GAN 을 활용한 비트코인 불법거래 과표본 추출 기법

한정수°, 이채현, 고경찬, 우종수, 홍원기
포항공과대학교,정보통신대학원

{saw1515, chlee0211, kkc90, woojs, jwkhong}@postech.ac.kr

Bitcoin Illegal Transaction Oversampling Technique with GAN

Jungsu Han°, Chaehyeon Lee, Kyungchan Ko, Jongsu Woo, James Won-Ki Hong
Department of Computer Science and Engineering, POSTECH
Graduate School of Information Technology, POSTECH

요 약

비트코인 거래는 익명성이 보장되어 범죄와 관련된 불법거래 사례가 급증하고있다. 이를 해결하기 위한 연구들이 활발히 진행되고 있지만 불법거래를 수집하는 것과 그것을 정상거래와 구별하는 것이 어렵기 때문에 불법거래 표본의 갯수가 정상 거래 갯수보다 부족한 문제가 있다. 이와 같이 클래스 불균형 데이터셋을 다루는 방법으로 과표본추출(SMOTE) 모델을 사용하는 경우가 대다수이지만 해당 모델의 경우 데이터의 다양성을 충분히 표현하지 못할 가능성이 존재한다. 본 논문에서는 적대적 신경망(GAN)을 이용하여 비트코인 불법거래 탐지에 사용되는 데이터의 클래스 불균형 문제를 해결하는 방법을 제안한다. 그리고 생성된 데이터와 실제 데이터의 유사도를 검증하기 위해 XGBoost classification 을 활용하여 모델 성능을 정량적으로 평가하였다.

Ⅰ. 서 론

암호화폐란 암호화된 공개키를 이용하여 블록체인 네트워크상에서 거래가 가능하며, 해시 함수를 통해 쉽게 소유권을 증명할 수 있는 디지털 자산이다. 암호화폐의 핵심 기술인 블록체인은 분산장부(Distributed ledger)를 이용하여 개인간의 모든 거래 정보와 계좌 정보를 공개하지만 각각의 정보에 대한 신상은 제공하지 않음으로써 거래의 익명성을 보장한다[1]. 특히, 블록체인 네트워크에서 개인을 식별하는 것은 계좌의 주소만을 이용해서 가능한데 이 같은 계좌 주소를 개인이 여러 개를 생성하거나 소유 할 수 있다는 특징 때문에 익명성을 보장받게 된다. 하지만 이 같은 블록체인 거래의 익명성을 이용하여, 악의적인 목적을 가진 네트워크 참여자가 불법적인 거래를 일으킬 수 있는 문제가 있다. Sean[2]의 조사에 따르면 가장 대표적인 암호화폐인 비트코인의 경우 미・유럽을 중심으로 무기, 약물, 돈세탁 등을 목적으로 매년 76 억 달러에 해당하는 거래가 일어나고 있다고 분석하였다. 그렇기 때문에 블록체인이 차세대 금융시스템으로 발전하기 위해서는 불법・사기 거래들에 대한 탐지와 분류가 필수적이다.

분류 문제(classification problem)란 기존의 데이터 셋의 패턴을 분석하여 새로운 입력 데이터에 대한 클래스를 예측하는 문제이다. 일반적으로 기계학습을 통해 분류 모델을 개발하는데, 이때 학습할 데이터셋은 클래스의 갯수에 맞게 균등한 것이 이상적이지만 실제로 수집되는 대부분의 데이터의 경우 클래스가 불균형한 형태이다. 비트코인 불법 거래 분류문제에 사용되는 비트코인 거래 데이터의 경우에도 불법 거래(illegal transaction)의 주소를 확보하는 것이 어렵고 한정적이기 때문에 불법거래로 라벨링 된 거래보다 그렇지 않은 것이 훨씬 많이 존재하는 불균형 데이터셋이다. 대부분의 경우 불균형 데이터셋(imbalanced dataset)으로 인해 모델의 성능이 저하되는 것을 해결하기 위해 SMOTE(Synthetic Minority Over-sampling TEchnique)와 같은 리샘플링(resampling) 방법들을 사용하게 된다[3]. 그렇지만 SMOTE 의 경우 다른 클래스의 데이터와의 거리를 고려하지 않아 생성된 데이터의 클래스 중첩(overlapping of class)이 일어날 수 있는 문제가 있고 고차원 데이터(high dimensional data)에는 적합하지 않다. 본 논문에서는 이 문제를 해결하기 위해 적대적 생성 신경망(GAN, Generative Adversarial Network)을 이용하여 비트코인 불법거래 탐지에 사용되는 불균형 데이터 문제를 해결할 방안을 제시한다.

