합성 곱 신경망 기반 비트코인 블록 단위 트랜잭션 수 증감 예측

지세현, 백의준, 김보선, 강민규, 김명섭

고려대학교

{sxzer, pb1069, boseon12, cxz361, tmskim}@korea.ac.kr

Prediction of increase/decrease in the number of transactions per bitcoin blokc based on convolution neural network

Se-Hyun Ji, Ui-Jun Baek, Boseon Kim, Mingyu Gang, Myung-Sup Kim

Korea Univ.

요 약

2009년 비트코인이 발행된 이후 비트코인은 모두의 관심을 끌고 있다. 비트코인이 활발하게 채굴되고 거래에 쓰이면서 비트코인 네트워크는 발전하고 있다. 하지만 비트코인 네트워크가 발전함에 따라 부작용도 발생하고 있다. 비트코인 트랜잭션 수를 예측하는 것은 비트코인 네트워크가 안전하게 유지하고, 성장하는 데 있어 필수적이다. 본 논문은 기계학습 알고리즘 중 하나인 합성 곱 신경망을 이용한 비트코인 블록 단위 트랜잭션 수의 증감을 예측하는 방법을 제안한다. 합성 곱 신경망 알고리즘을 사용하면 비트코인 데이터 특징 중 하나인 시계열 특징을 고려하지 않지만, 비트코인 블록 및 트랜잭션 데이터의 특징을 합성 곱 신경망의 관점에서 분류를 통한 예측 가능성을 검토하고, 예측정확도라는 객관적인 지표를 보여준다. 제안하는 방법은 데이터 분석, 데이터 전처리, 합성 곱 신경망 모델 구성, 학습 및 검증 실험을 통해 적합성을 검증한다.

Ⅰ. 서론

2009년 사토시 나카모토에 의해 개발된 비트코인은 p2p 방식을 사용하는 최초의 온라인 암호화폐이다. 비트코인은 정부, 금융기관의 개입 없이 개인 간의 빠르고 안전한 거래를 하기 위한 목적으로 발행되었다. 오늘날 비트코인은 개인, 정부 기업, 금융기관 등 모두의 관심을 끌고 있다.

비트코인이 최초 발행된 이후 일 단위로 확인된 비트코인 트랜잭션 수는 갈수록 증가하는 추세를 보인다. 비트코인 트랜잭션 수의 증가와 함께 비트코인 네트워크는 급속도로 발전하고 있다. 그러나 비트코인 네트워크가 급속도로 발전하고 있지만, 그에 따른 부작용이 발생한다. 예를 들자면, 비트코인 트랜잭션을 처리하는 비용은 증가했지만, 트랜잭션이 처리되는 시간은 지연되고 있다[1]. 이러한 부작용을 방지하기 위해 비트코인 트랜잭션 수를 예측하는 것은 필수적이다.

예측하는 데 있어 주로 사용되는 기계학습 알고리즘은 순환신경망 혹은 LSTM(Long Short Term Memory) 모델을 사용하는 것이다. 그러나 RNN, LSTM 알고리즘의 성능은 실제 값과 예측값의 상대적인 오차를 통해서 결정되기 때문에 객관적으로 성능을 평가하기 어렵다. 객관적으로 성능을 평가하기 위해 본 논문은 기계학습 알고리즘 중 하나인 합성 곱 신경망 모델을 설계하여 비트코인 블록에 담긴 트랜잭션 수의 증감을 예측하는 방법을 제안한다. 본 연구팀은 비트코인 코어로부터 비트코인 블록 및 트랜잭션 원시데이터를 수집한 뒤, 블록 및 트랜잭션 단위로 구분하여 약 300종류의 비트코인 블록 및 트랜잭션 통계데이터를 수집했다. 수집한 데이터를 합성 곱 신경망 모델의 학습데이터로 만들기 위한 전처리 과정을 거친 뒤, 합성 곱 신경망을 설계하여 실험을 진행한다.

본 논문에서는 서론에 이어 2장에서 제안하는 방법론을 설명한다. 3장에서 실험을 통해 합성 곱 신경망 모델의 성능을 보여주고, 마지막 5장에서 결론 및 향후 연구를 언급한 뒤 논문을 마친다.

