KIPS Trans. Comp. and Comm. Sys.
Vol.12, No.5 pp.165~170
ISSN: 2287-5891 (Print), ISSN 2734-049X (Online)

https://doi.org/10.3745/KTCCS.2023.12.5.165

Ethereum Phishing Scam Detection based on Graph Embedding and Semi-Supervised Learning

Yoo-Young Cheong† · Gyoung-Tae Kim† · Dong-Hyuk Im††

ABSTRACT

With the recent rise of blockchain technology, cryptocurrency platforms using it are increasing, and currency transactions are being actively conducted. However, crimes that abuse the characteristics of cryptocurrency are also increasing, which is a problem. In particular, phishing scams account for more than a majority of Ethereum cybercrime and are considered a major security threat. Therefore, effective phishing scams detection methods are urgently needed. However, it is difficult to provide sufficient data for supervised learning due to the problem of data imbalance caused by the lack of phishing addresses labeled in the Ethereum participating account address. To address this, this paper proposes a phishing scams detection method that uses both Trans2vec, an effective graph embedding techique considering Ethereum transaction networks, and semi-supervised learning model Tri-training to make the most of not only labeled data but also unlabeled data.

Keywords : Blockchain, Ethereum, Phishing Scam, Graph Embedding, Semi-supervised Learning

그래프 임베딩 및 준지도 기반의 이더리움 피싱 스캠 탐지

정 유 영† · 김 경 태† · 임 동 혁††

요 약

최근 블록체인 기술이 부상하면서 이를 이용한 암호화폐 플랫폼이 늘어나며 화폐 거래가 활발이 이뤄지고 있다. 그러나 암호화폐의 특성을 악용한 범죄 또한 늘어나 문제가 되고 있다. 특히 피싱 스캠은 이더리움 사이버 범죄의 과반수 이상을 차지하며 주요 보안 위협원으로 여겨지고 있다. 따라서 효과적인 피싱 스캠 탐지 방법이 시급하다. 그러나 전체 이더리움 참여 계정 주소에서 라벨링된 피싱 주소의 부족으로 인한 데이터 불균형 문제로 지도학습에 충분한 데이터 제공이 어려운 상황이다. 이를 해결하기 위하여 본 논문에서는 이더리움 트랜잭션 네트워크를 고려한 효과적인 그래프 임베딩 기법인 trans2vec과 준지도 학습 모델 tri-training을 함께 사용하여 라벨링된 데이터 뿐만 아니라 라벨링되지 않은 데이터도 최대한 활용하는 피싱 스캠 탐지 방법을 제안한다.

키워드 : 블록체인, 이더리움, 피싱, 그래프 임베딩, 준지도 학습

1. 서 론

블록체인은 양 당사자 간의 거래와 관련된 정보를 중앙 시스템 없이 검증 가능하고 영구적으로 기록하는 분산형 공개 원장이다[1]. 나카모토 사토시는 비트코인 프로젝트를 시작하여 블록체인을 이용한 최초의 성공적인 암호화폐를 소개하였다[2]. 비탈릭 부테린이 비트코인에 영감을 받아 개발한 이더리움[3]은 퍼블릭 블록체인 플랫폼으로 스마트 컨트랙트를 지원하며 암호화폐 이더(ether)를 기반으로 고속 성장하여 현재 비트코인과 함께 최대 규모의 플랫폼이다. 최근 이러한 블록체인 암호화폐에 대한 관심이 높아지며 사이버 범죄의 대상으로 문제가 되고 있다. 특히 이더리움에서는 2017년 이후 피싱(Phishing) 사기 사건이 전체 사이버 범죄의 50% 이상을 차지하고 있으며, 이더리움 거래 보안의 주요 위협원이 되고 있다[4].

