모형 구축

가격정보를 기반으로 예측모형을 구축하기 위해 1개월 전, 2개월 전, 3개월 전의 가격정보를 입력변수로 하였다. 즉, t시점을 예측하기 위해 t-1, t-2, t-3의 변동수익률을 입력변수로 사용하였다. 수익률 변동 데이터는 시간의 흐름에 따른 시계열 데이터이기 때문

에 학습용은 2018년부터 2020년까지로 하였으며 검증용은 2021년 데이터로 분류하고 python으로 로짓모형(P_Logit), 인공신경망 모형(P_ANN), SVM 모형(P_SVM), LSTM 모형(P_LSTM)을 구축하였다. 인공신경망의 은닉노드 수는 1부터 20까지 변화시켜가며 우수한 모형을 선정하였으며 SVM은 RBF 커널함수를 사용하여 SVM 모형의 모수인 C와 σ 는 그리드 탐색 기법을 적용하여 최적의 값을 찾아 적용하였다. LSTM은 시계열 데이터에 널리 사용되는 딥러닝 기법으로 최적화를 위해AdamOptimizer알고리즘을 사용하였으며 Loss함수는 binary_crossentropy, 활성화 함수는 ReLU와 Tanh를 이용하였다.

4.3. 뉴스정보 기반의 단일 예측모형 구축

예측모형에 뉴스의 정보를 반영하기 위해 뉴스의 헤드라인 정보에 대한 감성분석(sentiment analysis)을 수행하였다. 감성분석은 텍스트에 내재되어 있는 감성을 파악하여 작성자의 감성을 추출하는 기법으로 긍정인지 부정인지 분류할 수 있다(Liu, 2010). 뉴스정보는 주식시장의 투자 의사결정에 영향을 미치는 중요한 요소 중 하나이며(강현모, 2017; Schumaker et al., 2009; Bollen et al., 2011), 비트코인은 주식과 유사한 성격을 가진다고 보기도 한다(Jun and Yeo, 2014). 임병진(2019)는 종합주가지수와 비트코인 가격이 높은

<표 1> 입력변수의 기술통계량

변수명	설명	평균	표준편차
X1	t-1 의 비트코인 가격 변동률	0.0019	0.0401
X2	t-2 의 비트코인 가격 변동률	0.0020	0.0402
X3	t-3 의 비트코인 가격 변동률	0.0020	0.0402
X4	전날 뉴스의 극성값 합계	0.6455	7.4710
X5	전날 뉴스의 감성값(부정=-1, 중립=0, 긍정=1)	0.0857	0.9015
X6	전날 긍정뉴스의 비율	0.3327	0.2087
X7	전날 부정뉴스의 비율	0.3151	0.2091

상관관계를 가지고 있음을 검증하였다. 따라서 긍정적인 뉴스는 투자심리를 자극할 수 있으며 본 연구에서는 뉴스의 감성정보를 반영하여 비트코인의 가격변화를 예측하고자 한다.

감성분석을 위해 어휘기반(lexicon-based) 접근방법을 적용하였으며 SentiWordNet 감성사전을 적용하였다. 감성사전은 부정적인 단어와 긍정적인 단어를 분류하고 각 단어에 대한 극성값을 제시하고 있으며 극성값이 0보다 크면 긍정, 0보다 작으면 부정, 0이면 중립으로 분류한다. 본 연구에서는 뉴스에 대한 감성분석을 위해 뉴스의 극성값을 기준으로 일별 극성값을 계산하였으며, 각 뉴스에 대해 긍정뉴스 혹은 부정뉴스로 분류하고 전체 뉴스 중 긍정뉴스의 비율과 부정뉴스의 비율을 일별로 계산하였다. 뉴스는 실시간으로 제공되는 정보이기에 1일의 시차를 두고 입력변수로 사용하였으며 1일 전, 2일 전, 3일 전의 가격변동률도 함께 적용하여 예측모형을 구축하였다. 뉴스정보기반의 예측모형은 로짓모형(N_Logit), 인공신경망모형(N_ANN), SVM 모형(N_SVM), LSTM 모형(N_LSTM)을 python으로 구축하였으며 가격정보기반의 예측모형과 동일하게 실험을 설계하였다.

