Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.29, NO.1, Feb. 2019 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2019.29.1.225

기계학습 기반 비트코인 채굴 난이도 예측 연구*

이 준 원,† 권 태 경‡
연세대학교 정보보호연구실

Machine Learning Based Prediction of Bitcoin Mining Difficulty*

Joon-won Lee,† Taekyoung Kwon‡
Information Security Lab, Graduation School of Information, Yonsei University

요 약

비트코인은 탈중앙화와 분산원장을 특징으로 하는 암호화폐로서 "작업증명"이라는 채굴시스템을 통해 유지된다. 채굴 시스템에서는 블록 생성시간을 일정하게 유지하기 위해 채굴 난이도를 조정하게 되는데, 기존의 채굴 난이도 변경 방식은 미래의 해시파워를 반영할 수 없다는 문제가 있다. 따라서 실제시간과 예정시간 사이에 발생하는 오차로 인해 블록생성과 실세계 시간의 불일치를 가중시키게 되고, 결국 거래 기한을 맞추지 못하거나 코인 호핑 공격에 취약점을 노출시키게 된다. 블록 생성시간을 일정하게 유지시키기 위한 기존 연구도 여전히 오차 문제를 갖는다. 본 연구에서는 이러한 오차를 줄이기 위한 기계학습 기반 채굴 난이도 예측 방안을 제시한다. 이전 해시파워를 학습하여 미래의 해시파워를 예측하고 예측한 값을 이용하여 채굴 난이도를 조정한다. 우리의 실험 결과는 이와 같은 경우 기존 채굴 난이도 조정방식보다 오차율을 약 36% 더 줄일 수 있음을 보여준다.

ABSTRACT

Bitcoin is a cryptocurrency with characteristics such as de-centralization and distributed ledger, and these features are maintained through a mining system called "proof of work". In the mining system, mining difficulty is adjusted to keep the block generation time constant. However, Bitcoin's current method to update mining difficulty does not reflect the future hash power, so the block generation time can not be kept constant and the error occurs between designed time and real time. This increases the inconsistency between block generation and real world and causes problems such as not meeting deadlines of transaction and exposing the vulnerability to coin-hopping attack. Previous studies to keep the block generation time constant still have the error. In this paper, we propose a machine-learning based method to reduce the error. By training with the previous hash power, we predict the future hash power and adjust the mining difficulty. Our experimental result shows that the error rate can be reduced by about 36% compared with the current method.

Keywords: Bitcoin, Mining difficulty, Time-series analysis, Predictive model, Machine learning

I. 서 론

비트코인[1]은 2008년 S.Nakamoto에 의해 개발된 암호화폐로서 신뢰할 수 있는 중계자의 개입 없이 개인 간에 화폐를 안전하게 주고받을 수 있는 시스템이다. 비트코인 시스템을 이용하면 정부나 은

Received(11. 21. 2018), Accepted(11. 23. 2018)
* 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음(IITP-2018-2016-0-00304). 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No. 2017-0-00380, 차세대 인증 기술 개발)
† 주저자, withjoon@yonsei.ac.kr
‡ 교신저자, taekyoung@yonsei.ac.kr(Corresponding author)

행의 개입 없이도 안전한 금융거래가 가능하다. 이는 비트코인이 가지고 있는 탈중앙화(de-centralization)와 분산원장이라는 특징에 기인한 것으로, 비트코인에서는 특히 이 특징들의 유지를 위해 채굴 시스템을 도입하였다. 비트코인에서 채굴이란 부분 해시 충돌[2]을 찾는 작업증명(proof of work) 과정이며, 먼저 채굴한 채굴자만이 생성된 블록을 블록체인에 추가할 권한을 가지게 된다. 부분 해시 충돌을 찾기 위한 계산능력을 해시파워(hash power) 또는 마이닝 파워(mining power)라고 부르는데 채굴에 투입되는 해시파워가 클수록 채굴시간은 빨라지게 된다. 2009년 비트코인 채굴이 시작된 이후 채굴에 투입되는 해시파워는 지속적으로 증가해 왔다.

블록이 생성되는 것은 비트코인 시스템에서 중요한 의미를 가진다. 각 블록은 다수의 금융거래 정보를 지니게 되며 이 블록이 순조롭게 생성되어야만 비트코인을 이용한 거래를 완료시킬 수 있다. 이러한 블록 생성은 채굴 과정을 통해서만 가능하기 때문에 채굴을 위해 투입되는 해시파워는 블록생성 시간에 영향을 주게 된다. 비트코인에서는 10분에 한 개 씩 블록을 생성하도록 설계되어 있으며 해시파워의 변동성에 대처하기 위해 채굴 난이도란 개념을 사용한다. 즉 2016개의 블록 생성 후 평균 블록생성 시간이 10분에 못 미치는 경우 해시파워가 증가하였다고 간주하여 그 비율만큼 채굴 난이도를 높여 채굴 속도를 조절한다.