피싱 사기는 신뢰할 수 있는 개체로 위장하여 사용자의 민감 정보를 얻으려고 시도한다. 이더리움에서의 피싱 사기는 피해자를 통해 거래 취소나 변경이 불가능한 이더를 피의자 측으로 이전시키는 것으로 발생할 수 있다[5]. 피싱 스캠 탐지 문제는 다양하게 논의되었으며, 여러 가지 방법이 제안되었다[6]. 그러나 기존 피싱 사기와 비교했을 때 이더리움에서

※ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.NRF-2021R1F1A1054739). 또한, 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터지원사업의 연구결과로 수행되었음(IITP-2023-2018-0-01417).
※ 이 논문은 2022년 한국정보처리학회 ACK 2022의 우수논문으로 "그래프 임베딩 기반의 이더리움 피싱 스캠 탐지 연구"의 제목으로 발표된 논문을 확장한 것임.
† 준 회 원 : 광운대학교 인공지능응용학과 석사과정
†† 종신회원 : 광운대학교 정보융합학부 부교수
Manuscript Received : December 19, 2022
First Revision : February 1, 2023
Accepted : February 14, 2023
* Corresponding Author : Dong-Hyuk Im(dhim@kw.ac.kr)

의 피싱 사기는 다른 특성을 보인다.

첫째, 현금이 아닌 암호화폐가 대상이 되면서 이더리움에서의 피싱 범죄자들은 취득한 암호화폐를 실제 화폐와 교환하여 현금화해야 한다. 둘째, 이더리움과 같은 퍼블릭 블록체인 시스템은 모든 거래 기록에 공개적으로 접근할 수 있어 서로 다른 사용자 간의 거래 유형을 파악할 수 있는 완전한 데이터 소스를 제공하므로 피싱 탐지가 용이하다. 마지막으로, 기존의 피싱 사기는 사용자의 민감한 정보를 얻기 위해 피싱 이메일과 웹 사이트에 의존하기 때문에 기존의 피싱 탐지 방법은 일반적으로 피싱 사기 정보가 포함된 이메일 또는 웹 사이트를 탐지하는 방법에 초점을 맞추고 있다[6]. 따라서 이더리움과 같은 블록체인 플랫폼에서 기존의 피싱 사기 탐지 방식을 적용할 수 없기 때문에 블록체인 플랫폼에서의 피싱 탐지 문제를 해결하기 위한 연구들이 진행되고 있다[1, 7].

데이터 투명성과 무결성이 특징인 블록체인 암호화폐는 거래 기록에서 정보 추출이 가능하여[8] 추출한 정보를 이용하여 피싱을 탐지한다. 트랜잭션 이력을 네트워크로 모델링하면, 노드는 고유한 주소이고 노드의 엣지는 두 주소 사이에 하나 이상의 이더 전송이 존재하는 것을 의미한다. 그러나 피싱 탐지를 위한 트랜잭션 네트워크 사용은 극심한 데이터 불균형 문제가 존재한다. 이더리움의 블록 탐색 및 분석 플랫폼인 etherscan.io 에 따르면, 총 참여 계정 주소 및 트랜잭션 수는 각각 5 억개와 38 억 개를 초과하는 반면, 라벨링된 피싱 주소는 2041개에 불과하다[1]. 따라서 데이터 불균형은 피싱 탐지 문제에서 지도 학습 접근법을 사용할 경우 성능에 영향을 미칠 수 있다. 또한 이러한 대규모 네트워크 데이터를 학습에 사용할 경우 feature 선택이 성능에 중요하다.

Y. Y. Cheong et al[9]은 이더리움 트랜잭션 이력을 네트워크로 모델링한 트랜잭션 네트워크와 준지도 학습 알고리즘을 이용하여 피싱 노드를 탐지하는 방법을 제안하였다. 본 논문에서는 이 연구를 확장하여 트랜잭션 네트워크에서 그래프 임베딩을 사용하여 feature를 추출하고, 라벨링되지 않은 데이터의 레이블을 생성하여 training 데이터로 사용하는 준지도 학습 알고리즘으로 피싱 노드를 탐지하는 방법을 제안한다. 제안된 방법의 기여는 다음과 같이 요약할 수 있다.

