

기계학습 기반 비트코인 네트워크 트랜잭션 수 예측에 관한 연구

지 세 현*, 백 의 준*, 신 무 곤*, 박 준 상**, 김 명 섭^o

A Study on the Prediction of Number of Bitcoin Network Transactions Based on Machine Learning

Se-Hyun Ji*, Ui-Jun Baek, Mu-Gon Shin, Jun-Sang Park, Myung-Sup Kim^o

요 약

블록체인 기술을 기반으로 만들어진 비트코인은 Satoshi Nagamoto에 의해 개발된 온라인 암호화폐이다. 2009년 1월 3일 최초로 발행된 비트코인은 트랜잭션 수의 증가와 함께 급속도로 발전 중이다. 그러나 비트코인 트랜잭션 수의 증가에 따른 부작용이 발생하고 있다. 비트코인 트랜잭션 수를 예측하는 것은 비트코인 네트워크에 발생하는 부작용에 대비하기 위해 중요하다. 본 논문은 두 가지 기계학습 알고리즘을 적용하여 비트코인 트랜잭션 수를 예측하는 모델을 설계한 뒤, 실험을 통해 비트코인 트랜잭션 수를 예측하는 모델을 제안한다.

Key Words : Blockchain, Bitcoin transaction, Machine Learning

ABSTRACT

Bitcoin, based on the blockchain technology is an online crypto-currency developed by Satoshi Nagamoto. Bitcoin, which was first issued on January 3, 2009, is rapidly evolving with increasing number of transactions. However, untoward incidents are occurring due to an increase in the number of Bitcoin transactions. Predicting the number of Bitcoin transactions is important to prepare for any issues that can occur in the Bitcoin network. This paper proposes to design model for predicting the number of Bitcoin transactions by applying two machine learning algorithms and then a model for predicting the number of Bitcoin transactions through experiments.

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

• Korea University Department of Computer and Information Science, sxzer@korea.ac.kr

^o Corresponding Author : Korea University Department of Computer and Information Science, tmskim@korea.ac.kr

* Korea University of Department of Computer and Information Science, {pb1069, tm0309}@korea.ac.kr

** LG Electronics CTO, jusang.park@lge.com

논문번호 : KNOM2019-01-14, Received June 24, 2019; Revised July 6, 2019; Accepted August 15, 2019

I. 서론

Satoshi Nagamoto에 의해 개발된 비트코인은 정부, 금융기관의 개입 없이 개인 간의 빠르고 안전한 거래를 하기 위한 목적으로 발행된 온라인 암호화폐이다[1]. blockchain.com에서 제공한 그림 1에서 보듯이 24시간 단위로 확인된 비트코인 트랜잭션 수는 갈수록 증가하는 추세를 보인다[2]. 비트코인 트랜잭션 수가 증가와 함께 비트코인 네트워크는 급속도로 발전 중이다.

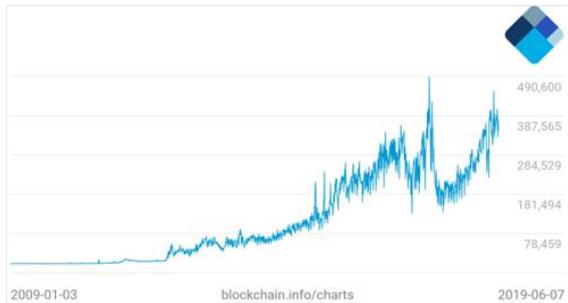


그림 1. 24시간 단위 확인된 비트코인 트랜잭션 수
Fig. 1. Bitcoin Transaction Counts in 24-hour Increments

그러나 비트코인 네트워크가 급속도로 발전하고 있지만, 그에 따른 부작용이 발생한다. 예를 들자면, 비트코인 트랜잭션을 처리하는 비용은 증가했지만, 트랜잭션을 확인하는 시간은 지연되고 있다[3]. 비트코인 네트워크에서 발생한 트랜잭션 수는 비트코인의 가치에 영향을 끼치는 요소이다. 이러한 이유로 비트코인 트랜잭션 수를 예측하는 것은 비트코인 네트워크에서 발생하는 부작용에 대비 및 유지하기 위해 중요하다. 비트코인 트랜잭션 수는 시간의 흐름 순으로 발생하는 시계열(Time-Series) 데이터이다. 보편적으로 시계열 데이터를 예측하는 데 사용하는 기법은 ARIMA(Auto-regressive Integrated Moving Average)모형을 이용한다. 그러나 ARIMA 모델은 안정적 시계열(Stationary Series) 데이터에만 우수한 성능을 보일 뿐만 아니라, 비안정적 시계열(Non Stationary Series) 데이터에 대해서는 안정화를 거쳐 예측을 시도해야 하는 복잡함이 있다. 따라서 비안정적 시계열 데이터인 비트코인 트랜잭션 수를 기계학습을 적용하여 예측하는 것은 불규칙적인 패턴으로 발생하는 트랜잭션 수를 자동으로 학습하여 기존의 ARIMA 모델보다 우수한 성능을 보일 수 있다[4]. 본 논문은 기계학습을 적용하여

