

# 국내 가상자산 거래소의 이더리움 지갑주소 식별을 위한 머신러닝 알고리즘 성능 비교

신용희, 배근우, 장대일, 지승구

한국인터넷진흥원

itmater@kisa.or.kr, bkwl402@kisa.or.kr, dale@kisa.or.kr, eleca@kisa.or.kr

## Performance Comparison of Machine Learning Algorithms for Identifying Ethereum Wallet Addresses in Korea Virtual Asset Exchange

Shin Yong-hee, Bae Keun-woo, Jang Dae-il, Ji Seung-goo

Korea Internet & Security Agency

### 요 약

최근 마약 및 음란물 유통, 범죄자금 은닉 등 기존 범죄가 가상자산을 이용한 사이버 범죄로 발전하고 있고, 가상자산 중 이더리움을 이용한 범죄가 다수 일어나고 있다. 이렇게 이더리움을 이용한 범죄의 범죄자 검거와 범죄수익 환수를 위해서는 범죄에 사용된 이더리움 지갑주소가 어느 거래소의 지갑주소인지 식별할 수 있는 기술이 필요하다고 판단하였다. 본 논문에서는 실제 이더리움 거래를 통해 수집한 국내 가상자산 거래소 5곳의 지갑주소와 이더리움 클라이언트를 통해 무작위로 수집한 지갑주소를 Dataset으로 구성하고, 4가지 머신러닝 알고리즘을 이용하여 각 거래소 지갑주소에 대한 식별율을 측정하는 실험을 진행하여 어떤 알고리즘이 가장 높은 식별율을 보이는지에 대한 결과를 도출하였다.

### I. 서 론

최근 가상자산을 이용한 사이버 범죄가 급증하고 범죄의 수준이 고도화되고 있으며, 마약 및 음란물 유통, 범죄자금 은닉 등 기존 범죄가 가상자산을 이용한 사이버 범죄로 발전하고 있다. 특히, 2019년에는 국내 거래소 해킹으로 580억원 규모의 이더리움이 도난 당하는 사건이 발생하여[1], 한국 경찰은 미국 법무부와 공조수사를 통해 해킹그룹 라자루스(Lazarus)가 범인이라고 발표[2]하였다. 2020년에는 최소 500억원 이상의 '이더월렛' 이더리움 투자사기 사건[3]이 발생하여, 경찰이 수사를 통해 범인을 검거하였다.

이와 같은 이더리움을 악용한 범죄의 범죄자 검거와 범죄수익 환수를 위해서는 범죄에 사용된 이더리움 지갑주소가 어느 거래소의 지갑주소인지 식별할 수 있는 기술이 필요하다.

### II. 본론

#### 2.1 실험대상 지갑주소의 수집

본 논문에서는 이더리움 지갑주소를 식별하기 위한 대상이 되는 거래소를 국내 가상자산 거래소 5곳으로 결정하고, 거래소 5곳에 대한 지갑주소와 기타 지갑주소(거래소 5곳에 속하지 않는 랜덤한 지갑주소)를 수집하기로 하였다. 우선 거래소 지갑주소의 수집을 위해 각 거래소에 개인 지갑주소를 생성하고, 실제 거래를 통해 거래소들의 거래패턴을 분석한 결과 5곳의 거래소들이 개인지갑 주소에 송금된 이더리움을 거래소의 핫월렛으로 자동이체를 진행하고 있는 패턴을 확인하여 각 거래소들의 핫월렛과 핫월렛에 연결된 다수의 지갑주소를 수집하였다. 기타 지갑주소의 수집은 이더리움 클라이언트[4]에서 무작위로 추출하여 수집하였다. 각 거래소별로 수집한 지갑주소와 기타로 수집한 지갑주소 Dataset은 <표1>과 같다.

<표1> 실험용 Dataset

구 분	수집한 지갑주소 개수
A 거래소	192,385개
B 거래소	2,849개
C 거래소	109개
D 거래소	1,801개
E 거래소	1,086개
기타 지갑주소	111,036개
합계	309,266개

#### 2.2 학습을 위한 Feature 및 머신러닝 학습 알고리즘 선정

우리는 수집한 이더리움 지갑주소 Dataset에 머신러닝 알고리즘 별로 학습을 통한 거래소 식별율을 측정하기 위해 이더리움의 거래 특징 중 <표2>와 같이 14가지의 Feature를 선정하였다.

