

뉴스와 비트코인 가격변동 간의 상관관계에 관한 연구

오동혁, 박상원
한국외국어대학교 정보통신공학과
donggar97@naver.com, swpark@hufs.ac.kr

A Study on the Correlation between News and Bitcoin Price Changes

DongHyeok OH, SangWon Park
Dept. of Information and Communication Engineering,
Hankuk University of Foreign Studies

요 약

2017년 가치가 급상승하며 전 세계적으로 큰 이슈를 끈 비트코인은 최근 많은 사람들의 재테크 수단으로 이용되고 있다. 그러나 비트코인은 비슷한 재테크 수단인 주식과 다르게 24시간 내내 거래되고, 기사 하나하나에 의해 가격변동의 폭이 굉장히 크다. 이는 가격이 급변하는 비트코인 시장에서 가격을 예측하는데 어렵게 작용한다. 본 논문에서는 직접적인 가격 예측은 어렵다고 판단해 비트코인 가격변동에 영향을 주는 요소들을 딥러닝 모델을 통해 일일 단위 종가 가격의 등락을 예측해 위의 요소들이 비트코인 가격변동과 상관관계를 가지는지 확인한다.

1. 서론

최근 많은 사람들은 재테크 방식으로 비트코인을 많이 선택하고 있다. 비트코인은 2017년 가치가 급상승하면서 전 세계적으로 큰 이슈로 주목을 받으며 특히 젊은 세대들이 재테크 수단으로 선택하고 있다. 하지만 비트코인의 경우 등락폭이 굉장히 크고, 24시간 내내 시장이 열려 있기 때문에 언제, 어떤 일이 일어날지 몰라 안정적인 수익을 얻는 것은 굉장히 힘들다. 이와 같이 변동이 크고 불안정한 비트코인 시장에서 미래의 가격을 예측한다는 것은 매우 힘든 일이다. 그러나 많은 연구들은 다양한 모델을 사용하거나, 모델 튜닝 및 추가적으로 비트코인 가격에 영향을 주는 요인들을 가정해 입력 데이터로 사용한 모델을 통해 미래의 비트코인의 정확한 가격을 예측하려고 하지만 좋은 성과를 내고 있지는 못하고 있다.

여기에 본질적인 문제로 왜 재테크를 하는지가 중요하다. 재테크를 통해 큰 수익을 얻든, 적은 수익을 얻든, 재테크를 하는 사람들은 모두 수익을 얻기위해 자신의 자본을 투자한다. 이를 보면, 결국 얼마의 수익을 얻는 것 보다는 최대한 손실을 보지 않고, 수익을 얻어가는 것이 더 중요한 포인트라고 생각한다. 이 점을 생각해 이 연구에서는 다른 방식으로 비트코

인 가격 예측에 접근했다.

본 연구는 가격을 예측하기 굉장히 어렵다는 점과 더불어 투자를 할 때는 가까운 미래의 가격이 현재보다 올라갈지 또는 내려갈지에 대한 비트코인 가격의 방향성을 중요하게 생각해 다른 방법으로 예측 성과를 측정한다. 이전 연구들과는 다른 방법으로 하루 단위로 비트코인 종가 가격 등락 예측을 진행하며, 이때 비트코인 가격변동에 영향을 주는 요소로 판단되는 심리적 요인을 비트코인 가격변동 예측을 하는데 사용한다. 비트코인 가격에 변동을 주는 대표적인 심리적 요인으로는 국제 비트코인 시장과 금융, 경제적인 상황 등을 실시간으로 알려주는 비트코인 관련 뉴스 기사에서 키워드를 추출해 데이터로 사용한다. 뉴스 기사에서 유의미한 키워드들을 추출하기 위해 텍스트 마이닝 기법을 사용해 각 키워드들에 가중치를 부여하며, 부여된 가중치를 통해 하루 동안의 뉴스 기사 점수를 구해준다. 이와 같이, 비트코인 가격변동에 영향을 주는 요소로 판단되는 뉴스 기사를 사용해 비트코인 가격변동 예측을 진행해 비트코인 가격변동과 뉴스 기사가 어떤 상관관계를 가지는지 알아본다.

