

# 뉴스 데이터로 LSTM 활용하여 이더리움 트랜잭션 개수 예측

\*김태연, 백의준, 신무곤, 김명섭  
고려대학교

{ingcubators, tm0309, pb1069, tmskim}@korea.ac.kr

## Design of LSTM Model for Estimation of Volume of Ethereum Transaction

Tae-Yuen Kim, Mu-Gon Shin, Uijun Baek, Myung-Sup Kim  
Korea Univ.

### 요 약

블록체인 기술을 기반으로 만들어진 암호 화폐 플랫폼 중 하나인 이더리움은 Peer-to-Peer 기술을 사용하는 비트코인의 2.0 버전이라고도 불리는 온라인 암호화폐이다. 뉴스에서 암호화폐를 언급한 경우에 그 가치에 따라서 암호화폐의 가격이 급격한 등락으로 반응이 나타나며, 이는 거래소에 투기자들이 급격하게 모여 거래를 한다는 이야기가 된다. 이때 트랜잭션(거래량)이 가격의 변동에 비례하여 일일 거래량의 수가 급증하고 이를 통해 뉴스 데이터를 통하여 딥러닝 중 예측이 가능한 LSTM(Long Short Term Memory)으로 학습을 시켜 예측이 가능하리라 추측해 보았다. 본 논문은 딥러닝 알고리즘 LSTM 을 사용하여 이더리움 트랜잭션의 변동 예측을 실행해 본다.

### Abstract

Ethereum, one of the crypto-currency platforms built on block chain technology, is an online crypto-currency, also called the 2.0 version of Bitcoin that uses peer-to-peer technology. If the news mentions the crypto, the price of the crypto will fluctuate sharply depending on its value, which means that speculators gather on the exchange to make transactions. At this time, it was assumed that the transaction volume would increase in proportion to the change in price, which would allow the daily transaction volume to be predicted by learning from the news data into the predictable LSTM (Long Short Term Memory) during deep learning. This paper uses deep learning algorithm LSTM to execute the prediction of variation in Ethereum transaction.

### I. 서 론

많은 투자자가 비트코인 출범 이후, 이익을 위해 암호화폐 시장에 뛰어들었고 암호화폐 거래소에서는 트랜잭션, 블록체인 등 많은 요청을 처리한다. 이러한 상황에서 투자자들이 특정 매체를 이용하여 정보를 취득 후 특정 시간대에 거래가 급증하게 되어 암호화폐 거래소의 서버 처리속도가 느려지는 등의 결과로 인해 지연이 생기는 일이 생긴다. 이러한 일을 방지하기 위해 더 큰 비용을 들여 좋은 서비스를 제공하는 경우도 있지만, 상시로 큰 비용을 지불하는 것에 거래소 입장에서는 별 좋은 선택이 아닐 것이다. 하지만 미리 처리해야 하는 트랜잭션 수량과 연산을 예측할 수 있게 된다면, 저렴한 유지 비용을 지불하면서 최상의 서비스를 제공할 수 있으리라 생각 된다. 이러한 예측을 하기 위해서는 딥러닝 알고리즘 중 LSTM 을 적용하는 것을 제안한다.

### II. 관련 연구

본 논문의 서론에서와같이 이더리움 트랜잭션 수를 예측하기 위해서는 딥러닝(Deep Learning) 기술 중 시간을 이용하여 데이터를 구분하는 시계열 데이터 예측

알고리즘이 필요하다고 판단되었다. 대표적인 시계열 데이터 예측 알고리즘은 RNN 과 LSTM 이 있다.

#### 2.1 순환 신경망: RNN

최근에 발생한 과거로부터 앞으로 어떻게 될 것인지 예측을 시도하는 것이 첫 시작이었다. 딥러닝 모델 중에서 RNN 은 일반 신경망인 NN 에서의 시계열 개념이 추가되었고 은닉계층(Hidden\_Layer)에서 이전 정보를 기억 할 수 있도록 바로 전 학습한 결과와 현재 입력받은 정보를 합쳐서 같이 학습하게 되어 시간상으로 입력되는 순서도 중요해지는 시계열 데이터에 관점에서 적합한 모델이다. 이러한 흐름을 그림 1 번에서 표현하였다.