비트코인 트랜잭션 수를 예측하는 모델을 설계하여 실험을 통해 제안한다. 기계학습 모델의 성능을 결정하는 것은 학습데이터와 모델을 구성하는 Hyper-parameter이다. 학습데이터의 종류와 모델을 구성하는 Hyper-parameter의 설정에 따라 모델의 성능은 달라지기 때문에, 적합한 학습데이터와 Hyper-parameter를 설정하는 것은 중요하다[5].

본 논문은 1장 서론에서 연구 배경과 목표를 서술하고, 2장 관련 연구에서 기계학습을 적용하여 비트코인 데이터를 예측하는 연구에 관하여 서술한다. 3장 본문에서 기계학습 기반 비트코인 트랜잭션 수 예측 방법에 관하여 서술하고, 4장 실험 및 결과에서 두 가지 방법의 기계학습을 적용한 모델의 실험 결과를 보여준다. 마지막 5장에서 결론 및 향후 연구를 언급한 뒤 논문을 마친다.

II. 관련 연구

기계학습을 적용하여 비트코인 데이터를 예측하는 다양한 연구가 진행 중이다. 예측에 사용된 기계학습 모델은 선형회귀 모델, Recurrent Neural Network(RNN) 모델과 RNN 모델의 구조를 개선한 Long-Short-Term-Memory(LSTM) 모델이다.

1. 선형회귀 모델 기반 예측

선형회귀는 종속변수와 한 개 이상의 독립변수와의 선형 상관관계를 분석하는 기계학습 기법이다[6]. 선형회귀 모델의 구조는 그림 2와 같다. 선형회귀 모델은 종속변수 Y 와 한 개 이상의 독립변수 x 에 대해 가설을 세운 뒤, 학습을 통해 가설에 가장 적합한 W 와 b 값을 찾는 방법이다.

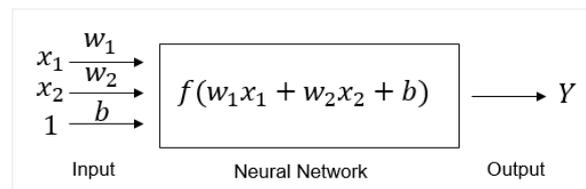


그림 2. 선형회귀 모델 구조
Fig. 2. Structure of Linear Regression Model

[7]은 비트코인 트랜잭션 데이터를 선형회귀 모델의 학습데이터를 사용하여 비트코인의 가격을 예측한다. 학습된 모델은 비트코인 가격의 상승 또는 하락에 대해 55%의 예측 정확도를 나타낸다. [8]은 비트코인 가격을 예측하는 선형회귀 모델을 설계한

뒤, 선형회귀 모델의 예측한 결과를 바탕으로 비트코인에 투자한다. 학습된 모델을 이용한 투자는 기존의 기계학습을 적용하지 않은 기존 투자방식보다 향상된 결과를 얻는다. 또한, 선형회귀 모델의 학습 데이터의 종류에 따라 모델의 성능이 바뀌는 것을 확인한다.

2. RNN 모델 기반 예측

RNN은 시계열 데이터를 학습하는데 특화된 기계 학습 기법이다^[9]. RNN 모델의 구조는 그림 3과 같다. RNN Cell 내부에는 쌍곡선 탄젠트(Hyperbolic Tangent) 함수를 사용하며, t 는 시간의 흐름, x 는 입력 값, y 는 결과 값 h 는 RNN Cell의 출력 값, w 는 가중치를 의미한다. 따라서 시간 t 의 학습 결과인 h_t 는 $t-1$ 에서의 학습을 통해 나온 h_{t-1} 과 입력 값 x_t 에 대해 각각의 가중치 w_x , w_h 와의 연산을 통해 나온다.