※ 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)

Ⅱ. 본론

본 장에서 합성 곱 신경망 기반 비트코인 블록 단위 트랜잭션 수 증감 예측방법에 대해 설명한다. 비트코인 원시데이터 수집 및 통계처리 및 분석, 데이터 라벨링 및 분석을 통해 학습 및 검증데이터를 완성한다. 합성 곱 신경망은 학습데이터를 통해 학습한 뒤, 성능 평가를 위해 검증데이터를 이용해 합성 곱 신경망 모델의 성능을 평가한다.

비트코인 코어로부터 비트코인 원시데이터를 수집한다. 수집한 비트코인 원시데이터에 대한 자세한 설명은 표 1과 같다[2]. 비트코인 블록 및 트랜잭션으로부터 12가지 원시데이터를 수집한다.

표 1. 비트코인 원시 데이터

데이터 명	의미
ntx	블록에 포함된 트랜잭션 수
Weight	블록의 Weight
Size	블록의 크기
vSize	블록의 가상크기
nVin	트랜잭션이 포함하고 있는 입력 수
nVout	트랜잭션이 포함하고 있는 출력 수
Value	트랜잭션의 거래 금액
Fee	트랜잭션 수수료
Tx.Size	트랜잭션의 크기
Tx.Vsize	트랜잭션의 가상크기
Vin.Value	트랜잭션의 입력 값
Vout.Value	트랜잭션의 출력 값

표 1로부터 수집한 비트코인 원시데이터로부터 통계정보를 얻기 위해 통계처리를 한다. 통계처리를 통해 수집한 데이터에 대한 자세한 설명은 표 2와 같다. 트랜잭션 단위의 데이터 6가지 항목은 한 번의 통계처리를 하고, 2가지 항목은 두 번의 통계처리를 한다. 총 312종류의 비트코인 블록 및 트랜잭션 단위의 통계데이터로 구성한다.

표 2 비트코인 통계데이터

데이터 단위	데이터 항목	1st 통계처리	2nd 통계처리	항목 수
블록	nTx			1
	Weight			1
	Size			1
	vSize			1
트랜잭션	nVin		통계정보	11
	nVout			11
	Value			11
	Fee			11
	Tx.Size			11
	Tv.vSize			11
	Vin.Value	통계정보		121
	Vout.Value			121

통계데이터를 합성 곱 신경망 모델의 학습 및 검증데이터로 생성하기 위해 데이터 분석을 한 후 데이터 라벨링을 한다. 데이터 분석을 한 결과는 표 3과 같고, 데이터 라벨링에 대한 정보는 표 4와 같다. 표3의 정보를 기준으로 증가, 유지, 감소 그룹으로 데이터 라벨링을 한다. 증가 그룹의 경우 다음번 블록에 담긴 트랜잭션 수가 현재 블록에 담긴 트랜잭션 수보다 50개를 초과하여 증가하는 경우이고, 유지 그룹의 경우 다음번 블록에 담긴 트랜잭션 수의 변화가 50개 이하인 경우이고, 감소 그룹의 경우는 그 이외의 다음번 블록의 담긴 트랜잭션 수가 현재 블록의 담긴 수보다 50개를 초과하여 감소하는 경우이다. 트랜잭션 변화량을 50개를 기준으로 분류하는 50개 이하 트랜잭션 수의 변화는 비트코인 블록의 특징에 큰 변화가 없기 때문이다.

표 3. 비트코인 블록 단위 트랜잭션 수 분석 결과

항목	값
비트코인 블록의 높이	500,000 ~ 600,000
최대 트랜잭션 수	4,225
최소 트랜잭션 수	1
평균 트랜잭션 수	1,827

표 4. 비트코인 블록 단위 통계데이터 그룹 정보

항목	개수
증가 그룹의 수	48,408
유지 그룹의 수	5,358
감소 그룹의 수	46,234

증가, 유지, 감소 그룹의 비트코인 통계데이터 특징들에 대해 상관분석을 한다. 상관분석은 각 통계데이터 특징들에 대한 분포를 파악한 뒤, 정규분포를 갖는 특징들에 대해서만 상관분석을 한다. 분석을 통해 비트코인 통계데이터의 특징을 선택한다. 총 312종류의 통계데이터 특징 중 정규분포의 형태를 보이는 데이터는 68개이다. 68개의 정규분포의 형태를 보이는 데이터를 대상으로 2가지 경우의 피어슨, 스피어만 상관분석을 실시한다.