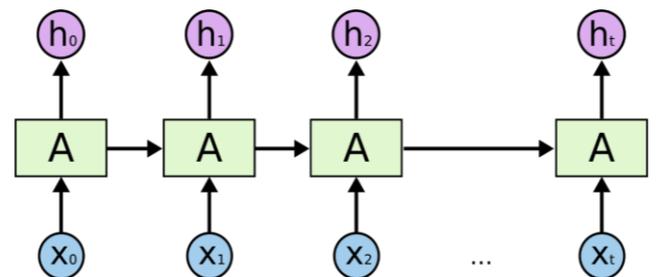


그림 1. RNN 흐름의 표현

#### 2.2 장기의존형 순환신경망: LSTM

LSTM 은 학습 시 중심이 되는 부분인 기울기가 중간에 의해 소실되어 생기는 문제를 해결하기 위해

이 논문은 2018 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2018R1D1A1B07045742)과 2018 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2018-0-00539-001, 블록체인의 트랜잭션 모니터링 및 분석 기술개발)

과거 데이터를 장기적으로 의존하는 장기 의존성을 학습할 수 있는 특수한 목적을 지닌 순환 신경망(RNN)이다. 이러한 기울기가 손실되는 것을 방지하기 위해 RNN 구조에서 신경망 사이에 셀(Cell) 메모리 장치를 적용하며 이 구성을 유지하기 위한 게이트(Gate)가 입력되어 있다. 게이트는 기능별로 망각, 입출력을 담당하며 망각 게이트인 경우 이전에 입력된 셀의 정보 중 어느 내용 얼마큼 손실을 내는가에 대한 게이트이며, 입력 게이트는 새로 입력되는 값 중 전체 혹은 일부만 받아들일지 결정하는 게이트이다. 또한, 출력 게이트는 반대로 얼마나 출력해 낼지 결정하는 게이트이다. 이러한 표현은 그림 2에서 표현하고 있다.

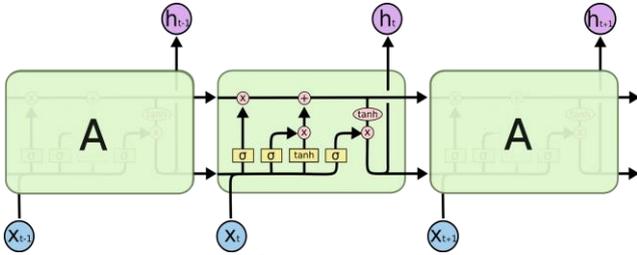


그림 2. LSTM(순방향)

하지만 이러한 순방향적인 은닉계층을 가지게 될 경우 입력 순서에 따라 결과물은 과거 데이터의 학습 패턴만을 기반으로만 의존하는 한계가 보인다. 이러한 문제점을 해결하기 위해 역방향으로 진행되는 은닉계층을 추가하여 성능을 향상 시켰다. 이러한 방식을 양방향 LSTM 이라 명칭 한다. 이러한 흐름은 그림 3에서 표현하고 있다.

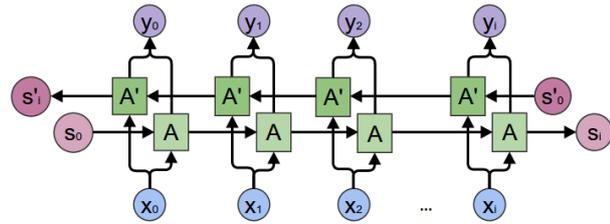


그림 3. LSTM(양방향)

### III. 본론

#### 3.1 데이터 수집

트랜잭션 거래 개수인 경우 EtherScan 홈페이지에서 제공하는 Daliy Transactions Chart 부분에서 공개되어 있는 트랜잭션의 총 일일 거래량을 자료를 받아온다.