- 본 연구는 이더리움 네트워크 구조에서 참여 노드들의 특징을 기반으로 새로운 피싱 스캠 탐지 모델을 설계한다. 본 연구의 모델은 그래프 임베딩 알고리즘과 준지도 알고리즘을 결합하여 이더리움 네트워크에서 피싱 노드를 효과적으로 식별한다. 실험은 그래프 임베딩을 기반으로 한 네트워크 feature가 이더리움에서 피싱 노드를 정확하게 식별하는 데 도움이 될 수 있음을 증명한다.
- 준지도 학습 알고리즘인 Tri-training을 사용하면 라벨링 되지 않은 데이터를 최대한 활용하여 라벨링된 데이터에 대한 필요성을 줄이고 모델의 일반화 능력을 향상시킬 수 있다. 본 연구에서는 Tri-training을 사용하여 이더리움 피싱 스캠 탐지에서 지도 학습 모델의 성능 문제를 효과적으로 해결한다.

2. 연구 배경 및 관련 연구

2.1 이더리움

이더리움[3]은 암호화폐와 탈중앙화 애플리케이션(DApp)을 모두 지원하는 차세대 블록체인 플랫폼이다. 전용 암호화폐인 이더(Ether)를 사용하여 구동하는 가상 머신을 통해 point-to-point 계약을 처리하는 퍼블릭 블록체인 플랫폼으로 사용자 간에 신뢰할 수 있는 거래를 지원한다. 이더리움에는 사용자가 통제하는 기본 거래 계정(Externally Owned Account)과 프로그램 코드를 저장하고 있는 계약 계정(Contract Account)가 존재한다.

2.2 피싱 스캠

피싱(Phishing)은 사용자의 개인 정보(사용자 이름, 비밀번호, 주민등록번호 등)를 취득하기 위하여 신뢰할 수 있는 기업의 웹 사이트를 사칭하는 기술로 정의된 온라인 위협의 한 형태를 말한다[10]. 기존의 피싱 사기는 피싱 이메일과 웹 사이트에 의존하기 때문에 기존의 탐지 방법은 피싱 웹 사이트를 파악하는 것을 목적으로 URL, 하이퍼링크, 피싱 가능성을 암시하는 민감한 단어 및 기타 콘텐츠 기반 feature를 추출한다. 그러나 블록체인 암호화폐에서의 피싱 스캠은 블록체인 참여 주소를 다양한 방식으로 유포하여 직접 돈을 사취한다[1]. 따라서 탐지 대상 및 데이터 등의 측면에서 기존 방식과 달리 본 연구에서 이더리움 피싱 스캠 탐지의 목적은 피싱자의 이더리움 주소를 탐지하는 것으로, 이더리움 주소 간의 공개된 거래 기록에 접근하여 주소 간 거래 행위를 마이닝하여 거래 내역에서 주소의 feature를 추출 및 학습하여 피싱 주소를 탐지한다.

2.3 그래프 임베딩

이더리움 네트워크에서 피싱 노드의 경우 거래가 특정 방향으로만 발생하는 현상을 보이는데, 피싱 노드는 일반적으로 짧은 기간 내에 여러 노드로부터 상당한 금액을 확보한 후 특정 주소로 빠르게 돈을 송금한 후 인출하는 경향을 보이는데 반해 일반 노드는 주변의 노드와 다양한 상호작용을 하며 거래한다[11]. 다른 노드와의 역방향 상호작용이 피싱 노드에는 관찰되지 않는 것이다. 따라서 이더리움 네트워크에서 노드 간 거래 특징을 효과적으로 추출하는 것이 중요하다. 네트워크 데이터에서 그래프 임베딩을 기반으로 한 이상 탐지 알고리즘은 현재 여러 업계에서 많은 관심을 끌고 있다.

그래프 임베딩은 노드 간 임베딩 매핑 기능을 학습하여 d 차원의 feature 공간에서 이웃 노드의 공존 가능성을 극대화하는 것을 목표로 한다. 노드 간의 구조적 관계를 포착하는 그래프 임베딩 방법은 (1) Factorization, (2) 랜덤 워크, (3) 딥 러닝 등 세 가지 방법으로 나눌 수 있다[12]. Factorization 기반 알고리즘은 노드 간 연결 정보를 사용하여 인접 행렬, 라플라시안 행렬 등 다양한 행렬을 구성하여 행렬을 인수 분해하여 임베딩을 얻는다. 랜덤 워크는 그래프를 부분적

