

LSTM 기반 감성분석을 이용한 비트코인 가격 등락 예측

강민규, 김보선, 신문곤, 백의준, 김명섭

고려대학교

{cxz3619, boseon12, tm0309, pb1069, tmskim}@korea.ac.kr

LSTM-based Prediction of Bitcoin Price Fluctuation using Sentiment Analysis

Mingyu Gang, Boseon Kim, Mu-Gon Shin, Ui-Jun Baek, Myung-sup Kim

Korea Univ

요약

사토시 나카모토에 의해 처음 소개된 비트코인은 현재 가장 가치 있는 통화로 간주되며 놀라운 성장을 이뤄냈다. 비트코인의 성장에 따라 가격 또한 큰 폭으로 증가했으며, 가격을 예측하기 위해 트위터, 가격지표들을 이용한 연구들이 선행되었다. 그에 비해 뉴스 기사를 활용한 연구들은 수가 적다. 본 논문은 비트코인의 가격등락 예측을 목표로 하며 가격예측을 위한 데이터로 전날의 비트코인 관련 기사들을 이용한다. 가격예측을 위한 방법으로 감성 분석과 딥 러닝 알고리즘 중 하나인 LSTM 모델을 사용하여 비트코인의 가격등락을 예측한다.

I. 서론

2009년 사토시 나카모토를 통해 처음 소개된 비트코인은 최초의 탈중앙화 암호 화폐이며 다른 알트코인, 가상화폐의 근본이 되는 암호화 화폐이다. 그 후 2010년 약 1달러였던 가격이 2020년 8월 11000달러에 이를 만큼 놀라운 성장을 이뤘으나 가격 변동성 또한 커지면서 비트코인의 가격을 예측하기 위한 연구들이 진행되었다.

선행된 연구 중에 [1]은 베이저안 최적화를 기반으로 ARIMA 모델, RNN 모델, LSTM 모델간의 성능차이를 분석해 LSTM 모델 성능의 우세함을 보여주었고, [2]는 비트코인의 Close, Open, High 및 Low 등의 가격지표를 이용하여 LSTM 모델과 GRU모델의 성능을 비교했다. [3]은 트윗을 전처리하여 감성분석기를 이용해 긍정적인 트윗과 부정적인 트윗을 구분한다. 그 후 RNN 모델을 이용해 77.62%의 가격 예측 정확도를 보였다. [4]는 비트코인 관련 Google Trend의 데이터와 트위터의 양과 양의 상관관계에 있음을 밝혔다. 비트코인 가격예측에 있어 가격지표와 트위터를 이용한 연구들은 많지만 뉴스를 활용한 연구들은 수가 적었다. 따라서 본 논문은 가장 큰 웹사이트 중 하나인 Google의 기사들을 활용해 연구를 진행했다.

본 논문의 목표는 전날의 비트코인 관련 기사들을 이용해 비트코인의 가격등락을 예측하는 것이다. 먼저 전날의 비트코인 관련 기사들을 수집해 감성 분석 처리한다. 처리된 값들은 부정/긍정적인 기사를 분류하고 1일 평균 감성 수치를 제공한다. 이 값들을 이용해 딥 러닝 모델 중 LSTM 모델을

이용해 비트코인 가격등락을 예측한다.

본 논문은 1장 서론에서 연구 배경과 목표를 서술하고, 2장 본론에서 LSTM 기반 감성 분석을 이용한 비트코인 가격등락 예측 방법에 대해서 서술한다. 3장 실험 및 결과에서 실험의 하이퍼 파라미터들과 결과에 대해 서술하며, 4장에서는 결론 및 향후연구를 언급하며 논문을 마친다.

II. 본론

본 장은 비트코인 가격 등락을 예측하기 위한 방법론을 설명한다. 방법론에는 데이터 수집, 감성분석을 이용한 데이터 전처리, 딥 러닝 과정으로 이루어져 있다. 구조도는 다음과 같다.

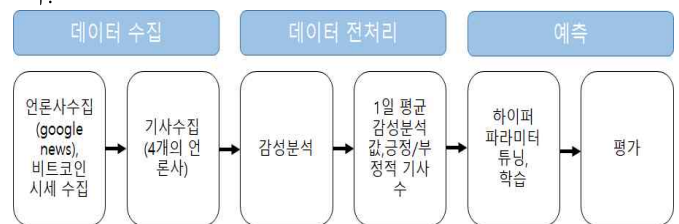


그림 1. 구조도

2.1 데이터 수집

본 절에서는 웹크롤러에 대해 언급한다. 웹 크롤러란 조직적, 자동화된 방법으로 월드 와이드 웹을 탐색하는 컴퓨터 프로그램이다. 주로 Web crawler는 방문한 사이트의 복사본을 생성하거나 이메일을 자동으로 수집하는 곳에 사용된다.

본 방법에서는 크롤링 과정에서 다른 방법을 제안한다. 구글 뉴스에서 직접 뉴스들을 크롤링하기에는 두 가지 문제점이 있다. 첫째, 언론사 별 웹페이지 구조가 다르다. 두 번째로, 키워드 '비트코인' 입력 시 비트코인 관련기사 혹은 광고 또한 검색된다.

다음과 같은 이유로 Google news에서 2013.01.01.부터 2020.07.30.까지 비트코인을 언급한 언론사를 수집했다. 해당 기간 동안 키워드 비트코인을 많이 언급한 상위 6개의 언론

※ 이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2018R1D1A1B07045742)과 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2018-0-00539-003,블록체인의 트랜잭션 모니터링 및 분석 기술개발)

사는 다음과 같다.