뉴스 데이터인 경우 각 뉴스 사에 검색하여 얻어내는 것은 시간 대비 효율이 낮다고 생각하였고, 구글 뉴스 검색 기록을 활용하고자 크롤링 프로그램을 제작하였다. 크롤링 프로그램은 2 가지로 구분되어서 제작되었다. 먼저 이더리움을 언급한 뉴스사의 뉴스페이지 링크를 추출하기 위하여 구글 뉴스 페이지를 검색하고 크롤링하여 수집을 시작한다. 추출된 각 뉴스 페이지 링크를 가지고 다시 크롤링하여 뉴스 페이지 전체인 HTML 을 날짜와 순서대로 저장한다.

#### 3.2 데이터 가공

이 과정에서는 2 가지 방법을 통해 데이터가 가공된다. 데이터 수집에서 가져온 뉴스 데이터 전체를 단어별로 나누어 파이썬 모듈 중 konlpy.tag 에 집어넣게 되면 부사, 동사, 조사 등을 제거하여 명사만 남아 있게

만들며, 이때 링크로 걸려 있는 광고나 웹 상에 태그 등은 영어로 인식되어 제거 되기 때문에 이 과정에서 자동으로 삭제된다. 단어 수집이 완료되면 중복되어 있는 단어는 제거해 가며 단어 사전을 완성한다

단어 사전을 만든 이후 EtherScan 홈페이지에서 가져온 총 일일 거래량 자료를 가져와 일별 거래량 테이블을 제작한다. 테이블 작성 시 전날과 비교하여 전날 대비 감소한 경우 -1 로 표기하였으며 감소하지 않았을 경우 전날보다 증가한 퍼센트만큼의 숫자를 입력하였다.

이 이후 다시 뉴스 데이터 전체를 가져와 단어 사전에 있는 단어와 비교하며 일별 단어 카운트를 계산하고, 테이블 형태로 작성을 시작한다. 이때 거래량 테이블을 일별에 맞추면서 작성하며, 총거래량 단어 테이블을 만든다.

#### 3.3 LSTM

이 실험 환경에서는 Tensorflow v.2.3.0 에서 사용하였다. Learning rate 는 0.1 과 0.5 그리고 1.0 으로 3 가지 방법으로 설정하였으며, 200 번으로 반복 학습을 수행하였고, 데이터는 3.2 에서 가공한 총거래량 단어 테이블을 가지고 생성한다.

실험은 증가를 예측하는 경우와 감소를 예측하는 경우 2 가지로 예측을 시도하였고 그 결과는 표 1 에 표현하고 있다.

	Rate 0.1	Rate 0.5	Rate 1.0
최소 RMSE	0.0081	0.0070	0.0106
예측률	47.95%	48.06%	48.06%

표 1 평균 제곱근 오차(RMSE)및 예측률 결과

Learning rate 의 경우 0.5 로 설정했을 때 최소 RMSE 가 적게 나왔다. 예측률은 평균 약 48% 정도로 뉴스 데이터를 이용하여 가까운 미래의 거래량을 예측이 가능하다는 것을 증명하였다.

### IV. 결론

본 논문에서는 양방향 LSTM 순환 신경망을 이용하여 거래량 등락 예측 모델을 구현하고 실험을 통해 제안하는 방법의 성능을 측정하였다. 그 결과는 뉴스 데이터를 이용하여 특정 단어가 거래량을 등락을 결정하였고, LSTM 을 이용하여 가까운 미래에 등락을 예측이 가능하다는 것을 일부분으로 증명하였다. 양측을 동시에 예측할 경우 예측률이 급격하게 떨어지게 되는 데 이는 뉴스 등락만을 예측을 시도할 경우 다른 데이터가 필요하다는 결론을 얻었다. 향후 연구에서는 SNS 의 데이터를 추가하여 연구를 진행할 계획이다.

### 참고 문헌

- [1] 신동하, 최광호, and 김창복. "RNN 과 LSTM 을 이용한 주가 예측을 향상을 위한 딥러닝 모델." 한국정보기술학회논문지 15.10 (2017): 9-16.
- [2] 지세현, et al. "비트코인 트랜잭션 수 예측을 위한 LSTM 모델 설계." 한국통신학회 학술대회논문집 (2019): 472-473.