2. 관련 연구

기존 연구들은 시계열 데이터인 비트코인 데이터를 예측하기 위해 딥러닝 모델로 RNN을 사용했으며, 과거의 비트코인 데이터를 사용하기 위해 장기 의존성 문제를 보완한 장단기 메모리인 LSTM 모델을 사용했다[2]. LSTM 모델을 구축하면서 모델의 성능을 높이기 위해 은닉층의 수를 추가하거나 파라미터 튜닝을 통해 성능을 높여주었다.

뉴스 기사에서 대표적 키워드 또는 토픽을 추출하기 위한 기존 연구[2,3] 들은 텍스트 마이닝 기법을 사용한다. 해당 키워드가 얼마나 중요한 키워드인지 나타내는 통계적 수치로 TF-IDF(Term Frequency-Inverse Document Frequency)를 사용하며, 키워드들 간의 거리를 고려한 새로 고안된 TF-IDF로 의미 없는 키워드들을 제거했다.

3. 뉴스 데이터를 이용한 비트코인 가격 예측 모델

3.1 비트코인 관련 뉴스 키워드 수집

비트코인 가격변동 예측 모델 구축을 위해 비트코인 가격 데이터와 더불어 비트코인 가격 데이터 변동에 큰 영향을 주는 요소로 연구된 심리적 요인의 대표적인 예시인 뉴스 기사 데이터를 사용했다. 비트코인 가격 데이터는 세계 최대 거래소인 Binance에서 Binance API를 통해 1일 단위 종가 데이터를 사용했으며, 뉴스 기사 데이터는 한국의 주요 일간지 및 방송사의 뉴스 기사들에서 핵심 키워드와 핵심 내용들을 추출하기 위해 대표적인 뉴스 빅데이터 분석 서비스인 빅카인즈를 통해 데이터를 수집했다. 뉴스 기사의 경우 선정 기준으로 비트코인을 기본 키워드로 한 세계적인 경제 상황, 국제 금융, 비트코인 연관 기사들을 선정했다. 데이터 수집 기간은 비트코인이 폭발적인 관심을 받은 2017년 이후 어느정도 안정권에 접어들기 시작한 2018년부터 2022년 3월까지의 데이터를 수집했다. 이 중 2018년부터 2021년 9월까지의 데이터를 학습 데이터, 2021년 9월~12월 데이터를 검증 데이터, 2022년 1~3월 데이터를 테스트 데이터로 분류해 진행했다. 이 연구는 이 데이터들을 통해 다음날의 비트코인 종가가 전날의 종가에 비해 증가 되었는지, 감소 되었는지 확인했다.

3.2 키워드 사이의 거리를 고려한 TF-IDF 가중치

빅카인즈에서 뉴스 분석을 진행하면 각 기사에서 뉴스 기사 타이틀, 키워드, 본문 데이터, 특성추출 등을 활용해 데이터 전처리를 진행한다. 이 논문에서는

통계적 수치를 계산하기 위한 TF-IDF 가중치 출을 진행하는데 추출된 키워드를 키워드 간의 거리를 활용한 변형된 새로운 TF-IDF 가중치를 사용한다. 새로 고안된 TF-IDF 가중치를 사용하면 기존 가중치 추출방법보다 의미 없는 키워드들을 제거하는데 도움이 된다. 이렇게 계산된 가중치 중에서 특정 값 이상인 키워드들만 추출한다. 이때, TF-IDF 가중치 값이 높다는 것은 다른 날 뉴스 기사에 자주 나오지 않지만, 해당 날짜의 뉴스 기사에는 자주 등장하며, 주변 키워드들과의 연관성이 높다는 것을 의미한다. 이를 통해 해당 단어가 얼마나 그 날 비트코인 가격변동에 영향을 주는지를 가중치로 판단해 뉴스 기사 점수를 부여한다.