<표2> 머신러닝 모델 학습을 위한 Feature

특 정	설 명
sentCount	송금 트랜잭션 수
receivedCount	수신 트랜잭션 수
contractCreation	컨트랙트 생성 트랜잭션 수
uniqueFrom	해당 지갑주소로 전송한 다른 지갑주소 수
uniqueTo	해당 지갑주소가 전송한 다른 지갑주소 수
minValueReceived	수신한 최소금액
maxValueReceived	수신한 최대금액
avgValueReceived	수신한 평균금액
minValueSent	송신한 최소금액
maxValueSent	송신한 최대금액

avgValueSent	송신한 평균금액
totalEthSent	총 송신금액
totalEthReceived	총 수신금액
totalEthBalance	잔액

실험을 위한 머신러닝 알고리즘으로는 KNeighbors, DecisionTree, RandomForest, GradientBoosting으로 총 4개의 알고리즘을 선정하였다.

### 2.3 실험방법 및 결과

위에서 수집한 Dataset과 Feature를 가지고 총 4개의 머신러닝 알고리즘을 통해 거래소 5종의 지갑주소 식별율을 측정하는 실험을 진행하였다. 실험을 위해 Dataset을 TrainingSet(80%), TestSet(20%)으로 나누고 각 머신러닝 알고리즘 별로 accuracy(정확도), precision(정밀도), recall(재현율), f1-score 측정결과를 <표3>과 같이 도출하였다.

<표3> 머신러닝 모델별 거래소 5종의 지갑주소 식별율 측정결과

구 분	KNeighbors	Decision Tree	Random Forest	Gradient Boosting
accuracy	0.9019	0.9148	0.9317	0.8791
precision	0.9001	0.9152	0.9278	0.8773
recall	0.9019	0.9148	0.9317	0.8791
f1-score	0.8972	0.9149	0.9289	0.8669

본 연구의 Accuracy(정확도)란, 전체 데이터 중 각 거래소들의 지갑주소를 얼마나 정확하게 분류하였는지를 측정하는 지표로 Accuracy가 1인 경우 100%의 정확도로 분류하였음을 의미한다.

Precision(정밀도)이란, 학습 알고리즘이 지갑주소 데이터 중에서 A거래소의 지갑주소라고 분류한 지갑주소가 실제 A거래소 지갑주소가 맞는지를 측정하는 지표로 Precision이 1인 경우 학습 알고리즘이 분류한 각 거래소의 지갑주소가 실제 해당 거래소의 지갑주소일 확률이 100%임을 의미한다.

Recall(재현율)이란, 실제 A거래소의 지갑주소를 학습 알고리즘이 A거래소의 지갑주소라고 분류하였는지를 측정한 지표로 Recall이 1인 경우 실제 각 거래소의 지갑주소를 학습 알고리즘이 해당 거래소라고 분류할 확률이 100%임을 의미한다.

F1-Score는 Precision과 Recall의 조화평균으로 본 실험에서와 같이 분류하고자 하는 데이터의 개수가 불균형한 경우 가장 정확한 식별율을 나타내는 평가지표로, 아래 <그림1>과 같은 수식으로 구할 수 있다.

<그림1> F1-Score 수식

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

실험결과, 4개의 머신러닝 알고리즘 중에서 RandomForest 알고리즘이 가장 좋은 식별율을 보이는 것을 확인할 수 있었다.

## III. 결론

본 논문에서는 실제 거래를 통하여 수집한 국내 가상자산 거래소 5곳의 이더리움 지갑주소 식별율에 대한 accuracy, precision, recall, f1-score 결과를 4가지의 머신러닝 알고리즘을 통해 측정하여 비교하였다. 실험결과, 우리가 선정한 14가지의 Feature를 통해서 RandomForest 모델이 가장 좋은 식별율을 보이는 것을 확인할 수 있었다.

향후 연구에서는 식별 대상 거래소를 늘리고, 딥러닝 알고리즘을

추가하여 10개 거래소의 이더리움 지갑주소 식별율을 측정하는 실험을 진행할 예정이다.

## ACKNOWLEDGMENT

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-00901, 가상자산 부정거래 등 사이버범죄 활동 정보 추적 기술)

## 참 고 문 헌

- [1] Upbit, [https://upbit.com/service\\_center/notice?id=1085](https://upbit.com/service_center/notice?id=1085)
- [2] Department of Justice, <https://www.justice.gov/opa/press-release/file/1253491/download>
- [3] 중앙일보, 2020.06.12, <https://www.joongang.co.kr/article/23799948#home>
- [4] <https://geth.ethereum.org/downloads/>