이러한 방식은 단순하다는 장점이 있지만 미래에 투입될 해시파워를 예측할 수 없기 때문에 투입되는 해시파워가 지속적으로 변화하는 실제 상황에서 블록 생성시간을 일정하게 유지시킬 수 없게 된다. 이렇게 블록 생성시간을 일정하게 유지하지 못할 경우 비트코인을 이용한 거래 시 기한 문제가 발생하게 된다. 즉 블록 생성의 예정시간을 기준으로 거래 기한을 제한하였지만 실제 블록 생성시간이 빨라지거나 느려지면서 기한을 넘겨 손해를 끼칠 수 있게 된다. 또한 코인 호핑(coin-hopping) 공격[3]에 대한 취약점을 노출하게 된다. 이것은 악의적인 채굴자들이 두 개 이상의 코인들 중에서 코인 가격과 채굴 난이도를 비교하여 수익성이 높은 코인만을 번갈아 가면서 채굴하는 공격 방법인데 주기적인 해시파워의 가감을 채굴 난이도에 반영하지 못하게 되므로 악의적인 채굴자에게 더 많은 이익을 제공하게 된다.

이러한 비트코인의 기존 채굴 난이도 변경 방식의 문제점을 개선하기 위해 일부 연구들이 진행되었으나[4][6], 여전히 예정시간 간의 오차는 존재한다.

본 연구에서는 머신러닝 기법을 활용하여 미래에 투입될 해시파워를 예측하는 방식으로 채굴 난이도를 계산하는 모델을 제시하려고 한다. 아직까지 채굴 난이도 계산을 위해 기계학습을 활용하는 연구는 없었다. 본 연구를 통해 비트코인 채굴 난이도 계산하기 위해 기계학습을 적용하는 방법을 제시하고 최적의 방식을 찾아 기존 연구 결과와 비교하도록 한다.

본 논문의 구성은 6장으로 되어 있으며 각 장의 내용은 다음과 같다. 1장에서는 연구 배경 및 목표에 대하여 언급하고 2장에서는 기존 비트코인 채굴 난이도 방식이 야기하는 취약성에 대하여 설명한다. 3장에서는 비트코인 채굴 난이도 및 기계학습 관련 연구들을 살펴본다. 4장에서는 제안하는 실험 방법과 데이터 처리 및 관련 기술에 대하여 설명하고 5장에서 실험 결과에 대하여 기술한다. 마지막으로 6장을 통해 본 논문의 결론을 맺는다.

II. 해시파워와 블록 생성시간

2.1 비트코인 채굴 해시파워의 비선형적 증가

현재 사용되는 비트코인 채굴 난이도 조정 방식은 지수적으로 증가하는 해시파워에 대해 취약하다는 문제가 있다. 예를 들어 매일 해시파워가 10%씩 지속적으로 증가한다면 블록 생성시간은 평균 6.3분이 되어 결국 예정된 10분과 많은 차이를 보이게 된다[4].

실제로 2009년 1월 비트코인이 최초로 채굴된 이래 해시파워는 계속 증가되어 왔으며, 지수적 증가세를 보이는 구간들도 관찰되고 있다. 2010년부터 2018년까지 월별 평균 블록 생성시간을 보면 Fig.1과 같이 불규칙하며 대부분의 경우 10분 미만인 것

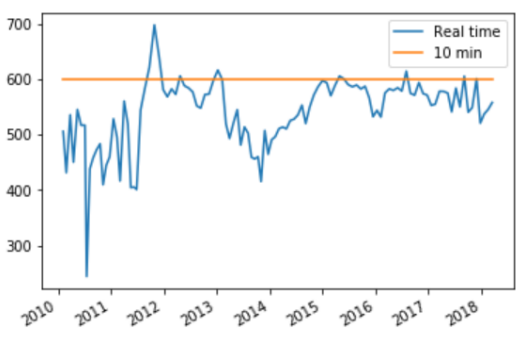


Fig. 1. Inter-Block time series graph

을 확인할 수 있다.

2.2 블록 생성시간이 영향을 미치는 요소

비트코인 시스템에서는 거래를 체결하고 가치를 보존하기 위해 여러 가지 활동을 수행한다. 이러한 활동들은 시간 기준이 아닌 생성된 블록의 수를 기준으로 활성화된다. 다음은 블록생성을 기준으로 활성화되는 비트코인 활동들이다.

거래 완료: 거래가 포함된 블록이 생성되었을 때 거래가 이루어진다. 하지만 해당 블록 생성 후, 5개의 블록이 추가로 생성되어야만 확률적으로 거래가 변경될 수 없음을 보장하고 거래가 확정되었다고 간주한다.

채굴 난이도 조정: 2016개의 블록이 생성될 때마다 블록 생성시간을 안정화시키기 위하여 채굴 난이도가 조정된다.

반감기: 채굴을 진행하는 채굴자에게 제공되는 보상은 210,000블록마다 절반으로 줄어든다. 초기 보상으로는 비트코인 50개씩 제공되었으나, 2018년 현재는 12.5개만 제공되고 있다.

합의: 비트코인 시스템 정책 변경은 민주적인 방식인 투표로 결정된다. 투표방식은 블록의 버전 정보에 찬성여부의 정보를 담아서 채굴을 진행하고, 채굴을 진행한 블록 내의 버전 정보를 통해 동의여부 비율로 결정한다[5]. 2017년 SegWit 적용시에도 투표를 통해 결정하였다.

잠금시간(locktime): 비트코인을 거래할 때 거래될 수 있는 가장 빠른 시간을 설정할 수 있다. 이 기능을 통해 거래 방법의 다양성을 제공해 준다. 잠금시간은 시간 뿐 아니라 블록의 높이로도 설정할 수 있게 되어 블록 생성시간에 영향을 받게 된다.

2.3 블록 생성시간의 불안정성

비트코인은 생성된 블록의 수를 기준으로 활동을 진행하도록 설계되어 있어 시간을 중심으로 스케줄이 정해지는 현실 세계와 차이가 발생할 수밖에 없다. 특히 금전적인 거래와 관련된 경우 현실세계에서 손실을 유발시킬 수 있다.