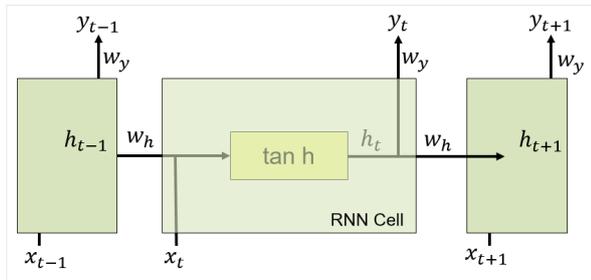


그림 3. RNN 모델 구조
Fig. 3. Structure of RNN Model

[10]은 비트코인의 환율을 예측하는 RNN 모델을 설계한다. RNN 모델의 학습데이터로 비트코인 블록데이터를 사용한다. 시계열 데이터인 비트코인 블록데이터가 RNN 모델에 효과적인 학습데이터로 사용할 수 있음을 보인다. [11]은 비트코인의 가격을 예측하기 위해 RNN 모델을 설계할 때 트위터에 대한 정서를 학습데이터로 사용한다. 설계된 RNN 모델의 가격예측 정확도는 77.62%를 나타낸다.

3. LSTM 모델 기반 예측

LSTM은 RNN 모델의 구조를 개선한 학습 기법이다^[12]. LSTM 모델의 구조는 그림 4와 같다. LSTM Cell은 기존 RNN Cell의 구조에서 시그모이드(Sigmoid) 함수와 곱셈 연산, 덧셈 연산을 추가한 구조이다. c 는 시간 t 일 때 LSTM Cell의 상태를 말한다.

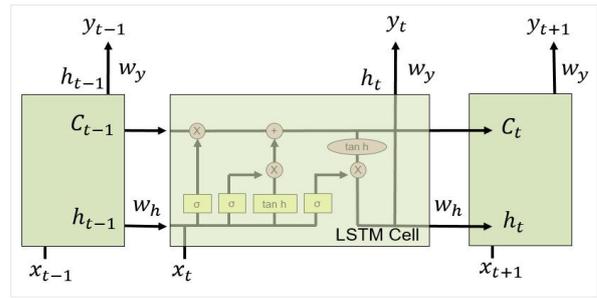


그림 4. LSTM 모델 구조
Fig. 4. Structure of LSTM Model

[13]은 비트코인 가격 관련 데이터를 LSTM 모델의 학습데이터로 사용하여 ARIMA 모델과의 성능을 비교한다. LSTM 모델의 성능이 ARIMA 모델보다 우수한 성능을 보인다. [14]는 선형회귀 모델을 설계하여 비트코인의 환율을 예측하였다. 선형회귀 모델에 사용된 학습데이터를 LSTM 모델에 사용한다면 더 좋은 성능을 보일 수 있음을 언급한다.

III. 본 론

기계학습 기반 비트코인 트랜잭션 수 예측모델 설계과정은 그림 5와 같다. 비트코인 블록 및 트랜잭션 데이터 수집, 통계처리, 데이터 분석, 학습데이터 생성, 기계학습 모델 선정, Hyper-parameter Optimization의 과정을 통해 기계학습 모델을 생성하고 검증 과정을 통해 기계학습 모델의 성능을 평가한다.

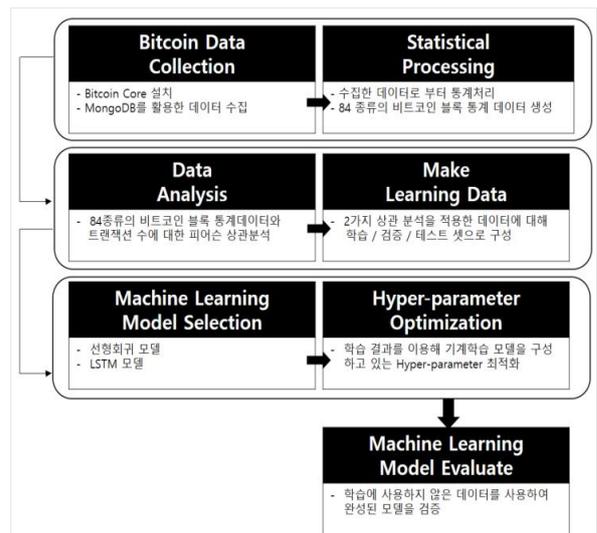


그림 5. 기계학습 모델 설계과정
Fig. 5. Design Process of Machine Learning Model

1. 비트코인 블록 및 트랜잭션 데이터 수집

비트코인 코어로부터 비트코인 블록에 담긴 블록 데이터와 트랜잭션 데이터를 수집하는 과정이다. 수집을 위해 CentOS 운영체제를 사용하고, 데이터베이스 소프트웨어는 MongoDB를 사용한다. 수집한 총 12종류의 블록 및 트랜잭션 데이터 종류에 대한 자세한 설명은 표 1과 같다.