언론사	언급횟수
Coindesk	3724
Bitcoin News	2633
coinTelegraph	1987
Forbes	1100
newsBTC	446
CriptoNoticias	397

표 1. 수집된 상위 6개 언론사

키워드 비트코인을 1000번 이상 언급한 언론사를 영향력 있고, 신뢰성 있는 언론사로 판단하여 해당 언론사에서 비트코인 관련 기사들을 모두 수집했다. 수집된 기사들은 일 단위로 파일별로 구분했으며 한 줄에 [제목, 본문] 형태로 구분했다.

2.2 데이터 전처리

감성 분석이란 일반 대중들의 기분과 감정을 해독하고 맥락에 대한 통찰력 있는 정보를 제공하는 것을 목적으로 한다. TextBlob은 파이썬의 라이브러리로 텍스트로 된 데이터에 감성을 복잡한 분석 및 작업을 통해 수치화된 데이터로 만들어준다. TextBlob 감성분석은 어휘 사전 기반 방식이며, 감정은 사전에 정의된 강도에 따라 정의된다. 모든 단어에 개별 점수를 할당한 후 평균, 가중치 등의 작업에 따라 문장의 최종 감정을 계산한다. TextBlob은 문장의 극성과 주관성을 반환하는데 극성은 -1에서 1까지의 값이며 -1은 부정적인 감정을 의미하고 1은 긍정적인 감정을 의미한다.

수집된 기사들은 텍스트 형태의 데이터로 저장되어 있다. 딥 러닝 모델의 Input 값으로는 수치로 된 값들이 들어가기 때문에 수집된 기사들을 직접 활용할 수 없다. 그렇기 때문에 기사들을 감성 분석을 통해 수치화된 값으로 변환해주는 과정이 필수적이다.

감성 분석을 하는 범위로는 기사 제목, 본문 전체에 감성 분석 처리하였다. 기사 하나에 대한 감성 분석 수치를 이용해 1일 단위의 감성 분석 평균, 긍정적인 기사의 비율을 도출하였다. 이때 긍정/부정적인 기사를 나누는 기준은 감성 분석 수치가 0 미만이라면 부정적인 기사, 0을 초과하면 긍정적인 기사로 분류했다.

2.3 예측

LSTM은 신경망의 일종으로 시계열 데이터를 다루는데 특화된 딥 러닝 모델이다. 1일 단위의 감성 분석 평균값은 그림2와 같이 시계열의 데이터 형태로 나타나기 때문에 LSTM 모델의 input 값으로 적합하다. LSTM 모델의 feature들로는 총 기사의 수, 긍정적인 기사의 수, 1일 단위 감성분석 평균값을 넣었으며 label로는 증가를 설정했다.



그림 2.시계열 형태의 1일단위 감성분석 평균 값

III. 실험 및 결과

딥 러닝 단계에 있어, 예측 모델의 하이퍼 파라미터에 따라 예측 모델의 성능이 크게 달라지므로 최적의 하이퍼 파라미터를 조정하는 과정은 필수적이다. 하이퍼 파라미터 조정 관련 실험으로 훈련 셋과 테스트 셋의 비율, 은닉 층의 유닛의 개수에 관한 실험을 진행했다. 훈련 셋과 테스트 셋 수를 훈련 셋 :1208, 테스트 셋:100 개로 두고 두 개의 은닉 층과 마지막 하나의 sigmoid 출력 층을 두고 은닉 층의 유닛의 수를 조절해가며 실험했다. 그 결과 두 개의 은닉 층과 유닛이 각각 4개일 때 가장 좋은 정확도인 58.75%의 정확도를 보여주었다.

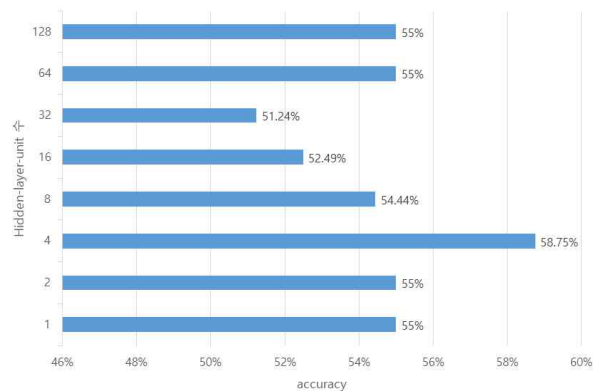


그림 3.은닉 층 유닛 수에 관한 accuracy

IV. 결론 및 향후 연구

실험 결과, 전날의 비트코인에 관한 기사가 다음날의 비트코인에 대해 어느 정도 영향을 끼침을 알 수 있다. 적은 수의 feature에도 불구하고 58.75% 정확도를 보여주는 것은 긍정적인 성과이다. 하지만 여전히 비트코인 시세와 관련 기사 간의 강한 상관관계를 보여주기에는 여전히 부족한 결과이다. 향후 연구계획으로는 Google news의 4개의 상위 언론사 외에 다른 언론사와 Yahoo의 기사를 수집한 연구를 진행할 것이다.

참 고 문 헌

[1] S. McNally, J. Roche and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning," (PDP), Cambridge, 2018, pp. 339-343

[2] T. Phaladisailoed and T. Numnonda, "Machine Learning Models Comparison for Bitcoin Price Prediction," (ICITEE),Kuta,2018, pp. 506-511

[3] D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel and B. K. Lama, "Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis," (ICCCS), Kathmandu, 2018, pp.128-132

[4]Martina Matta, Ilaria Lunesu, Michele Marches, 'Bitcoin Spread Prediction Using Social And WebSearch Media' UMAP workshops, 2015 - academia.edu.