4.4. 사례기반추론 기반의 통합모형

단일 예측모형의 성과를 향상시키기 위해 본 연구

에서는 사례기반추론을 적용한 통합모형을 제안하였다. 통합모형을 구축하기 위해 단일모형으로 구축한 로짓, 인공신경망, SVM, LSTM의 예측결과를 이용한다. 예측모형에 따라 예측결과는 상이하게 나타나기도 하며 예측성과에도 큰 차이를 보이기 때문에 데이터의 특성에 맞는 모형의 선정은 중요하다. 통합모형은 다양한 모형에 대한 예측결과를 하나의 사례집합으로 저장하여 적용하기 때문에 모형선정의 어려움을 해결할 수 있으며 예측성과를 향상시킬 수 있다.

사례기반추론은 사례집합에 저장되어 있는 특성과 새로운 값에 대한 특성에 대한 유사도를 측정하여 가장 유사한 사례를 찾아 결과를 제시하고 이는 다시 새로운 사례로 사례집합에 저장되어 다음 사례 매칭에 이용할 수 있도록 갱신된다. 유사도 측정은 k-최근접 이웃법(nearest-neighbor method)을 사용하였으며 유클리드 거리(Euclidean distance)를 사용하여 특성들 간의 유사성을 측정하였다. 사례 간의 거리가 가장 짧은 참조 셋을 첫 번째 근접이웃으로 다음은 두 번째 근접이웃으로 하여 최적의 k를 찾기 위해 k를 1부터 10까지의 범위에서 탐색하여 최적의 이웃 수를 정하였다.

4.5. 실험결과

본 연구에서는 비트코인의 수익률 변동을 예측하기 위해 단일모형으로 로짓모형, 인공신경망 모형, SVM

〈표 2〉 가격정보 기반의 예측모형 성과

P_Logit		P_ANN		P_SVM		P_LSTM		P_CBR
학습용	검증용	학습용	검증용	학습용	검증용	학습용	검증용	검증용
53.98%	49.17%	55.35%	49.17%	55.26%	50.00%	53.52%	51.67%	51.67%

〈표 3〉 뉴스정보 기반의 예측모형 성과

N_Logit		N_ANN		N_SVM		N_LSTM		N_CBR
학습용	검증용	학습용	검증용	학습용	검증용	학습용	검증용	검증용
55.08%	53.33%	55.90%	52.50%	54.35%	54.17%	54.71%	55.83%	56.67%

모형, LSTM 모형을 구축하였으며 예측성과를 향상시키기 위한 방안으로 사례기반추론을 적용하여 통합모형을 제시하였다. 가격정보를 기반으로 구축한 예측모형의 성과는 <표 2>와 같다. 시계열 데이터에서 널리 사용되고 있는 P_LSTM의 예측성과가 51.67%로 나타났으며 P_CBR의 예측성과도 P_LSTM과 동일하게 우수한 것으로 나타났다.

<표 3>은 감성분석을 적용한 뉴스정보를 반영한 예측모형의 성과이다. 단일 예측모형에서 N_LSTM이 55.83%로 가장 우수한 것으로 나타났으며 단일모형을 통합한 N_CBR은 56.67%로 N_LSTM 보다 우수한 성과를 보였다. 또한 <표 2>의 가격정보 기반의 예측모형보다 뉴스정보를 활용한 예측모형에서 예측성과가 모두 향상된 것을 확인할 수 있으며 비트코인 가격의 등·하락 예측을 위해 뉴스의 감성분석이 유용한 정보를 제공하고 있음을 확인할 수 있다. 통합모형은 단일모형보다 향상된 예측성과를 보이며 단일 예측모형에서는 최소 3%이상의 향상된 성과를 확인할 수 있으며 통합모형에서는 뉴스정보를 활용했을 때 예측성과가 5%이상 향상되었다.