표 1. 비트코인 블록/트랜잭션 데이터 종류
Table 1. Category of Bitcoin Block/Transaction Data

데이터 항목	의미
nTx	블록에 포함된 트랜잭션 수
Weight	블록의 Weight
Size	블록의 사이즈
vSize	블록의 가상 크기
nVin	트랜잭션이 포함하고 있는 input 의 수
nVout	트랜잭션이 포함하고 있는 Output 의 수
Value	트랜잭션의 거래 금액
Fee	트랜잭션 수수료
Tx.Size	트랜잭션의 크기
Tx.Vsize	트랜잭션의 가상 크기
Vin.value	트랜잭션의 입력 값
Vout.value	트랜잭션의 출력 값

표 1로부터 수집한 비트코인 블록 및 통계데이터로부터 통계정보를 얻기 위해 통계처리를 한다. 트랜잭션 단위에 대해서 통계처리를 적용하고, 통계처리의 종류는 합, 곱, 최대, 최소, 평균값을 구한다. 통계처리를 통해 수집한 데이터에 대한 자세한 설명은 표 2와 같다. 트랜잭션 단위의 데이터 6가지 항목은 1번의 통계처리를 하고, 2가지 항목은 2번의 통계처리를 한다. 총 84종류의 비트코인 블록 및 트랜잭션의 통계데이터로 구성한다.

표 2. 비트코인 블록/트랜잭션 통계데이터
Table 2. Statistical Data of Bitcoin Blcok/Transaction

데이터 단위	데이터 항목	1 st 통계처리	2 nd 통계처리	항목 수	
블록	nTx			1	
	Weight			1	
	Size			1	
	Vsize			1	
트랜잭션	nVin			5	
	nVout			5	
	Value			5	
	Fee		Sum Max	5	
	Tx.Size		Min Avg	5	
	Tx.Vsize		Stdv	5	
	Vin.value	Sum Max		25	
	Vout.value	Min Avg		25	
		Stdv			

2. 비트코인 통계데이터 상관분석

상관분석은 두 변수 사이의 선형적 관계 유/무를 분석하는 방법이다. 두 변수는 서로 독립적인 관계이거나 상관된 관계일 수 있으며 이때 두 변수 사이의 관계 강도를 상관관계라 한다[15]. 본 논문에서 기계학습 모델에 적합한 학습데이터 선정을 위해 2가지 상관분석을 이용한다. 첫 번째 상관분석은 피어슨 상관계수(Pearson correlation coefficient)를 이용하는 것이고, 두 번째 상관분석은 스피어만 상관계수(Spearman correlation coefficient)를 이용하는 것이다. 수집한 84종류의 비트코인 블록 및 트랜잭션 통계데이터에 상관분석을 적용하여 학습데이터를 선정한다.

2.1 피어슨 상관계수

피어슨 상관계수는 두 변수 사이의 선형 상관관계를 계량화한 수치이다[16]. 피어슨 상관계수는 -1부터 +1 사이의 값을 갖고 수식 1에 의해 값을 구한다. p는 피어슨 상관계수, B는 블록 통계데이터, T는 트랜잭션 수, n은 데이터의 수, S는 표준편차, bar는 표본 평균이다. 피어슨 상관계수의 값이 -1인 경우 완벽한 음의 선형적 관계, +1인 경우 완벽한 양의 선형적 관계를 의미한다.

$$p = \frac{\sum_{i=1}^n (B_i - \bar{B})(T_i - \bar{T})}{(n-1)S_B S_T} \quad (1)$$

2.2 스피어만 상관계수

스피어만 상관계수는 변수의 값 대신 순위를 이용하는 경우의 상관계수로서, 값을 작은 것부터 차례로 순위를 매겨 서열 순서로 바꾼 뒤 순위를 이용해 값을 구한다[17]. 스피어만 상관계수는 수식 2에 의해 값을 구한다. p는 스피어만 상관계수, n은 데이터의 수, Bi는 블록 통계데이터에서 i번째 데이터 순위, Ti는 트랜잭션 수에서 i번째 데이터 순위, bar는 표본 평균이다. 스피어만 상관계수의 값은 -1과 +1 사이의 값을 갖는다. -1인 경우 두 변수의 순위가 완전히 반대이고, +1인 경우 두 변수의 순위가 완벽히 일치함을 의미한다.