으로만 관측할 수 있거나 그래프가 너무 커서 전체를 확인할 수 없는 경우에 특히 유용하다. 노드 표현을 얻기 위해 그래프에 랜덤 워크를 사용하는 임베딩 기법으로는 DeepWalk [13], Node2vec[14] 등이 제안되었다. 이더리움 트랜잭션 데이터에서의 피싱 스캠 탐지에 그래프 임베딩을 적용한 연구로 J. Wu et al.[1]은 대규모 이더리움 트랜잭션 네트워크에서 거래 금액인 amount 값과 타임스탬프 값을 통합하여 트랜잭션 네트워크를 위한 새로운 그래프 임베딩 모델인 Trans2vec 기법을 제안하였다.

2.4 준지도 학습

라벨링된 데이터와 라벨링되지 않은 데이터를 모두 사용하는 준지도 학습 알고리즘인 Tri-training[15]은 세 개의 분류기를 사용하는 알고리즘으로, 라벨링되지 않은 데이터에 레이블을 지정한 후 새롭게 라벨링된 데이터를 포함한 데이터로 분류기들을 재학습하여 최종 분류를 결정하는 기법이다. He et al.[16]은 소셜 네트워크에서 온라인 마케팅 회사 등의 조직에 의해 악의적으로 댓글과 의견을 게시하는 그룹을 탐지하기 위하여 그래프 임베딩 기법과 준지도 알고리즘을 결합한 모델을 제안하였다. 그래프 임베딩 기법인 Node2vec으로 소셜 네트워크의 구조 feature를 추출한 후 준지도 알고리즘인 Tri-training으로 악성 사용자를 탐지하여 레이블이 지정된 데이터에 대한 의존성을 낮췄다. 그러나 블록체인 네트워크 피싱 스캠 탐지를 위한 모델이 아니며, 사용하는 그래프 임베딩 기법은 블록체인 트랜잭션 네트워크에서 구체적인 거래 정보를 효과적으로 추출하는 방법이 아니다.

따라서 본 연구에서는 이더리움 트랜잭션 네트워크에서 효과적으로 feature를 추출하기 위하여 Trans2vec 모델을 사용하며, 극심한 데이터 불균형 문제를 개선하기 위하여 Tri-training으로 라벨링되지 않은 데이터에 레이블을 생성한 후 피싱 및 비피싱 노드를 분류한다.

3. 이더리움 피싱 탐지 모델

본 논문에서 피싱 스캠 탐지 모델은 Fig. 1과 같이 세 가지 단계로 구성되어 있다. (1) 이더리움 클라이언트를 통해 수집된 트랜잭션 이력과 Etherscan.io, EtherScam DB에서 라벨링된 피싱 주소를 결합하여 노드는 피싱 및 일반 주소이며 엣지는 두 주소 간의 트랜잭션을 나타내는 이더리움 트랜잭션 네트워크를 구축한다. (2) 네트워크에서 거래 금액과 타임스탬프 정보 feature를 추출하기 위하여 Trans2vec으로 그래프 데이터를 벡터화 한다. (3) 준지도 학습 알고리즘인 Tri-training으로 피싱 및 비피싱 노드를 분류한다.

Trans2vec 단계에서 그래프 임베딩에서 랜덤 워크를 수행함으로써 대규모 네트워크는 샘플링된 노드 시퀀스 집합으로 변환된다. Wu et al.[1]이 제안한 Trans2vec은 랜덤 워크 기반 그래프 임베딩 기법으로, 거래 금액과 타임스탬프 정보에 편향된 샘플링을 한다. 소스 노드로부터 랜덤 워크를 수행하여 매 단계마다 관계가 더 강한 노드를 다음 노드로 채택하여 진행한다.