다음 예는 블록 생성시간이 일정하지 않을 때 현실세계에서 발생할 수 있는 문제를 보여준다: 먼저 Alice가 Bob에게 비트코인을 구매하기로 하고 은행을 통해 비용을 지불하기로 하였다. 은행 영업시간이 70분이 남은 상황에서 Bob은 Alice에게 비트코인을 보냈다. Alice는 확실한 거래를 위해 6블록이 생성되기를 기다렸으나 블록 생성시간이 예정보다 증가하여 70분을 넘어버렸고 은행 영업시간이 종료되어 돈을 보낼 수 없게 되었다. 블록생성시간이 예상보다 늦어지는 경우 대기시간이 길어지면서 현실 세계의 스케줄과 맞지 않아 문제가 발생한 경우이다.

반대로 예상보다 빨라지는 경우에도 문제가 발생할 수 있다. Bob은 비트코인 전송시 잠금시간을 6블록 후로 설정하여 미리 거래를 실행하고 Alice가 돈을 송금할 때까지 기다렸다. 하지만 50분을 기다렸으나 송금이 되지 않아 Alice에 연락하려고 하였으나 연락이 닿지 않았다. 이에 Bob은 거래를 취소하려고 보니 이미 6블록이 생성되어 취소를 할 수 없게 되었다.

두 경우 모두 간단한 예이지만 경우에 따라서 큰 손해를 입을 수도 있다. 비트코인 시스템을 통해 거래를 할 시에 상호간 이 같은 특성을 이해하고 조심해야 하겠지만, 채굴시간이 예정시간과 차이가 줄어든다면 해당 경우의 위험성은 줄어들 것이다.

2.4 코인 호핑 공격

코인 호핑 공격은 가상화폐 채굴 전략으로 2개 이상의 가상화폐의 가격과 난이도를 고려하여 수익성이 큰 가상화폐를 번갈아가면서(hopping) 채굴하는 공격 방법이다. 이 방법을 이용하면 정직하게 채굴하는 것보다 더 많은 수익을 올릴 수 있게 된다.[3][6] 더 큰 문제는 정직한 채굴자들의 수익이 감소하면서 채굴 의지를 낮춰 채굴 시스템이 붕괴될 수도 있다는 점이다.

이러한 코인 호핑 공격이 가능한 이유는 채굴난이도가 미래의 해시파워를 제대로 반영하지 못하기 때문이며, 따라서 주기적으로 발생되는 코인 호핑 공격에 대응할 수 있는 새로운 채굴난이도 변경 방식이 필요하다.

2.5 채굴 난이도 예측의 필요성

비트코인의 활동 단위인 블록 생성의 실제시간과 예정시간 간의 간극을 줄이기 위해서는 채굴에 투입되는 해시파워를 정확히 예측하여 채굴 난이도를 계산하여야 한다. 채굴 난이도가 예측을 통해 정확히 계산된다면 비트코인 시스템을 통한 거래시 현실 세

계의 시간 중심의 스케줄을 따라갈 수 있다. 또한 코인 호핑 공격이 주기적으로 발생하더라도 이를 예측하고 난이도를 조절해 줌으로써 정직한 채굴자들에게 수입을 보장해 주어 채굴 시스템을 안정적으로 유지시켜줄 수 있게 된다.

III. 관련 연구

3.1 채굴 시스템 공격

채굴 시스템 공격과 관련된 연구 중 대표적인 것은 블록체인에서 분기가 발생했을 때 유효한 분기를 선택하는 과정을 이용하여 이중지불이 가능하도록 하는 공격이다.

Rosenfeld, M는 공격자의 해시파워에 따른 이중지불 확률을 도출하여 다양한 방식으로 보여주고 6 컨펌(confirmation)이 안전하지 않을 수 있다는 사실을 보여주었다[16].

I. Eyal은 채굴된 블록들을 비공개로 채굴하다가 기존 블록들보다 더 길어질 때 공개하는 selfish mining 전략을 소개하였다. 이 전략을 이용하면 25%의 해시파워 만으로 이중지불이 가능하다는 것을 보여주었다[17].

Bahack, Lear는 충분히 긴 시간동안 정직한 채굴자보다 많은 블록을 생성하려고 시도하면 이중지불이 가능해지는 Difficulty Raising 공격을 소개하였다[18].

3.2 채굴 난이도 조정 알고리즘

채굴 시스템의 직접적인 공격 뿐 아니라 채굴 난이도 조정 방식의 문제점으로 인해 발생되는 블록 생성시간의 불안정성과 그로 인한 코인호핑 공격에 대해서도 연구되었다.

D. Kraft 등은 비트코인 채굴을 모델링하여 채굴을 시뮬레이션을 하고, 대안 알고리즘을 제안하였다[4]. 해시파워가 지수적 증가하는 경우에도 기존 알고리즘보다 거래시간의 안정성이 향상된다는 시뮬레이션 결과를 보여준다.

Meshkov 등은 채굴난이도 조정 알고리즘으로 간단한 선형 최소제곱법를 응용하여 계산하는 방법을 제안하였다[6]. 직전의 채굴 난이도들을 이용하여 현재 난이도를 계산하는 방식으로 블록 생성시간의 안정성과 함께 코인호핑 공격을 약화시킬 수 있었다.