$$p = \frac{\sum_{i=1}^n (B_i - \bar{B})(T_i - \bar{T})}{\sqrt{\sum_{i=1}^n (B_i - \bar{B})^2} \sqrt{\sum_{i=1}^n (T_i - \bar{T})^2}} \quad (2)$$

2.3 상관분석을 적용한 학습데이터 선정

초기 비트코인 블록은 비트코인 네트워크가 본격적으로 활성화되기 이전의 단순한 데이터만 포함하고 있다. 따라서 100,001 높이의 블록부터, 200,000 높이의 비트코인 블록데이터를 사용한다. 두 가지 상관분석을 통해 선정된 학습데이터는 그림 6, 7과 같다. 트랜잭션 수 예측을 위해 n번째 블록의 84종류의 비트코인 블록 및 트랜잭션 통계데이터와 n+1번째 비트코인 트랜잭션 수와의 상관분석을 한 뒤, 상관계수 값의 크기에 따라 오름차순으로 나열한다. 학습데이터에 따른 기계학습 모델의 성능 비교를 위해 피어슨 상관분석의 경우 양의 선형적 관계가 있는 데이터 3개와 선형적 관계가 없는 데이터 3개를 선정한다. 84종류의 비트코인 블록 및 트랜잭션 통계데이터 중 트랜잭션 수와 음의 선형적 관계를 나타내는 데이터는 없다. 스피어만 상관분석의 경우 -1 또는 +1에 가까운 4개의 데이터와 상관관계가 없는 데이터 3개를 선정한다. 두 가지 상관분석을 통한 상관계수의 값이 큰 상위 3개의 항목은 sum.nVout, sum.Vin, sum.vsize로 일치한다. 그러나 피어슨 상관분석의 경우 sum.nVin의 상관계수 값이 가장 크고, 스피어만 상관분석의 경우 sum.vsize의 상관계수 값이 가장 크다.

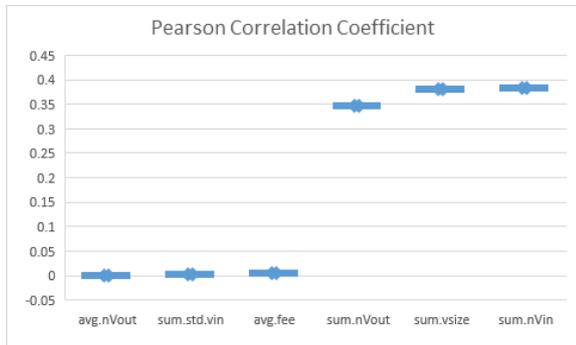


그림 6. 피어슨 상관분석을 적용한 학습데이터 선정항목
Fig. 6. Selected Data from the Pearson Analysis

3. 실험데이터

두 가지 상관분석을 적용해 선정된 학습데이터 선정항목에 대해 실험데이터로 구성한다. 실험데이터 구성에 대한 정보는 표 3과 같다. 기계학습 모델의 학습을 위한 학습데이터와 모델의 성능을 평가하기 위한 검증 및 시험데이터로 구성한다. 100,001 높이의 블록부터 180,000 높이의 블록은 학습데이터, 180,001 높이의 블록부터 190,000 높이의 블록은 검증 데이터, 190,001 높이의 블록부터

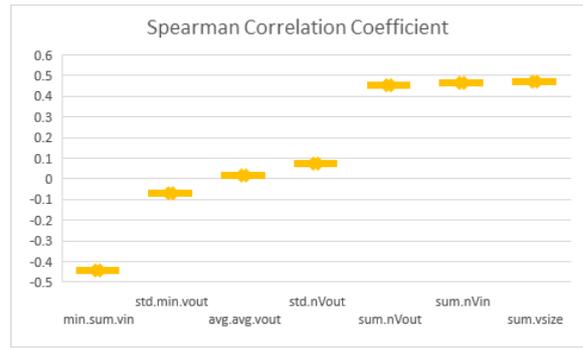


그림 7. 스피어만 상관분석을 적용한 학습데이터 선정항목
Fig. 7. Selected Data from the Spearman Analysis

200,000 높이의 블록은 시험데이터가 된다. 학습데이터를 통해 학습한 모델은 학습에 사용되지 않은 검증 및 시험데이터를 통해 성능을 평가한다.

4. 기계학습 모델 선정

본 논문에서 두 가지 기계학습 모델을 선정한다.