Amount 기반 샘플링 : 이더리움 트랜잭션 네트워크에서는 거래 금액 Amount 값이 클수록 두 노드 간의 관계가 강해진다는 것을 의미하며, V_u가 노드 u에서 직접 연결된 노드 집합일 때, Amount 기반 샘플링에서 노드 u에서 인접 노드 $x \in V_u$로의 전이 확률은 Equation (1)과 같다. 여기서 $A(u,x)$는 노드 u와 x 사이의 총 Amount 값을 나타낸다.

$$PA_{ux} = \frac{A(u,x)}{\sum_{x' \in V_u} A(u,x')} \tag{1}$$

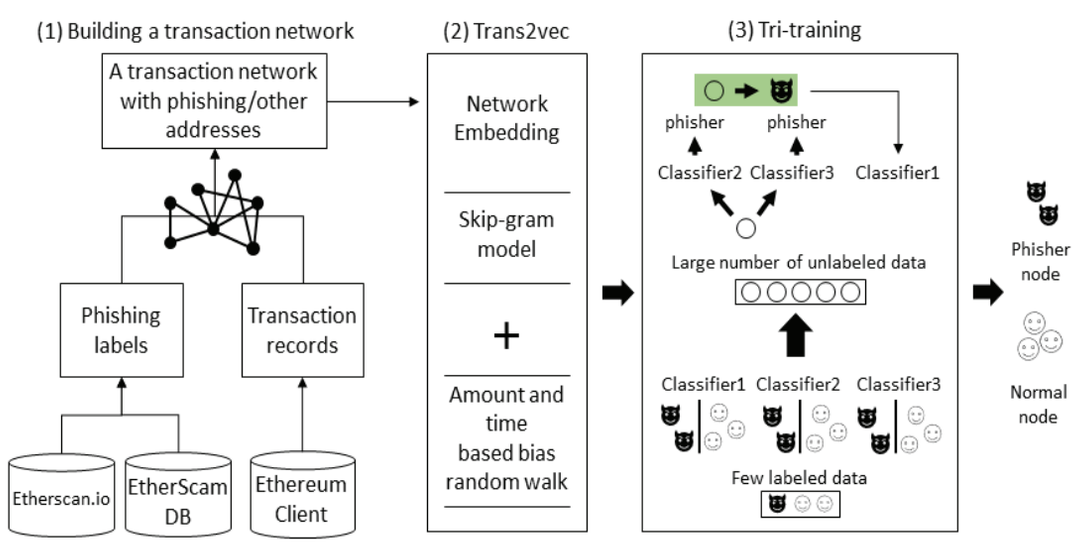


Fig. 1. Ethereum Phishing Scam Detection Structure

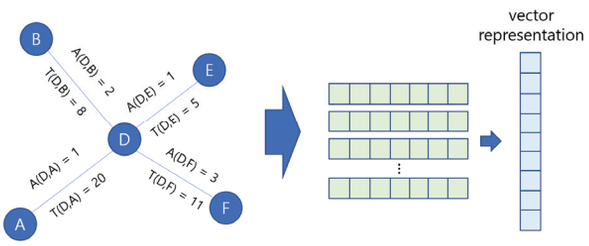


Fig. 2. Graph Embedding Structure

타임스탬프 기반 샘플링 : Wu et al.[1]는 Amount 이외에 타임스탬프 값도 노드 간 관계에 미치는 영향이 존재한다고 가정한다. Amount와 마찬가지로 노드 u에서 인접 노드 $x \in V_u$로의 전이 확률은 Equation (2)와 같다.

$$PT_{ux} = \frac{T(u,x)}{\sum_{x' \in V_u} T(u,x')} \tag{2}$$

Fig. 2에서 노드 D에서 $A(u,x)$값 만으로 다음 노드를 결정하는 경우 노드 F가 선택될 가능성이 높지만, $T(u,x)$ 값으로 고려할 경우 노드 A가 선택될 수 있다. 따라서 Trans2vec은 거래 금액 정보와 타임스탬프 정보를 모두 고려하기 위하여 매개 변수 $\alpha(0 \le \alpha \le 1)$값을 통해 트랜잭션 금액과 타임스탬프 사이의 값을 조절하여 균형을 맞춰 샘플링하여 확률은 Equation (3)과 같다.

$$\pi_{ux}(\alpha) = PA_{ux}^{\alpha} \cdot PT_{ux}^{1-\alpha} \tag{3}$$

또한 생성된 시퀀스로부터 word2vec 스킵그램 모델을 사용하여 특정 노드에 대한 이웃 노드들의 발생 확률을 최대화하는 함수를 최적화한다.