<식 1>은 키워드 사이간 거리를 고려하는 새롭게 변형된 TF-IDF 가중치 식이다. 이와 같이 계산된 키워드의 TF-IDF 가중치는 하루 동안 기사에 나오는 키워드들의 TF-IDF 가중치와 해당 날짜의 비트코인 가격변동 값과의 계산을 통해서 해당 날짜 하루의 뉴스 기사 점수를 부여한다.

<식 1>

$$f(t,d,l) = \sum (tf(t,d) \times \frac{1}{\sqrt{l}})$$

$$tf(t,d,l) = 0.5 + \frac{0.5 \times f(t,d,l)}{\max(f(w,d,l) : w \in d)}$$

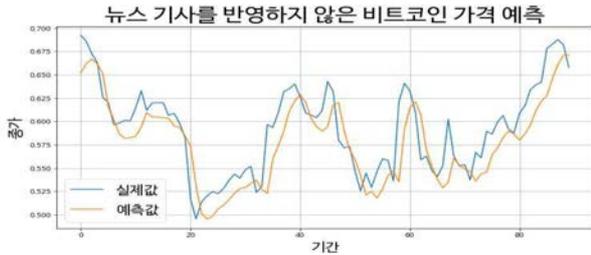
$$tf(t,d,l) \times idf(t,D)$$

3.3 비트코인 가격변동 예측 LSTM 모델

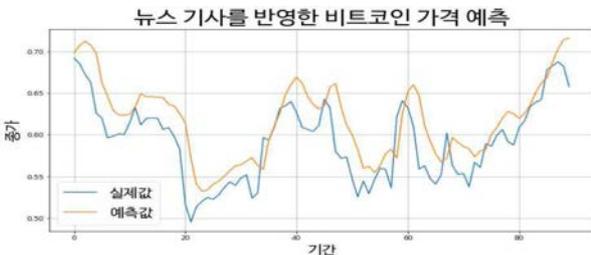
시계열 데이터인 비트코인 데이터 변동을 예측하기 위해 선행 연구[4, 5] 들에서 좋은 성과를 낸 모델인 LSTM 모델을 사용한다. 모델은 비트코인 종가 데이터만을 feature로 사용하는 모델, 뉴스 기사 데이터가 추가로 feature로 들어가는 모델 2개의 모델을 구축한다. 위의 2개의 모델을 통해 비트코인 가격 데이터만으로 예측을 한 경우보다 뉴스 기사 데이터가 들어갈 경우 예측이 얼마나 더 잘 되는지를 판단해 상관관계를 알아본다.

LSTM 모델을 구축하기 위해 먼저 비트코인 종가 데이터와 TF-IDF 가중치를 이용한 뉴스 기사 가중치 값을 MinMaxScaler을 사용해 정규화 시킨다. 정규화된 데이터에서 학습 데이터, 검증 데이터, 테스트 데이터로 분류를 한 뒤, 최적화를 위해서 지역 최소점에서 학습이 정제되는 기존 최적화 방법들의 문제를 개선하기 위해 AdamOptimizer을 사용하며 손

실함수로는 MSE(Mean Squared Error), 활성화 함수로는 하이퍼볼릭 탄젠트 함수를 이용한다. 추가로 은닉층을 추가해 성능을 향상하고, 학습, 검증 데이터를 활용해 Hyper-parameter 튜닝을 진행해 성능을 높였다.