Ⅱ. 배경지식

1. Oversampling

Haixiang [4]은 대표적인 과표본추출 기법으로 Random Oversampling(ROS), SMOTE, ADASYN 가 사용된다고 설명하였다. ROS 는 소수 클래스 데이터 샘플을 무작위로 복제하여 생성하는 방식이고 SMOTE 는 데이터 샘플 간의 K-nearest neighbor 를 선택하여 그 점들을 이어 새로운 데이터 샘플을 만들게 된다. ADASYN 은 SMOTE 에서 발전된 기법으로 SMOTE 의 방식으로 만들어진 데이터의 왜곡된 분포를 막기 밀도분포(density distribution)을 이용하여 더 현실성 있는 표본을 생성하는 방식이다.

2. GAN and CGAN

GAN(generative adversarial network)은 기존의 데이터를 바탕으로 실제 데이터 셋에 없는 새로운

데이터를 생성하는 모델로 일반적으로 생성자(generator)와 구분 자(discriminator) 두 심층 네트워크로 이루어져있다[5, 6]. 생성자는 생성된 데이터에 랜덤 노이즈를 주어 생성된 데이터가 실제 데이터 분포와 유사한 형태를 만드는 것이 목적이고 구분자는 생성자가 만들어낸 데이터와 실제 데이터를 구별하는 것이 목적인 모델이다. 이 두 모델은 경쟁적(adversarial)으로 학습을 진행하여 더 현실적인 데이터를 생성하게 된다.

CGAN(conditional-GAN)은 GAN 의 확장된 모델로 data space 에 additional space 를 추가한 형태이다. 예를 들면 MNIST 데이터셋에서 특정 숫자의 값을 CGAN 으로 생성하고 싶으면 그 숫자의 label 을 additional space 로 추가하여 모델을 학습하면 된다.

3. WGAN and WCGAN

기존의 GAN 에서는 데이터 분포 간의 거리를 재는 방식으로 KL divergence 나 JS divergence 를 사용하였는데 이 두 방식은 생성 모델이 데이터의 특정 값(예를 들면 MNIST 의 label 이 1 인 이미지)에만 생성하는 mode collapse 현상이 발생하는 문제가 있었다. 이를 해결하기 위해 Arjovsky [7]은 실제 데이터의 확률분포를 알기 위해 Earth Mover(EM) distance 를 사용하였다. 해당 논문에서는 EM distance 값을 최소화하기 위해 Wasserstein-GAN(WGAN)을 제안하여 WGAN 이 생성자와 구분자의 균형에 예민한 GAN 이 학습 중에 발생할 수 있는 문제를 해결할 수 있음을 보였다.

Ⅲ. 실험

비트코인의 트랜잭션 데이터를 수집하기 위해 자체 비트코인 풀노드(full node)를 구축하였다. 그리고 불법 트랜잭션을 수집하기 위해 WalletExplorer.com나 해외 포럼과 같은 곳에서 불법 거래가 이루어진 트랜잭션의 hash나 지갑 주소를 조사하고 그에 해당하는 트랜잭션을 풀노드에서 수집하여 레이블링 하였다. 이 때 일반 거래소에서 발생한 트랜잭션을 0으로 레이블링 하였으며 silkroad(비트코인 불법거래 사이트)에서 발생한 트랜잭션을 1로 레이블링 하였다. 또한 수집한 트랜잭션의 특징(feature)을 알기 위해 트랜잭션의 비트코인 전송량, 전송 횟수, lifetime, fee, sibling 갯수 등을 함께 수집하였다. 수집한 특징에 대하여 주성분 분석(PCA)를 적용하여 15개의 주성분 벡터와 class 레이블 총 16개의 특징으로 데이터를 전처리 하였다. 그림 1은 전처리 한 주성분 벡터 중 데이터 분류에서 특징 중요도(feature importance)가 높은 상위 5가지를 시각화한 그래프이다.

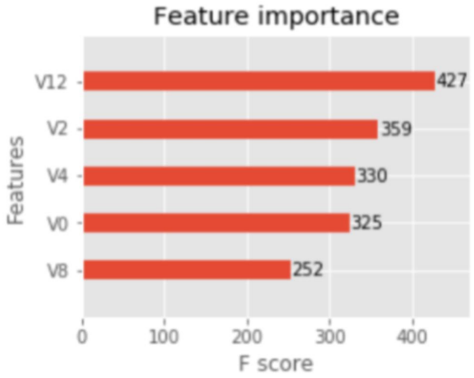

그림 1. Feature importance 상위 주성분 벡터

어떤 특징이 데이터 분류(classification)에 큰 영향을 주는지를 확인하기 위해 XGBoost algorithm[8]을 사용하였으며 분류에 영향력이 큰 상위 두 특징(PCA로 전처리된 V2, V12)을 이용해 GAN, WGAN 등으로 oversampling 데이터를 분석하였다. 또한 class 레이블을 추가한 CGAN, WCGAN을 구현하여 동일한 방식으로 분석하였다.