Zhou et al.[15]이 제안한 Tri-training은 세 가지 분류기를 사용하는 알고리즘이다. 라벨링된 데이터 집합을 L, 라벨링되지 않은 데이터 집합을 U 라고 할 때, 초기에 분류기 (M_1, M_2, M_3) 는 L에서 샘플링된 데이터 L_1, L_2, L_3로 학습된다. 이후 반복문에서 선택한 분류기 $M_i(i,j,k \in (1,2,3))$를 제외한 나머지 분류기M_j, M_k로 U에서 동일하게 샘플링된 데이터를 예측한다. 두 분류기의 예측 결과가 동일하면 해당 샘플 및 생성한 레이블은 높은 신뢰도를 갖는 것으로 간주하여 M_i의 라벨링된 훈련 세트에 추가한다. 세 분류기에 각각 위 과정을 반복한 후 최종적으로 분류기들의 분류 결과를 종합한다.

$$|L \cup L^t|\left(1 - 2\frac{\eta L|L| + \check{e}_1^t|L^t|}{|L \cup L^t|}\right)^2 > |L \cup L^{t-1}|\left(1 - 2\frac{\eta L|L| + \check{e}_1^{t-1}|L^{t-1}|}{|L \cup L^{t-1}|}\right)^2 \tag{4}$$

Zhou et al.[15]은 새로운 훈련 샘플이 Equation (4)에 정의된 조건을 충족한다면, 재훈련 후 분류 성능이 향상될 것이라는 것을 입증했다. L^t는 t번째 반복에서 라벨링된 샘플의 집합이고, ηL은 L의 분류 노이즈 비율을 나타낸다. 즉, 잘못 라벨링된 샘플은 $\eta L|L|$이다. $\check{e}_1^t$은 t번째 반복에서선택 분류기를 제외한 나머지 두 분류기의 분류 오류 상한선이다. ηL이 매우 작을 수 있고, $0 \le \check{e}_1^t, \check{e}_1^{t-1} < 0.5$ 라고 가정할 때, $|L^t| > |L^{t-1}|$ 이고, $\check{e}_1^t|L^t| < \check{e}_1^{t-1}|L^{t-1}|$ 인 경우 Equation (5)와 같이 정의할 수 있다.

$$0 < \frac{\check{e}_1^t}{\check{e}_1^{t-1}} < \frac{|L^{t-1}|}{|L^t|} < 1 \tag{5}$$

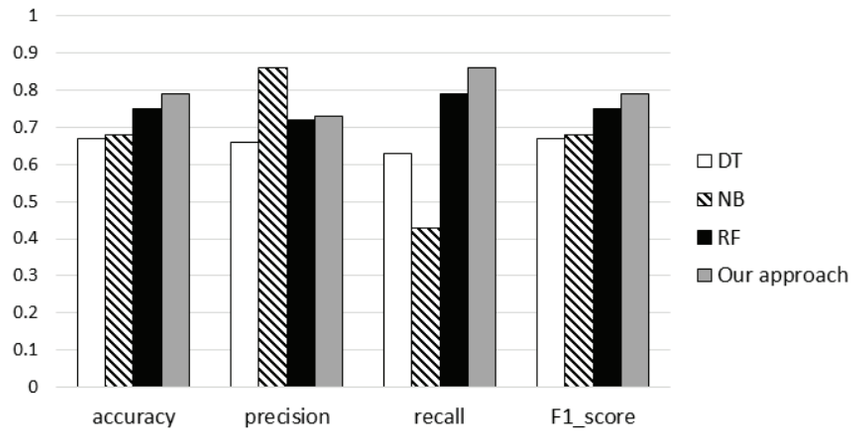
M_j, M_k가 동일하게 제공한 레이블 샘플 데이터인 L^t를 M_i의 훈련 데이터에 추가 가능 여부를 결정하기 위하여 Equation (5)를 사용한다.