표 3. 실험데이터 구성

Table 3. Set of the Experimental Data

블록의 높이	구분
100,000~180,000	학습데이터
180,001~190,000	검증데이터
190,001~200,000	시험데이터

기계학습 알고리즘 중 가장 단순한 알고리즘이 적용된 선형회귀 모델과 시계열 데이터예측에 특화된 LSTM 모델이다. 두 가지 모델을 사용하는 이유는 같은 학습데이터를 사용하는 것으로 인해 선형회귀 모델과 LSTM 모델의 성능을 비교할 수 있고, 상관분석을 통해 선정한 데이터 항목에 대한 각 모델의 성능 비교를 하기 위함이다. 두 가지 모델의 차이점은 표 4와 같다. 선형회귀 모델의 경우 학습할 때, 입력 데이터로 비트코인 블록 및 트랜잭션의 통계 데이터만을 사용하여 비트코인 트랜잭션 수 예측을 한다. 그러나 LSTM 모델의 경우 학습할 때, 비트코인 블록 및 트랜잭션 통계데이터와 트랜잭션 수 모두를 입력 데이터로 사용하여 예측한다.

표 4. 기계학습 모델의 입력/출력

Table 4. Input/Output of Machine Learning Model

기계학습 모델	입력	출력
선형회귀 모델	비트코인 블록 및 트랜잭션 통계데이터	트랜잭션 수
LSTM 모델	비트코인 블록 및 트랜잭션 통계데이터, 트랜잭션 수	트랜잭션 수

5. Hyper-parameter Optimization

Hyper-parameter는 기계학습 모델을 구성에서 임의로 설정할 수 있는 요소이다. 모델별 Hyper-parameter에 대한 정보는 표 5와 같다. 선형 회귀 모델의 경우 Neural Network의 수, 모델의 성능을 평가하는 Loss Function, Loss Function을 줄이기 위한 함수인 Optimize Function이 있다. LSTM 모델의 경우 1회 학습에 사용할 데이터의 길이인 Sequence Length, LSTM Cell에 존재하는 Hidden Unit의 수를 추가로 설정한다. 학습된 모델의 성능을 평가하기 위해 생성한 검증 및 시험데이터에 대한 Loss Function을 적용한 값을 분석하여 Hyper-parameter Optimization을 한다.

표 5. 기계학습 모델의 Hyper-parameter 항목
Table 5. Hyper-parameter of Machine Learning Model

	선형회귀 모델	LSTM 모델
Hyper parameter 항목	Number of Neural Networks	Number of Neural Network
	Loss Function	Loss Function
	Optimize Function	Optimize Function
	Sequence Length	Sequence Length
	Number of Hidden Units	Number of Hidden Units

IV. 실험 및 결과

상관분석을 적용한 실험데이터를 사용하여 실험을 진행한다. 데이터 항목별 성능을 비교한 뒤, 가장 좋은 성능을 보이는 데이터 항목 및 기계학습 모델을 사용하여 Hyper-parameter Optimization을 진행한다. 예측모델의 성능 평가를 위해 각 모델의 Loss Function은 Mean Square Error(MSE)를 사용한다. 선형회귀 모델과 LSTM 모델의 데이터 항목별 실험 결과는 표 6, 7과 같다. 표 6의 경우 피어슨 상관분석을 통해 선정된 데이터를 사용한 결과이고, 표 7의 경우 스피어만 상관분석을 통해 선정된 데이터를 사용한 결과이다.

피어슨 상관분석을 적용한 경우 두 가지 모델 모두 비트코인 트랜잭션 수와의 피어슨 상관계수의 값이 가장 큰 sum.nVin을 사용한 경우 가장 좋은 성능을 보인다. 스피어만 상관분석을 적용한 경우

표 6. 피어슨 상관분석을 적용한 실험 결과
Table 6. Performance of Machine Learning Model with the Pearson Correlation Analysis.

데이터 항목	실 험	선 형 회 귀	LSTM 모
	단계	모델 MSE	델 MSE
avg.nVout	검증	0.0201	0.0219
	시험	0.0227	0.0251
sum.std.vin	검증	0.0200	0.0204
	시험	0.0226	0.0235
avg.fee	검증	0.0199	0.0219
	시험	0.0225	0.0251
sum.nVout	검증	0.0177	0.0130
	시험	0.0201	0.0154
sum.vsize	검증	0.0164	0.0130
	시험	0.0188	0.0154
sum.nVin	검증	0.0163	0.0130
	시험	0.0187	0.0153

표 7. 스피어만 상관분석을 적용한 실험 결과
Table 7. Performance of Machine Learning Model with the Spearman Correlation Analysis.