[그림 1] 뉴스 기사를 반영하지 않은 비트코인 가격 예측



[그림 2] 뉴스 기사를 반영한 비트코인 가격 예측

4. 비트코인 가격변동 예측

모델의 성능은 뉴스 기사를 반영하지 않은 모델과 뉴스 기사를 반영한 모델의 파라미터 튜닝을 통해 가장 좋은 예측 성공률을 보여준 모델을 가장 좋은 성능을 보여주는 모델로 선정해서 2개의 모델의 성능을 비교했다. 여기서 예측 성공률은 오늘 비트코인 종가에 비해 다음날의 비트코인 종가의 변동을 맞추면 예측 성공률이 올라가는 방식이다. [표 1]에서는 뉴스 기사를 반영하지 않고 비트코인 가격 데이터만으로 학습된 첫 번째 모델은 52.81%의 예측 성공률을 보였으며, 뉴스 기사 데이터를 포함한 예측 결과는 55.18%의 예측 성과를 보여주면서 약 2%가량의 예측 성과 향상이 나타났다. 위의 [그림 1]과 [그림 2]는 각 모델이 예측한 결과와 실제 가격 데이터 변동에 대한 그래프이다. 파란색 그래프는 실제 비트코인 가격 그래프이고, 주황색 그래프는 예측된 비트코인 가격 그래프이다. 또한 3달 동안의 테스트 결과와 실제 비트코인 가격간의 편차는 뉴스 기사가 반영되지 않았을 때 0.0246, 뉴스 기사가 반영되었을 때 0.0184로 소폭 낮아진 것을 확인할 수 있다.

위와 같은 결과를 확인하면 비트코인 가격 변동 예측 성공률이 약 2%가량 증가하고, 실제 가격과 예

측 가격 간의 편차도 줄어들었지만, 예측하기 어렵다. 이를 통해 비트코인 가격변동은 무작위적인 성향을 보여주고 있으며, 뉴스 기사와 비트코인은 상관관계를 가지고 있다고 보기 어렵다.

	뉴스 기사를 반영하지 않은 경우	뉴스 기사를 반영한 경우
예측 성공률	52.81%	55.18%
실제 가격과 편차	0.0246	0.0184

[표 1] 뉴스 기상 반영 유무에 따른 예측 성과와 실제 비트코인 가격과 예측 가격의 편차

5. 결론 및 향후 연구

본 논문에서는 LSTM 모델을 구축해 비트코인 가격변동과 뉴스 기사가 어떤 상관관계를 가지는지 알아보았다. 뉴스 기사를 반영해 예측한 경우가 약 2%가량의 예측 성과 향상을 보여주었으나, 비트코인 가격변동의 주된 이유를 심리적 요인인 뉴스 기사 데이터를 반영했지만, 2%의 예측 향상으로는 상관관계를 가지고 있다고 보기 어렵다.

뉴스 기사는 경제 시장 및 국제 금융 관련 이슈가 있을 때 실시간으로 기사들이 나오고 그에 따라 가격변동이 급격히 일어나기 때문에 하루 뒤의 비트코인 가격이 하루의 뉴스 기사 데이터로 판단하기에는 한계가 있다. 이를 극복하기 위해서 하루 단위를 분 단위로 가격변동 예측을 진행할 수 있도록 실시간 모델을 구축한다면 더 높은 예측 성과 향상이 가능하다고 보인다. 또한 뉴스 기사 데이터뿐만 아니라 심리적 요인으로 작용하는 SNS 데이터를 추가해, 보다 향상된 예측 성과가 나오도록 연구를 진행할 예정이다.

참고문헌

[1] 박민규, 사회적 감성이 가상화폐 가격에 미치는 영향분석 : 실물화폐와의 비교를 중심으로, 대한산업공학회지, 47권, 1호, p.23-33, 2021.02
 [2] 이주영, 단어 사이의 거리를 활용한 Bi-gram 기반 주제-키워드 그래프 생성 시스템, 한국정보과학회, 온라인, 2020, p.404-406
 [3] 원종관, 텍스트 마이닝과 딥러닝을 활용한 비트코인 가격 예측 : 한국과 미국시장 비교, 한국지식경영학회, 온라인, 2021, p.1-17
 [4] 김준호, 비트코인 가격 예측을 위한 LSTM 모델의 Hyper-parameter 최적화 연구, 한국융합학회,
 [5] 강민규, LSTM 기반 감성분석을 이용한 비트코인 가격 등락 예측, 한국통신학회, 온라인, 2020, p.561-562