또한 정직한 채굴자들이 많지 않은 알트코인들은 블록 생성시간의 안정화와 코인호핑 공격을 방지하기 위해 자체 난이도 조정알고리즘을 연구하여 사용하고 있다.[12][13]

3.3 비트코인 가격 예측

기계학습을 이용하여 비트코인 채굴 난이도를 예측하는 연구는 없었으나 비트코인 가격 예측에 대한 연구들은 지속적으로 발표되고 있다.

McNally 등은 비트코인 가격을 RNN, LSTM과 같은 기계학습을 통해 예측하고, 그 결과와 전통적인 시계열 분석기법인 ARIMA 모델과 비교하였다[11]. 학습에 사용한 데이터는 가격정보 뿐 아니라 블록체인 내에서 얻을 수 있는 채굴 난이도와 해시정보, 가격의 단순이동평균을 사용하였고, 기계학습 방법 중 LSTM의 정확도가 가장 높았다.

Madan 등은 이항형 로지스틱 회귀, 서포트 벡터 머신, 랜덤 포레스트 기법으로 비트코인 가격을 예측하였다[14].

위 두 연구들은 가격의 수치적 예측보다는 가격 변동에 대한 분류의 정확성을 예측하였다.

Jang, H 등은 베이지안 기계학습을 통해 비트코인 가격을 수치적으로 예측하였다[15]. 이때 학습에 사용된 데이터는 비트코인 블록체인 내부 정보 뿐 아니라 주식시장 및 환율 데이터 등 외부 데이터도 활용하였다.

IV. 채굴 난이도 예측

이 장에서는 채굴 난이도 예측을 위한 방법을 제안하고 그 실험설계 및 데이터 처리 방법을 설명한다. 4.1절은 실험 설계의 관점에서 제안 방법을 설명하고 4.2절과 4.3절은 데이터 수집 및 가공에 대하여 설명하며 4.4절과 4.5절은 적용하는 기계학습 방식과 채굴 시뮬레이션 방법에 대하여 설명한다.

4.1 실험 설계

채굴 난이도 예측을 위해 Fig. 2와 같은 실험 방법을 설계하고 제안한다. 먼저 비트코인 블록체인을 다운받아 분석하여 블록 별로 투입된 해시파워를 계산하고 자질(features) 세트를 추출한 후 시계열 예측에 적합하도록 블록단위 데이터에서 시간기반 데이

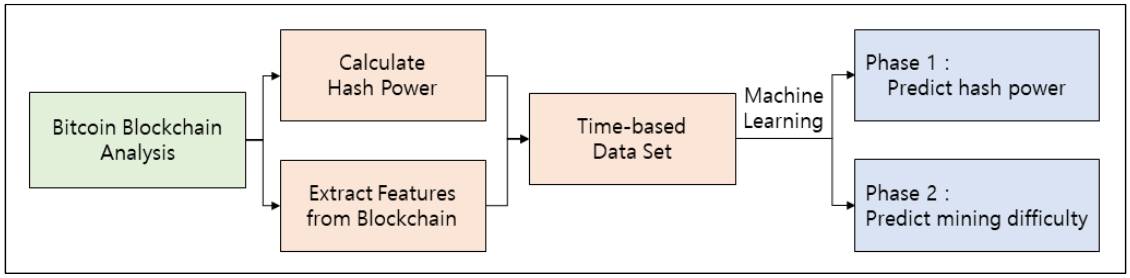


Fig. 2. Design of Experiments about Mining Difficulty Prediction

터로 가공한다. 가공된 데이터를 기반으로 본 연구에서는 두 가지 값에 대하여 예측을 수행한다.

첫 번째는 기계학습을 통해 단위시간 별 채굴에 투입되는 해시파워 예측한다. 기존 시계열 데이터 예측과 동일하게 단위시간 별로 예측하며 해시파워의 오차율을 측정한다. 이 단계에서는 실험을 통해 다양한 자질들과 기계학습 알고리즘들 중 해시파워 예측에 적합한 조합을 찾는 것을 목표로 한다.

두 번째는 첫 번째 단계에서 찾은 기계학습 조합을 이용하여 채굴 시뮬레이션을 통해 채굴 난이도를 예측한다. 이 단계에서는 2016블록이 생성되는 시간의 오차율을 측정하여 최적의 방법을 선별하고 기존 논문에서 소개하는 모델의 성능과 비교한다.

위 과정을 통해 채굴 난이도 예측에 적합한 데이터 형태와 기계학습 알고리즘을 찾고 기존 채굴 난이도 변경 방식뿐 아니라 기존 연구 방식과 성능 비교를 진행한다.

4.2 데이터 수집

비트코인 시스템은 다수의 채굴자들이 데이터를 주고 받으면서 동일한 블록체인을 유지해야 한다. 이를 위해서 외적요소가 아닌 블록체인 내의 데이터를 기반으로 채굴난이도가 계산되고 검증되어야 한다. 본 연구의 목표는 블록 생성시간을 안정화시킬 수 있는 채굴 난이도 조정 모델을 도출하는 것이므로, 채굴난이도 예측을 위해 사용하는 데이터는 비트코인 블록체인 내의 데이터로 한정한다.

비트코인은 2009년 1월 처음 발행된 이후 지속적으로 채굴이 진행되고 있다. 발행 첫 해는 비트코인의 초기 단계로 기계학습을 위한 데이터로 적합하지 않다고 판단되어, 2010년 1월부터 2018년 3월까지 8년간의 데이터를 대상으로 연구를 진행한다. 기간 내에 포함된 블록의 총 개수는 423,360개이다.