4. 실험 및 평가

실험에서 사용한 데이터는 Etherscan.io에서 제공하는 피싱 주소를 이용하여 라벨링된 주소와 해당 주소의 1차 거래 기록을 포함하는 데이터로 피싱 주소 445 개를 포함한 트랜잭션 네트워크를 구축하였다. Trans2vec으로 얻은 벡터 표현은 64 차원이며, Tri-training 알고리즘에서는 세 가지 분류기를 사용하므로 랜덤 포레스트, BP Neural networks 및 Naive Bayes 알고리즘을 사용하였다. 제안하는 알고리즘의 효과를 확인하기 위하여 기존의 지도학습 알고리즘을 단일로 사용하였을 경우와 비교하였다. 라벨링된 데이터의 20%만을 학습 데이터로 설정하여 Decision Tree(DT), Naive Bayes(NB), 랜덤 포레스트(RF) 알고리즘을 학습시킨 후 본 연구에서 제안한 방법과 성능을 비교하였다. 실험에서 성능을 평가하기 위해 Accuracy, Recall, Precision, F1-score 평가 지표를 사용하였으며, 결과는 Table 1과 같다. 본 연구에서 제안한 방법이 다른 알고리즘보다 precision을 제외하고 좋은 성능을 보였다. 이는 준지도 학습인 Tri-training 알고리즘의 경우 기존의 지도 학습 알고리즘과 비교하여 라벨링된 데이터의 수가 적을 때 더 나은 탐지 결과를 얻을 수 있음을 나타낸다.

또한 적은 양의 라벨링된 데이터에서 좋은 결과를 얻을 수 있는지 확인하기 위하여 제안하는 방법에서 라벨링된 데이터 비율을 10%부터 90%까지 증가시켜 실험하였다. 실험 결과는 Fig. 3과 같다. 10%의 비율에서만 약간 다른 수치를 보이지만 기본적으로 모든 평가지표가 비율 별로 변동하지 않았다. 이는 라벨링된 데이터가 적은 조건에서도 대략적인 탐지 결과를 얻을 수 있음을 나타낸다.

Table 1. Comparison with Conventional Machine Learning Algorithms



Method	Accuracy	Precision	Recall	F1_score
Decision Tree(DT)	0.67	0.66	0.63	0.67
Naïve Bayes(NB)	0.68	0.86	0.43	0.68
Random Forest(RF)	0.75	0.72	0.79	0.75
Our approach	**0.79**	**0.73**	**0.86**	**0.79**

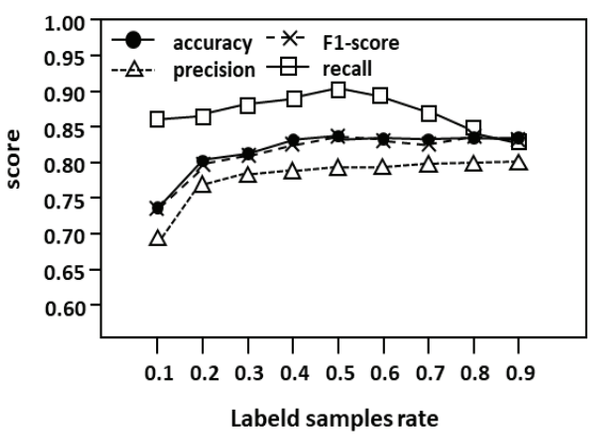


Fig. 3. The Influence of Different Labeled Data Ratio on Experimental results

5. 결 론

본 논문에서는 이더리움 트랜잭션의 특징에 맞는 Trans2vec 그래프 임베딩 기법과 라벨링되지 않은 데이터도 학습에 사용하기 위하여 Tri-training 준지도 학습 알고리즘을 함께 사용한 탐지 방법을 제안하였다. 실험을 통하여 레이블이 적은 암호화폐 트랜잭션 네트워크 데이터에 대하여 기존 라벨링된 데이터를 기반으로 세 가지 분류기의 시너지를 활용하여 레이블을 확장시켜, 라벨링된 데이터 비율이 적어도 탐지 결과를 얻을 수 있음을 확인하였다.

본 연구는 이더리움 플랫폼으로 특정하여 진행하였으나 추후 연구를 통하여 다른 블록체인 플랫폼으로 확장할 수 있을 것이며, 레이블 확장 생성에 있어 피싱 스캠 탐지에 좀 더 효과적인 모델이 될 수 있도록 성능 개선 연구를 진행할 것이다.