5. 결론 및 향후 연구과제

본 연구에서는 비트코인의 가격을 예측하기 위해 뉴스정보와 딥러닝을 적용한 통합모형을 제안하였다. 비트코인의 가격은 높은 변동성을 가지며 세계의 환경변화와 뉴스정보와 같은 심리적인 요인에 더 강한 영향을 받기 때문에(이기광 등, 2019) 실시간으로 다양한 정보를 제공하는 뉴스의 감성정보를 예측모형에 반영하였다. 비트코인은 계절성을 보이지 않기 때문에 기계학습을 적용한 예측모형이 보다 적합하며 (Chen et al., 2020) 로짓, 인공신경망, SVM, LSTM으로

예측모형을 구축하였다. 과거의 가격정보를 기반으로 구축한 예측모형과 뉴스의 감성정보를 반영한 예측모형을 구축하여 예측성과를 비교한 결과 뉴스의 감성정보를 반영한 예측모형의 성과가 우수하게 나타났으며 통합모형의 성과가 56.67%로 가장 우수한 것으로 나타났다.

본 연구의 이론적 시사점은 다음과 같다. 첫째, 비트코인 가격의 수익률 변동을 예측하기 위해 딥러닝 기법을 적용하였다. 비트코인 투자에 대한 의사결정 시 비트코인의 가격보다는 오를 것인지 내릴 것인지에 대한 수익률에 대해 보다 높은 관심을 가질 것이기에 비트코인 가격의 상승과 하락에 대한 예측모형을 구축하였다. 시계열 데이터의 예측에 많이 활용되고 있는 딥러닝 기법인 LSTM을 적용하여 우수한 예측성과를 확인할 수 있었으며 향후 비트코인 가격에 대한 방향성에 대한 정보를 제공할 수 있다. 둘째, 단일 예측모형을 기반으로 통합모형을 제안하여 예측성과를 향상시키고자 하였다. 본 연구에서는 사례기반추론을 기반으로 통합모형을 제안하여 단일모형보다 우수한 예측성과를 확인할 수 있었으며 모형선정에 대한 어려움도 해결하였다. 셋째, 뉴스의 감성정보를 활용하여 정량적인 정보와 정성적인 정보를 모두 예측모형에 적용하였다. 최형규와 이상용(2020)의 연구에서도 정량적 정보와 정성적 정보를 모두 활용한 예측모형에서 예측력이 증가한다고 하였으며 비정형화된 뉴스정보를 예측모형에 반영하여 예측모형의 성과를 향상시켰다.

본 연구의 실무적 시사점은 다음과 같다. 첫째, 뉴스의 정보를 모형에 반영하여 급변하는 환경을 즉각적으로 반영할 수 있도록 하였다. 비트코인의 가격은 세계의 환경변화에 민감하며 뉴스는 환경변화에 대한 정보를 실시간으로 파악할 수 있도록 한다. 급변하는 환경에 따라 비트코인의 가격 방향성을 사전에 알 수

있다면 투자자들의 무분별한 투자를 예방하는데 도움을 줄 수 있을 것이다. 둘째, 수익률 변동에 대한 예측으로 안전한 투자환경을 제공할 수 있다. 비트코인은 가격 변동성이 높은 만큼 높은 수익률을 보이지만 파산에 대한 위험부담도 높기에 대부분의 사람들은 투자를 망설이게 된다. 특히 투자 경험이 없는 초보 투자자들은 비트코인에 대한 지식도 부족하며 수익률에 대한 불확실성으로 투자의 기회를 가지지 못한다. 수익률 변동에 대한 예측모형은 전문적인 지식이 없더라도 수익률에 대한 정보를 제공하기 때문에 보다 안전한 환경에서 다양한 투자자들에게 투자 기회를 제공할 수 있다. 셋째, 무분별한 투자로 인해 야기되는 사회적 문제를 예방할 수 있다. 비트코인의 비정상적인 급등으로 무리하게 투자하는 투자자들에게 향후 가격의 방향성에 대한 정보를 제공하여 의사결정 시 참고할 수 있도록 한다면 무분별한 투자로 인해 야기되는 사회적 문제를 줄이는데 조금이나마 기여할 수 있을 것이다.

본 연구의 한계점 및 향후 연구방향은 다음과 같다. t시점의 가격을 예측하기 위해 t-1, t-2, t-3의 정보를 적용하였으나 향후 연구에서는 반영되는 과거의 가격 데이터의 기간에 따른 예측성과를 비교해 볼 수 있으며 예측모형의 구축에 보다 많은 정보를 활용할 수 있을 것이다. 또한 뉴스의 헤드라인 정보에 대한 감성정보를 반영하였으나 뉴스 본문의 내용을 예측모형에 반영하여 보다 풍부한 정보를 활용하며 뉴스를 제공하는 제공자에 따라 가중치를 다르게 하여 예측모형에 적용해 볼 수 있을 것이다. 비트코인 수익률 변동에 대한 예측모형을 통해 무분별한 투자를 예방하고 투자자들의 현명한 투자가 이루어질 수 있도록 유용한 정보를 제공할 수 있을 것이다.