해당 블록들 내에서 정보를 분석하고 추출해 내기 위해 비트코인 프레임워크인 "Blocksci"[7]를 이용한다. 이 프레임워크는 다운로드 받은 블록체인을 분석하고 데이터베이스화하여 빠르게 데이터를 얻을 수 있도록 도와준다. 더욱이 파이썬 인터페이스를 제공하여 데이터를 쉽게 처리할 수 있다.

4.3 데이터 가공

4.3.1 블록별 해시파워 계산

채굴에 투입되는 해시파워를 직접적으로 확인할 수는 없다. 하지만 블록체인 내에서 제공하는 블록별 nbits 값과 타임스탬프(timestamp)를 이용하여 간접적으로 계산할 수 있다[8]. nbits 값에서 계산해 낸 채굴난이도와 블록간의 타임스탬프의 시간 차이를 이용하여 해시파워를 계산한다. 이때 타임스탬프로 제공되는 시간정보는 블록체인 내에서 블록간의 순서를 따르지 않는 경우가 발생할 수 있다. 다시 말해 하나의 블록이 이전 블록보다 더 빠르게 생성된 것처럼 보일 수 있다[9]. 이러한 상황을 고려하여 블록별로 해시파워를 계산하지 않고 6개씩 블록을 묶어 해시파워를 계산한다. 비트코인의 기본 설계에 따라 약 10분에 한 개의 블록이 추가되는 것을 고려할 때 6개의 연속된 블록은 1시간동안 추가된 블록에 해당한다.

4.3.2 자질세트 추출

비트코인 블록체인에는 다양한 정보가 담겨 있다. 이중 해시파워 예측을 위해 블록체인 내에서 다음과 같이 7가지 자질(feature)들을 추출하였다.

Transaction : 거래 횟수
Fee : 블록생성 수수료

Incentive : 블록 생성시 주어지는 보상
Output : 거래되는 비트코인 수
Size : 데이터 총 크기
Blocks : 생성된 블록 개수
Miners : 채굴에 성공한 채굴자(주소) 수

각각의 값들은 해시파워 데이터와 같이 6개 블록 단위를 기준으로 계산했다.

4.3.3 시계열 데이터

본 연구는 해시파워 변동을 시간 기반으로 예측하려는 시계열 분석을 기반으로 한다. 하지만 추출된 데이터들은 6개의 블록 단위로 계산된 값이다. 블록 간의 시간간격은 투입되는 해시파워에 따라 달라질 수 있기 때문에 시계열 데이터 분석을 위해 그대로 사용하는 것은 부적합하며 시간 기반의 데이터로 변환해야 한다.

특히, 채굴 난이도 예측을 위해 채굴 난이도 변경 시점을 고려해서 데이터를 가공해야 한다. 비트코인은 2016블록마다 채굴 난이도가 변경되기 때문에 2016블록이 생성되는 이상적인 시간인 336시간(2016블록 × 10분) 단위로 데이터를 재계산한다.

4.3.4 슬라이딩 윈도우

기계학습을 위한 입력 값으로 336시간 단위로 계산된 자질들(features)을 사용하며, 출력 값으로 해시파워를 사용한다. 해시파워 예측 시 바로 이전의 데이터만을 사용할 수 있지만 몇 개씩 묶어서 입력 값으로 사용할 수도 있다. Fig.3.은 입력 값을 묶는 슬라이딩 윈도우(sliding window)의 과정을 보여준다. 윈도우 사이즈가 5인 경우 예측 전 데이터 5개씩을 하나의 윈도우로 묶고, 한 칸씩 이동(sliding)하면서 입력 값으로 사용될 윈도우를 구성한다.

본 연구에서는 기계학습 알고리즘마다 최적의 윈도우 사이즈를 구하여 오류율을 측정한다.

4.4 기계학습 알고리즘

시계열 데이터로 가공된 해시파워를 예측하기 위해 본 연구에서는 기계학습 알고리즘을 사용한다. 기계학습 툴로는 파이썬 인터페이스를 제공하는 텐서플로우와 싸이킷-런 프레임워크를 이용하였다.

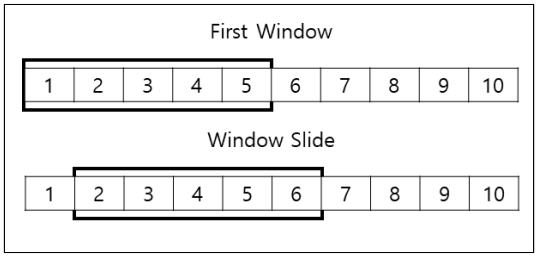


Fig. 3. Process of sliding window

추출한 자질세트와 해시파워를 8:2로 나누고, 이를 각각 학습 데이터와 테스트 데이터로 사용하여 오류율을 측정하였다. 본 연구에서는 해시파워 예측을 위한 기계학습 알고리즘으로 순환 신경망, 장단기 기억 신경망, 그리고 선형회기 세 가지 기법을 사용하여 비교한다.

4.4.1 순환 신경망

순환 신경망(Recurrent Neural Networks, RNN)은 데이터를 순차적으로 학습시키며 학습된 데이터를 학습 데이터로 재사용하는 특수한 형태의 신경망 네트워크이다. 신경망 네트워크는 입력 값의 크기에 민감하기 때문에 학습시에 해시파워의 log값을 취한 후 나머지 자질세트와 함께 MinMaxScaler로 정규화 시켰다.