데이터 항목	실 험	선 형 회 귀	LSTM 모
	단계	모델 MSE	델 MSE
min.sum.vin	검증	0.0199	0.0174
	시험	0.0225	0.0202
std.min.vout	검증	0.0200	0.0133
	시험	0.0226	0.0156
avg.avg.vout	검증	0.0200	0.0134
	시험	0.0226	0.0157
std.nVout	검증	0.0201	0.0139
	시험	0.0227	0.0163
sum.nVout	검증	0.0177	0.0130
	시험	0.0201	0.0154
sum.nVin	검증	0.0163	0.0130
	시험	0.0187	0.0153
sum.vsize	검증	0.0164	0.0130
	시험	0.0188	0.0154

보편적으로 스피어만 상관계수의 값이 큰 데이터 항목 좋은 성능을 나타내지만, 스피어만 상관계수가 가장 큰 sum.vsize보다 sum.nVin을 사용했을 때, 더 좋은 성능을 보인다. 또한, LSTM 모델을 사용한 경우 선형회귀 모델을 사용한 경우보다 더 좋은 성능을 보인다. 따라서 비트코인 트랜잭션 수 예측을 위한 데이터 분석방법은 피어슨 상관분석 방법이 적합하고, 선형회귀 모델보다는 LSTM 모델을 사용하는 것이 적합하다.

첫 번째 실험을 통해 비트코인 트랜잭션 수를 예측하는데 가장 좋은 성능을 나타내는 데이터 항목인 sum.nVin을 학습데이터로 사용하여 LSTM 모델의 Hyper-parameter Optimization을 위한 실험을 한다. LSTM 모델의 Sequence Length를 1로 설정하

고, Hidden Unit의 개수를 증가시키며 모델을 학습하여 적합한 Hidden Unit의 개수를 구한 뒤, Sequence Length를 증가시켜 모델을 학습하여 적합한 Sequence Length를 구한다. 실험 결과는 표 8, 9와 같다.

표 8. Hidden Unit의 수 최적화 실험 결과
Table 8. Experimental Results for Optimizing the Number of Hidden Units

Hidden Unit의 수	실험 단계	MSE
1	검증	0.0130
	시험	0.0153
2	검증	0.0127
	시험	0.0151
4	검증	0.0124
	시험	0.0149
8	검증	0.0123
	시험	0.0148
16	검증	0.0123
	시험	0.0148
32	검증	0.0123
	시험	0.0148

표 9. Sequence Length 최적화 실험 결과
Table 9. Experimental Results for Optimizing the Sequence Length

Sequence Length	실험 단계	MSE
1	검증	0.0123
	시험	0.0148
2	검증	0.0115
	시험	0.0135
4	검증	0.0109
	시험	0.0122
8	검증	0.0105
	시험	0.0112
16	검증	0.0103
	시험	0.0111
32	검증	0.0102
	시험	0.0110
64	검증	0.0102
	시험	0.0110

Hidden Unit의 수의 경우 8개까지는 MSE 값이 줄어들지만, 8개를 초과하는 개수를 설정해도 값이 줄지 않는다. 따라서 가장 적합한 Hidden Unit의 수는 8개이다. Hidden Unit의 수를 8개인 상태에서 Sequence Length를 늘리는 경우 32까지는 MSE 값이 줄어들지만, 32를 초과해도 값이 줄지 않는다. 따라서 적합한 Sequence Length는 32이다. 두 개의 실험에서의 MSE 값의 변화를 보았을 때, Sequence

Length가 Hidden Unit의 수보다 LSTM 모델의 성능에 더 영향을 주는 parameter임을 알 수 있다.

V. 결론 및 향후 연구

본 논문은 비트코인 네트워크를 유지 및 성장시키는 데 중요한 비트코인 트랜잭션 수를 예측하는 방법으로 두 가지 기계학습 방법을 적용하여 실험하였고, 실험을 통해 적합한 기계학습 모델을 설계하였다. 향후 연구로는 Stacked-LSTM, Stateful-LSTM 모델을 설계하여 현재 모델보다 성능이 더 좋은 비트코인 트랜잭션 수 예측을 하는 모델을 설계할 계획이다.