〈참고문헌〉

[국내 문헌]

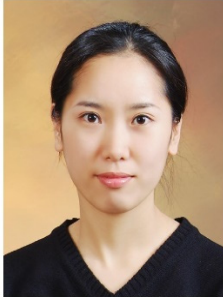
1. 강현모(2017). 투자자의 손실회피 성향과 해석수준이 금융상품 태도에 미치는 영향. *지식경영연구*, 18(1), 49-65.
2. 김명중 (2009). 정량 추론과 정성 추론의 통합 메카니즘: 추가 예측의 적용. *지식경영연구*, 10(2), 35-48.
3. 송한진, 최홍식, 김선웅, 오수훈 (2019). AI의 LSTM기법을 이용한 금융시계열 데이터 변동성 예측방법 연구. *한국지식정보기술학회 논문지*, 14(6), 665-673.
4. 신동하, 최광호, 김창복 (2017). RNN과 LSTM을 이용한 추가 예측을 향상을 위한 딥러닝 모델. *한국정보기술학회논문지*, 15(10), 9-16.
5. 신태수, 홍태호 (2004). 인공지능경망과 로짓모형을 통합한 부실 확률맵기반 신용등급화에 관한 연구. *회계저널*, 13(3), 1-26.
6. 이기광, 조수지, 민경수, 양철원 (2019). 비트코인 가격의 결정 요인: 한국시장에 대한 실증분석. *한국증권학회지*, 48(4), 393-415.
7. 이준식, 김건우, 박도형 (2018). 비트코인 가격 변화에 관한 실증분석: 소비자, 산업, 그리고 거시변수를 중심으로. *지능정보연구*, 24(2), 195-220.
8. 임병진 (2019). 비트코인의 가격변화가 한국 국고채 시장에 미친 상호 영향에 관한 실증적 연구. *e-비즈니스연구*, 20(5), 143-154.
9. 최형규, 이상용 (2020). 재무제표 주석의 텍스트 분석 통한 재무 비율 예측 향상 연구. *지식경영연구*, 21(2), 177-196.
10. 홍태호, 김은미 (2010). 데이터마이닝을 이용한 세분화된 고객집단의 프로모션 고객반응 예측. *Information Systems Review*, 12(2), 75-88.
11. 홍태호, 박지영 (2009). 사례기반추론을 이용한 다이렉트 마케팅의 고객반응예측모형의 통합. *정보시스템연구*, 18(3), 375-399.

[국외 문헌]

12. Bakar, N. A., Rosbi, S., & Uzaki, K. (2017). Cryptocurrency Framework Diagnostics from Islamic Finance Perspective: A New Insight of Bitcoin System Transaction. *International Journal of Management Science and Business Administration*, 4(1), 19-28.
13. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
14. Chen, Z., Li, C., & Sun, W. (2020) Bitcoin price prediction using machine learning: An approach to sample dimension engineering, *Journal of Computation and Applied Mathematics*, 365, 112395.
15. Garg, S. (2018). Autoregressive integrated moving average model based prediction of bitcoin close price. *In 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. 473-478.
16. Hayes, A. S. (2018). Bitcoin price and its marginal cost of production: support for a fundamental value, *Applied Economics Letters*, 26(7), 554-560.
17. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
18. Jang, H., & Lee, J.(2018). An empirical study on modeling and prediction of Bitcoin prices with bayesian neural networks based on Blockchain information, *IEEE Access*, 6(1), 5427-5437.
19. Ji, S., Kim, J., & Im, H. (2019). A comparative study of Bitcoin price prediction using deep learning. *Mathematics*, 7(10), 898.
20. Jun, J., & Yeo, E. (2014). Understanding Bitcoin: From the Perspective of Monetary Economics, *Korea Business Review*, 18(4), 21-239.
21. Kristoufek, L. (2015) What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE*, 10(4), 1-15.
22. Lamothe-Fernández, P., Alaminos, D., Lamothe-López, P., & Fernández-Gámez, M. A. (2020). Deep Learning Methods for Modeling Bitcoin Price. *Mathematics*, 8(8), 1245.
23. Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of Natural Language Processing*, N. Indurkha and F. J. Damerau, eds.
24. Mallqui, D. C., & Fernandes, R. A. (2019). Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning

- techniques. *Applied Soft Computing*, 75(1), 596–606
25. McNally, S., Roche, J., & Caton, S. (2018). Predicting the price of bitcoin using machine learning. *In 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing*. IEEE, 339–343.
26. Mittal, A., Dhiman, V., Singh, A., & Prakash, C. (2019). Short-term bitcoin price fluctuation prediction using social media and web search data. *2019 Twelfth International Conference on Contemporary Computing (IC3)*. IEEE 1–6.
27. Nakamoto, S., & Bitcoin, A. (2008). A peer-to-peer electronic cash system. *Bitcoin*. URL: <https://bitcoin.org/bitcoin.pdf>.
28. Patel, M. M., Tanwar, S., Gupta, R., & Kumar, N. (2020). A deep learning-based cryptocurrency price prediction scheme for financial institutions. *Journal of Information Security and Applications*, 55, 102583.
29. Roy, S., Nanjiba, S., & Chakrabarty, A. (2018). Bitcoin price forecasting using time series analysis. *In 2018 21st International Conference of Computer and Information Technology (ICCIIT)*. IEEE, 1–5.
30. Schumaker, R. P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. *Information Processing & Management*, 45(5), 571–583.
31. Shin, D. H., & Kim, Y. M. (2016). The Factors Influencing Intention to Use Bit Coin of Domestic Consumers. *The Journal of the Korea Contents Association*, 16(1), 24–41.
32. Thies, S., & Molnár, P. (2018). Bayesian change point analysis of Bitcoin returns. *Finance Research Letters*, 27, 223–227
33. Versace, M., Bhatt, R., Hinds, O., & Shiffer, M. (2004). Predicting the exchange traded fund DIA with a combination of genetic algorithm and neural networks. *Expert Systems with Applications*, 27(3), 417–425.
34. Wirawan, I. M., Widiyaningtyas, T., & Hasan, M. M. (2019). Short term prediction on bitcoin price using arima method. *In 2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*. IEEE, 260–5.
- [URL]
35. <http://www.investing.com/>

● **저 자 소 개** ●



김 은 미 (Eunmi Kim)

현재 경희대학교 스마트관광 연구소의 학술 연구교수로 재직 중이다. 부산대학교 경영학과에서 경영학 박사 학위를 취득하였다. 주요 관심분야는 지식경영, 비즈니스 애널리틱스, 딥러닝, 소셜미디어, CRM 등이다. 지금까지 Expert Systems with Applications, 정보시스템연구, 인터넷전자상거래연구 등 주요 학술지에 논문을 발표하였다.

〈 Abstract 〉

An Integrated Model for Predicting Changes in Cryptocurrency Return Based on News Sentiment Analysis and Deep Learning

Eunmi Kim*

Bitcoin, a representative cryptocurrency, is receiving a lot of attention around the world, and the price of Bitcoin shows high volatility. High volatility is a risk factor for investors and causes social problems caused by reckless investment. Since the price of Bitcoin responds quickly to changes in the world environment, we propose to predict the price volatility of Bitcoin by utilizing news information that provides a variety of information in real-time. In other words, positive news stimulates investor sentiment and negative news weakens investor sentiment. Therefore, in this study, sentiment information of news and deep learning were applied to predict the change in Bitcoin yield. A single predictive model of logit, artificial neural network, SVM, and LSTM was built, and an integrated model was proposed as a method to improve predictive performance. As a result of comparing the performance of the prediction model built on the historical price information and the prediction model reflecting the sentiment information of the news, it was found that the integrated model based on the sentiment information of the news was the best. This study will be able to prevent reckless investment and provide useful information to investors to make wise investments through a predictive model.

Key Words: Cryptocurrency, Sentiment Analysis, Deep Learning, LSTM, Integrated Model

* Kyung Hee University Smart Tourism Research Center