4.4.2 장단기 기억 신경망

장단기 기억 신경망(Long Short-Term Memory, LSTM)은 RNN 모델에서 길이가 긴 데이터를 학습하는데 발생되는 vanishing gradient 문제를 해결하기 위해 등장한 네트워크이다. 많은 경우 시계열 데이터 예측을 위한 기계 학습에서 RNN보다 우수한 결과를 가져왔다[10][11]. RNN과 동일하게 log값을 취한 해시파워와 MinMaxScaler로 정규화 된 자질세트를 학습데이터로 사용한다.

4.4.3 선형회기

선형회기(linear regression)는 학습 데이터에서 오차를 고려한 선형 관계를 모델링하는 학습방법이다. 별도의 정규화는 진행하지 않고 자질세트와 해시파워를 그대로 사용하여 학습을 진행한다.

4.5 난이도 예측 시뮬레이션

시계열 데이터 예측에서는 시간별 데이터를 예측하기 때문에 Fig 4.(a)와 같이 다구간 예측시 이전 구간 예측에 오차가 발생하여도 다음구간 예측을 위한 입력 값은 변경하지 않는다. 이는 구간별로 예측의 오류를 측정하기 때문이다.

하지만 채굴 난이도 예측시에는 난이도 예측의 오차에 의해서 블록생성시간이 변경된다. 따라서 채굴 난이도 변경시점도 변경되기 때문에 다단계로 예측을 위해서는 매번 입력 값을 Fig.4.(b)와 같이 다시 계산해 주어야 한다. 정확한 난이도 예측 및 오류 측정을 위해 시간별 해시파워와 채굴 난이도 계산으로 블록생성에 대한 시뮬레이션을 진행해야 한다.

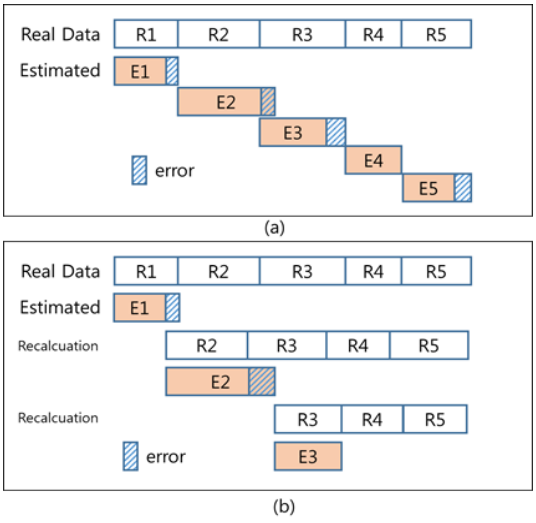


Fig. 4. (a) Prediction of Hash Power (b) Prediction of Mining Difficulty

V. 결과분석

5.1 실데이터 상의 해시파워 예측

해시파워 예측시 기계학습에 적합한 자질들을 찾기 위하여 학습할 입력데이터를 해시파워를 제외한 자질세트를 이용하는 경우, 해시파워를 포함한 자질세트를 이용한 경우와 해시파워만 사용한 경우 3가지 경우로 나누어 학습을 진행하였다.

성능 측정을 위해서는 비트코인 데이터 상에서의 실제로 투입된 해시파워와 예측한 해시파워를 비교하여 RMSE(Root Mean Square Error)와 MAPE(Mean Absolute Percentage Error)를 구하였다.

각각의 계산방법은 다음과 같다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad (1)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y_i}}{y_i}\right| \quad (2)$$

N은 예측 구간의 개수, y_i는 실제 값, $\hat{y_i}$는 예측된 값이다.

3가지 입력 데이터에 대하여 RNN, LSTM와 선형회기(LR) 방식으로 측정한 값은 각각 Table 1,

Table 1. Training without hash power

Response var.		Window Size	Error
RMSE	RNN	4	90.9689e+23
	LSTM	1	100.5807e+23
	LR	10	96.6023e+23
MAPE	RNN	4	42.8177%
	LSTM	1	86.3425%
	LR	10	65.6196%

Table 2. Training with all features

Response var.		Window Size	Error
RMSE	RNN	1	66.4484e+23
	LSTM	1	39.1920e+23
	LR	11	6.6352e+23
MAPE	RNN	1	37.4765%
	LSTM	1	16.0383%
	LR	11	4.3249%

Table 3. Training with only hash power

Response var.		Window Size	Error
RMSE	RNN	4	29.0475e+23
	LSTM	1	7.0495e+23
	LR	11	6.6584e+23
MAPE	RNN	4	24.1886%
	LSTM	1	4.5373%
	LR	11	4.3221%

Table 2, Table 3에서 확인할 수 있다. 측정 시 각각에 대하여 최적의 윈도우 사이즈를 선택하였다.

위 결과에서 보면 해시파워 이외의 자질세트가 포함되는 경우 오차가 큰 것을 확인할 수 있다. 기계학습 알고리즘 중에서는 선형회기를 사용했을 때 성능이 가장 좋게 나왔다.

5.2 실데이터 상의 난이도 예측

해시파워 예측과 달리 난이도 예측을 위해서는 이전 결과의 오차도 고려해야 한다. 전 단계 예측의 오차를 다음 단계에 반영해야 전체 단계에 대한 결과를 구할 수 있다. 이때 각 단계마다 학습 입력 값과 출력 값을 보정하고 다시 학습해야 한다. 따라서 난이도 예측 시뮬레이션에서는 매 단계 입·출력 값을 보정하여 새로운 학습을 통해 난이도를 예측하게 된다.