References

[1] J. Wu et al., "Who are the phishers? phishing scam detection on ethereum via network embedding," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol.52, No.2, pp.1156-1166, 2022.

[2] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," [Internet], https://bitcoin.org/bitcoin.pdf, 2008.

[3] V. Buterin, "A next-generation smart contract and decentralized application platform," *White Paper*, Vol.3, No.37, pp.1-36, 2014.

[4] M. Conti, E. S. Kumar, C. Lal, and S. Ruj, "A servey on security and privacy issues of bitcoin," *IEEE Communications Serveys & Tutorials*, Vol.20, No.4, pp.3416-3452, 2018.

[5] K. F. K. Low and E. Teo, "Legal risk of owning crytocurrencies," Handbook of Blockchain, Digital Finance, and Inclusion, Vol.1, London: Academic Press, 2008.

[6] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: A literature survey," *IEEE Communications Surveys & Tutorials*, Vol.15, No.4, pp.2091-2121, 2013.

[7] L. Chen, J. Peng, Y. Liu, J. Li, F. Xie, and Z. Zheng, "Phishing scams detection in ethereum transaction network," *ACM Transactions on Internet Technology(TOIT)*, Vol.21, No.1, pp.1-16, 2020.

[8] D. Lin, J. Wu, Q. Yuan, and Z. Zheng, "Modeling and understanding ethereum transaction records via a complex network approach," *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol.67, No.11, pp.2737-2741, 2020.

[9] Y. Y. Cheong, K. T. Kim, and D. H. Im, "Ethereum phishing scam detection based on graph embedding," *in Proceedings of the Annual Conference of Korea Information Processing Society Conference (KIPS) 2022*, Vol.29, No.2, 2022.

[10] N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection based associative classification data mining," *Expert Systems with Applications*, Vol.41, No.13, pp.5948-5959, 2014.

[11] T. Yu, X. Chen, Z. Xu, and J. Xu, "MP-GCN: A phishing nodes detection approach via graph convolution network for ethereum," *Applied Sciences*, Vol.12, No.14, pp.7294, 2022.

[12] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowledge-Based Systems*, Vol.151, pp.78-94, 2018.

[13] B. Perozzi, R. Al-Rfou, S. Skiena, "DeepWalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.701-710, 2014.

[14] A. Grover, J. Leskovec, "Node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.855-864, 2016.

[15] Z. H. Zhou and M. Li, "Tri-training: Exploting unlabeled data using three classifiers," *IEEE Transactions on Knowledge and Data Engineering*, Vol.17, No.11, pp.1529-1541, 2005.

[16] Y. He, P. Yang, and P. Cheng, "Semi-supervised internet water army detection based on graph embedding," *Multimedia Tools and Applications*, Vol.82, No.7, pp.9891-9912, 2023.

정 유 영

https://orcid.org/0000-0003-2786-5592
e-mail : yycheong@kw.ac.kr
2015년 덕성여자대학교 심리학과(학사)
2022년 ~ 현 재 광운대학교
인공지능응용학과 석사과정
관심분야 : 블록체인, 이상탐지, 머신러닝

김 경 태

https://orcid.org/0000-0003-0723-7898
e-mail : kkt9601@kw.ac.kr
2022년 호서대학교 컴퓨터공학과(학사)
2022년 ~ 현 재 광운대학교
인공지능응용학과 석사과정
관심분야 : NLP, 지식공학, 추천 시스템

임 동 혁

https://orcid.org/0000-0002-0290-755X
e-mail : dhim@kw.ac.kr
2003년 고려대학교 컴퓨터교육과(학사)
2005년 서울대학교 컴퓨터공학과(석사)
2011년 서울대학교 컴퓨터공학과(박사)
2011년 ~ 2012년 서울대학교 의생명지식공학연구실 선임연구원
2013년 ~ 2020년 호서대학교 컴퓨터공학과 조교수/부교수
2020년 ~ 현 재 광운대학교 정보융합학부 부교수
관심분야 : 빅데이터, 머신러닝, 시공간 데이터 분석