References

- [1] NAKAMOTO, Satoshi, et al. Bitcoin: A peer-to-peer electronic cash system. 2008.
- [2] Blockchain.com, last Modified June 17, 2019, accessed June 17, 2019, <https://www.blockchain.com/ko/charts/n-transactions>.
- [3] Gabriel Bianconi, Mahesh Agrawal, Predicting Bitcoin Transactions with Network Analysis, snap.stanford.edu, last modified Sep 10, 2018, accessed May 20, 2019, <https://snap.stanford.edu/class/cs224w-2017/projects/cs224w-65-final.pdf>.
- [4] 지세현; 심규석; 김명섭. 효율적인 통신망 관리를 위한 기계학습 알고리즘을 적용한 트래픽 발생량 예측. 한국통신학회 학술대회논문집, 2018, 824-825.
- [5] 지세현, 백의준, 신무곤, 김명섭, “합성곱 신경망 기반 웹 응용 트래픽 분류 모델 설계”, *The Journal of Korean Institute of Communications and Information Sciences(J-KICS)*, accepted April 21, 2019.
- [6] “선형 회귀” Wikipedia, last modified April, 19, 2019, accessed June, 17, 2019, https://ko.wikipedia.org/wiki/%EC%84%A0%ED%98%95_%ED%9A%8C%EA%B7%80.
- [7] GREAVES, Alex; AU, Benjamin. Using the bitcoin transaction graph to predict the price of bitcoin. No Data, 2015.
- [8] NAKANO, Masafumi; TAKAHASHI, Akihiko; TAKAHASHI, Soichiro. Bitcoin technical trading

with an artificial neural network. *Physica A: Statistical Mechanics and its Applications*, 2018, 510: 587-609.

- [9] "Recurrent neural network" Wikipedia, last modified May 03, 2019, accessed May 15, 2019, https://en.wikipedia.org/wiki/Recurrent_neural_network.
- [10] KODAMA, Osamu; PICHL, Lukáš; KAIZOJI, Taisei. Regime change and trend prediction for Bitcoin time series data. In: *CBU International Conference Proceedings*. 2017. p. 384-388.
- [11] PANT, Dibakar Raj, et al. Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis. In: *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*. IEEE, 2018. p. 128-132.
- [12] "Long short-term memory" Wikipedia, last modified June 7, 2019, accessed June 17, 2019, https://en.wikipedia.org/wiki/Long_short-term_memory.
- [13] KARAKOYUN, E. S.; CIBIKDIKEN, A. O. Comparison of ARIMA Time Series Model and LSTM Deep Learning Algorithm for Bitcoin Price Forecasting. In: *The 13th Multidisciplinary Academic Conference in Prague 2018 (The 13th MAC 2018)*. 2018. p. 171-180.
- [14] PICHL, Lukáš; KAIZOJI, Taisei. Volatility analysis of bitcoin. *Quantitative Finance and Economics*, 2017, 1: 474-485.
- [15] "상관 분석", Wikipedia, last modified May 11, 2019, last accessed June 17, 2019, https://ko.wikipedia.org/wiki/%EC%83%81%EA%B4%80_%EB%B6%84%EC%84%9D.
- [16] "Pearson correlation coefficient" Wikipedia, last modified April 30, 2019, accessed May 15, 2019, https://en.wikipedia.org/wiki/Pearson_correlation_coefficient#cite_note-1.
- [17] "Spearman's rank correlation coefficient" Wikipedia, last modified April 2, accessed May 15, 2019, https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient.

지 세 현 (Se-Hyun Ji)



2018년 : 고려대학교 컴퓨터정보학과 학사
 2018년~현재 : 고려대학교 컴퓨터정보학과 석사과정
 <관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석
 <관심분야> 전자공학, 통신공학, 광통신 공학

백 의 준 (Ui-Jun Beak)



2018년 : 고려대학교 컴퓨터정보학과 학사
 2018년~현재 : 고려대학교 컴퓨터정보학과 석사과정
 <관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석

신 무 곤 (Mu-Gon Shin)



2019년 : 고려대학교 컴퓨터정보학과 학사
 2019년~현재 : 고려대학교 컴퓨터정보학과 석사과정
 <관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석

박 준 상 (Jun-Sang Park)



2008년 : 고려대학교 컴퓨터정보학과 학사
 2011년 : 고려대학교 컴퓨터정보학과 석사
 2014년 : 고려대학교 컴퓨터정보학과 박사
 2015년 ~ 현재 : LG

Electronics, CTO
 <관심분야> 네트워크 관리 및 보안, 트래픽 분석

김 명 섭 (Myung-Sup Kim)



1998년 : 포항공과대학교 전자
계산 학과 학사

2000년 : 포항공과대학교 전자
계산 학과 석사

2004년 : 포항공과대학교 전자
계산 학과 박사

2006년 : Dept. of ECS, Univ
of Toronto Canada

2006년~현재 : 고려대학교 컴퓨터정보학과 교수

<관심분야> 네트워크 관리 및 보안, 트래픽 모니터
링 및 분석, 멀티미디어 네트워크