난이도 예측 실험에서는 이전에 진행했던 해시파워 예측 실험에서 성능이 좋았던 방식만을 사용했다. 자질 세트가 포함된 경우와 RNN을 수행한 경우 오류율이 상대적으로 높았기 때문에 해시파워 만을 사용한 LSTM과 선형회기 방식만을 사용했다.

성능 측정은 예측된 난이도상에서의 블록생성 시간과 설계된 예정시간 10분간의 차이를 비교하여 RMSE와 MAPE를 계산하였다.

다음 Table 4.에서는 채굴 시뮬레이션을 통한 채굴 난이도 예측 결과를 보여준다.

윈도우 사이즈가 1인 선형회기에서 가장 좋은 성능을 보여주었다. 기존 비트코인 채굴난이도 변경 방식으로 측정해보면 같은 기간동안 RMSE는 48.0490, MAPE는 6.5343%의 오류율을 보여 기계학습을 이용한 방식보다 성능이 좋지 않았다. 평균 블록 생성시간을 계산해 보면 윈도우 사이즈가 1인 선형회기 방법을 사용하는 경우 594초, 기존 채굴난이도 방식의 경우 566초로 기계학습을 사용한 방식이 블록 생성시간 안정화에 더 좋은 결과를 가져왔다.

Table 4. Prediction of Mining difficulty

Response var.		Window Size	Error
RMSE	LSTM	1	37.9917
	LR	1	33.9358
	LR	2	36.0563
MAPE	LSTM	1	4.6128%
	LR	1	4.1870%
	LR	2	4.4599%

5.3 상황별 난이도 예측

Meshkov 등의 연구에서는 대안알고리즘을 제시하고 두 가지 상황을 가정하여 블록 생성시간의 안정성을 검토하였다[6].

첫 번째는 채굴 난이도가 변경될 때마다 해시파워가 10%씩 증가한다는 가정이다. 이 경우 기존 비트코인 채굴변경 방식의 경우 정상 난이도수치의 차이는 9.1%이며, 해당 연구에서 제시한 대안알고리즘의 경우 1.9%의 차이를 보인다고 말하고 있다.

본 연구에서 가장 좋은 성능을 내는 선형회기 방식을 이용하는 경우 윈도우 사이즈가 2일 때 0.65%의 오차율을 보여준다. Fig.5.에서는 실제 채굴 난이도(파란색), 비트코인 내 난이도 변경방식(오렌지색), 해당 연구의 대안 알고리즘(녹색)과 본 연구의 선형회기를 이용한 난이도(빨간색) 예측 결과를 비교하여 보여준다. 해시파워가 10%씩 증가하는 경우에는 선형회기를 이용한 방식은 실제 채굴 난이도와 차이가 거의 없음을 알 수 있다.

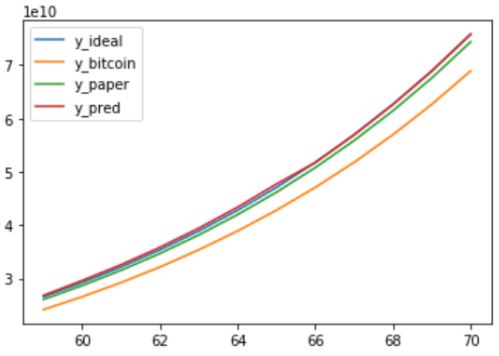


Fig. 5. Comparison of difficulty values

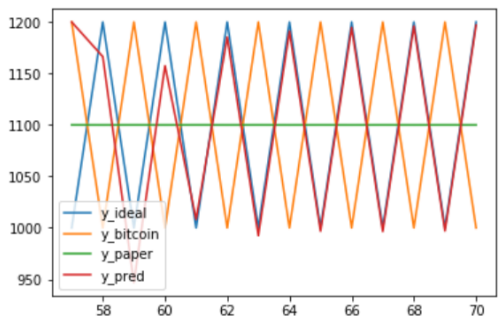


Fig. 6. Comparison of mining difficulties in the coin-hopping attack

두 번째 가정은 악의적인 채굴자가 전체 해시파워의 20%를 가지고 있을 때 coin-hopping 공격을 하는 경우이다. 평균 블록 생성시간이 기존방식에서는 10분 10초가 걸리지만 해당 연구의 대안알고리즘에서는 10분 5초로 줄기 때문에 공격자의 수입을 줄였다고 말하고 있다.

본 연구의 선형회기 방식을 이용하는 경우에는 9분 57초로 기존 연구보다 좋은 성능을 보여준다. Fig.6.은 coin-hopping 공격이 지속적으로 이루어지는 경우의 채굴 난이도를 비교한 것이다.

VI. 결 론

본 논문에서는 비트코인 채굴을 위해 투입되는 해시파워와 채굴 난이도를 예측하기 위한 기계학습 기반 연구를 수행하였다. 제안한 방법에 따른 실험결과 선형회기 방식을 이용한 예측이 가장 좋은 성능을 보였다. 이 방식을 사용하는 경우 기존 비트코인 채굴 난이도 변경 방식보다 오차율을 36% 더 줄일 수 있었으며 코인호핑 공격이 주기적으로 발생되는 경우에도 블록생성시간을 예정시간인 10분에 가깝도록 채굴 난이도를 조정하여 공격자의 수입을 최소화 시켰다.