68개의 비트코인 통계데이터 특징 중 합성 곱 신경망의 실험데이터로 만들기 위해 상관계수의 값이 0에 가까운 64개의 통계데이터를 내림차순으로 추출하여 이미지화를 한다. 이미지는 0~255 사이 정수형태의 음영으로 구성되기 때문에 통계데이터 특징값을 0~255 사이의 정수로 정규화하는 과정을 통해 완성한다. 완성된 합성 곱 신경망의 실험데이터는 그림 1과 같다. 유지 그룹 개수와 같은 학습데이터 개수를 생성하기 위해 그룹별 약 5,000개씩의 이미지를 추출한다. 생성된 실험데이터 중 3,500개는 학습데이터, 1,500개는 검증데이터로 구분한다. 완성된 실험데이터는 [3]의 합성 곱 신경망 모델을 참조하여 합성 곱 신경망을 구성한 뒤 학습 및 검증 실험을 진행한다. 실험에 사용된 합성 곱 신경망의 정보는 표 5와 같다.

증가그룹 유지그룹 감소그룹

그림 1 비트코인 통계데이터 이미지

표 5. 합성 곱 신경망 정보

Hyper-parameter 종류	정보
Convolution Activation Function	Leaky_ReLU
Neural Network Activation Function	Leaky_ReLU
Convolution Filter Size	3x3
Pooling Method	Max Pooling
Padding Option	Valid
Number of Neural Network	1024

Ⅲ. 실험 및 결과

제안하는 방법의 적합성을 검증하기 위해 실험을 진행한다. 2가지 상관분석에 대해 구성된 실험데이터를 이용해 학습한 뒤, 학습에 사용되지 않은 데이터를 통해 검증 실험을 한다. 실험 결과는 표 6과 같다. 두 가지 경우 모두 학습정확도는 100% 가까운 성능을 보이며 제안하는 방법을 통해 만들어진 실험데이터는 합성 곱 신경망의 학습데이터로 적합하다는 것을 알 수 있다. 검증 실험을 통해 피어슨 상관분석을 통해 만들어진 실험데이터가 스피어만 상관분석을 통해 만들어진 것보다 약 2% 높은 예측성능을 보이는 것을 알 수 있다.

표 6. 학습 및 검증 실험 결과

상관분석 방법	학습정확도	검증정확도
피어슨 상관분석	99.7%	95.8%
스피어만 상관분석	99.3%	93.3%

Ⅳ. 결론 및 향후 연구

본 논문은 비트코인 합성곱 신경망 기반 블록 단위 트랜잭션 수 증감을 예측하는 방법을 제안하였고, 실험을 통해 제안하는 방법의 적합성을 검증하였다. 향후 연구로는 더 정교한 분석을 통해 현재 모델보다 성능이 더 좋은 예측 모델을 설계할 계획이다.

참 고 문 헌

[1] Gabriel Bianconi, Mahesh Agrawal, Prediction Bitcoin Transactions with Network Analysis, snap.stanford.edu, last modified Sep 10, 2018, accessed January, 7, 2021, https://snap.stanford.edu/class/cs224w-2017/projects/cs224w-65-final.pdf

[2] 백의준, et al. "비트코인 네트워크 트랜잭션 이상 탐지를 위한 특징선택 방법." 기계학습 기반의 가상 네트워크 기능 자원 수요 예측방법: 18.

[3] 지세현, et al. "합성 곱 신경망 기반 웹 응용 트래픽 분류 모델 설계." 한국통신학회논문지 44.6 (2019): 1113-1120.