실험에서 비트코인 블록체인 내부의 데이터들은 채굴 난이도와 상관관계가 부족하여 실제 난이도 예측에 이용할 수 없다는 결과를 보여주었고, 따라서 LSTM을 이용한 방식도 선형회기 방식보다 좋은 성능을 보여주지 못하였다. 하지만 실제 비트코인 시스템 내에 채굴 난이도 조정이 실시간으로 이루어져야 하기 때문에 수행속도가 빨라야 한다. 이러한 측면에서 데이터를 추출하고 학습에 많은 시간을 소비하는 신경망 네트워크 방식 보다는 해시파워 만을 계산하여 선형회기로 예측하는 방식이 더 적합하다고 판단된다.

본 연구는 채굴난이도와 비트코인 내부 데이터의 상관관계를 좀 더 명확하게 보여주지 못한 한계를 지니고 있다. 추후 채굴 난이도와 내부 정보에 대한 상관관계를 명확히 밝히고, 비트코인 가격 등의 외부 데이터를 이용한 예측 방식에 대하여서도 추가적인 연구가 필요할 것이다.

References

[1] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," Oct. 2008.

[2] Back, A, "A partial hash collision based postage scheme," http://www.hashcash.org/papers/announce.txt, last accessed 2018/10/27.

[3] Király, Tamás, and Lilla Lomoschitz. "Profitability of the coin-hopping strategy," EGRES quick proof, no. 2018-03, Mar. 2018.

[4] D. Kraft, "Difficulty control for blockchain-based consensus systems," Peer-to-Peer Networking and Applications, vol. 9, no. 2, pp.397-413, Mar. 2016.

[5] BIP 34 : Consensus (soft fork), https://github.com/bitcoin/bips/blob/master/bip-0034.mediawiki, last accessed 2018/10/27.

[6] Meshkov, Dmitry, Alexander Chepurnoy, and Marc Jansen. "Short paper: revisiting difficulty control for blockchain systems," Data Privacy Management, Cryptocurrencies and Blockchain Technology. Springer, Cham, vol. 10436, pp. 429-436, Sep. 2017.

[7] Kalodner, Harry, et al. "BlockSci: Design and applications of a blockchain analysis platform," arXiv preprint arXiv:1709.02489, Sep. 2017.

[8] Bitcoin Wiki : Difficulty, https://en.bitcoin.it/wiki/Difficulty, last accessed 2018/10/27.

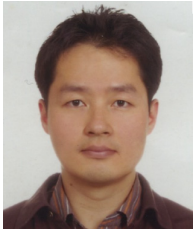
[9] Bitcoin Wiki : Block timestamp, https://en.bitcoin.it/wiki/Block_timestamp, last accessed 2018/10/27.

[10] Akita, Ryo, et al. "Deep learning for stock prediction using numerical and textual information," Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Confe-

rence on. IEEE, pp. 1-6, Jun. 2016.
〔11〕 McNally, Sean, Jason Roche, and Simon Caton. "Predicting the price of Bitcoin using Machine Learning," Parallel, Distributed and Network-based Processing (PDP), 2018 26th Euromicro International Conference on. IEEE, pp.339-343, Mar. 2018.
〔12〕 Bitcoin cash : Difficulty Adjustment Algorithm Update, https://www.bitcoinabc.org/2017-11-01-DAA/, last accessed 2018/10/27.
〔13〕 https://github.com/zawy12/difficulty-algorithms/issues, last accessed 2018/10/27.
〔14〕 Madan, Isaac, Shaurya Saluja, and Aojia Zhao. "Automated bitcoin trading via machine learning algorithms," Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech Rep, https://pdfs.semanticscholar.org/e065/3631b4a476abf5276a264f6bbff40b132061.pdf, last accessed 2018/10/27.
〔15〕 Jang, Huisu, and Jaewook Lee. "An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information," IEEE Access, vol. 6, pp. 5427-5437. Dec. 2017.
〔16〕 Rosenfeld, Meni. "Analysis of hashrate-based double spending," arXiv preprint arXiv:1402.2009, Feb. 2014.
〔17〕 Ittay Eyal and Emin Gün Sirer, "Majority is not enough: Bitcoin mining is vulnerable," Financial Cryptography and Data Security. vol. 8437, pp. 436-454, Mar. 2014.
〔18〕 Bahack, Lear. "Theoretical Bitcoin Attacks with less than Half of the Computational Power (draft)," arXiv preprint arXiv:1312.7013, Dec. 2013.

〈저자소개〉

이 준 원 (Joon-won Lee) 학생회원
2000년 2월: 연세대학교 컴퓨터과학과 학사
2017년 3월~현재: 연세대학교 정보대학원 석사과정
〈관심분야〉 정보보호, 기계학습, 블록체인 등



권 태 경 (Taekyoung Kwon) 종신회원
1992년 2월: 연세대학교 컴퓨터과학과 학사
1995년 2월: 연세대학교 컴퓨터과학과 석사
1999년 8월: 연세대학교 컴퓨터과학과 박사
1999년~2000년: U.C. Berkely Post-Doc.
2001년~2013년 8월: 세종대학교 컴퓨터공학과 교수
2007년~2008년: Univ. Maryland at College Park 교환교수
2013년 9월~현재: 연세대학교 정보대학원 교수
〈관심분야〉 암호프로토콜, Usable Security, 소프트웨어/시스템보안, 기계학습과보안 등