GAN 구조에 따른 생성 데이터는 그림 2와 같이 시각화 할 수 있다. 그림에서 초기 학습단계에서는 실제 데이터의 형태와 유사한 형태로 데이터를 생성하지만 학습 단계가 1000 step이 넘기 시작하자 일부 데이터 분포로 수렴하기 시작하였다. 이는 학습이 특정(non-optimal)한 샘플 분포에 수렴하는 mode collapse 때문으로 보인다. CGAN의 경우 class 값에 따라 각각 특정 분포로 수렴하게 된다. 데이터 간의 분포를 EM distance를 이용하여 분석하는 WGAN, WCGAN의 경우 GAN과 달리 class 정보 유무에 상관없이 모두 mode collapse를 보이지 않는 것으로 확인되었다.

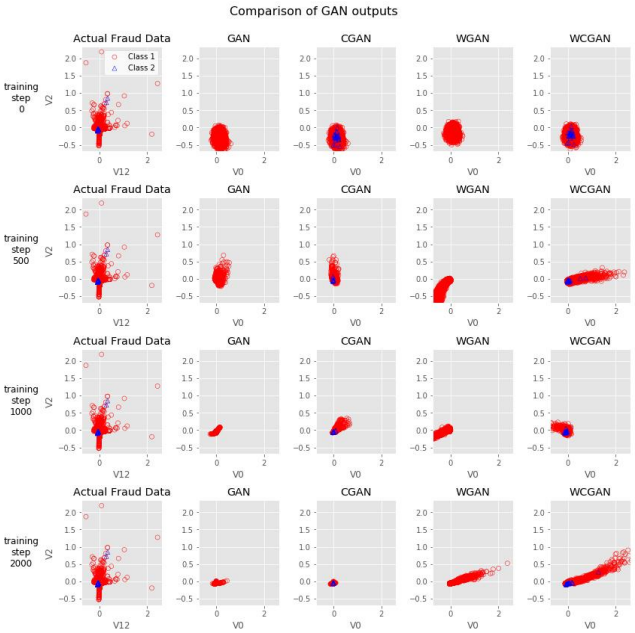

그림 1. GAN 종류에 따른 생성 데이터 시각화

GAN 모델 중 성능이 가장 좋게 나온 WCGAN과 SMOTE 방식으로 oversampling 된 각각의 데이터

셋을 이용하여 XGBoost classifier을 이용한 불법 트랜잭션 분류 결과를 정량적으로 비교하였다. 이때 학습되지 않은 WGAN 모델로 생성된 데이터와 실제 불법거래 데이터를 추가 하였을 때의 XGBoost classifier의 성능을 함께 표 1과 같이 정량적으로 분석하였다.

	auc	precision	recall	roc_auc
Untrained	0.9282	1.0	0.2826	0.9833
WCGAN	0.9280	0.9978	0.2808	0.9843
SMOTE	0.9291	0.9989	0.2919	0.9848
Real	0.9946	0.9693	0.9774	0.9993

표 1. XGBoost Classification 성능

WCGAN, Untrained WCGAN, SMOTE 방식으로 생성된 가짜 데이터를 활용하여 XGBoost classifier의 성능 변화를 테스트하였다. 만약 classifier의 성능에 변화가 있다면 생성된 데이터가 실제 데이터와 유사하다는 것을 의미한다. 그림 3을 보면 생성된 데이터가 추가됨에 따른 재현율(recall) 값의 변화 확인할 수 있다. Untrained WGAN의 경우 생성된 데이터가 추가됨에 따라 재현율의 변화가 크게 없는 것으로 확인되었다. 이는 생성된 데이터가 실제데이터와 유사하지 않다는 것을 의미한다. 하지만 WCGAN과 SMOTE 모델의 경우에도 생성된 데이터를 통한 클래스 분류의 재현율이 큰 차이가 나지 않는 것을 확인하였다. 생성된 데이터가 아닌 실제 데이터만을 사용하였을 경우 데이터 수에 따라 재현율이 0.9774까지 증가하는 것과 비교하였을 때 데이터 생성 모델(WCGAN, SMOTE)의 재현율 기준선(baseline)이 0.3 내외로 크게 차이 나는 것을 확인하였다

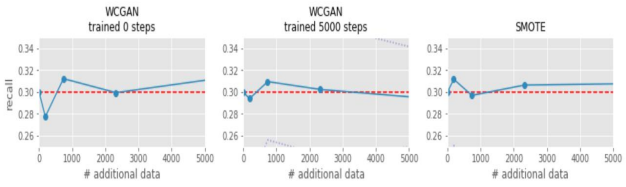

그림 3. WGAN vs SMOTE 비교

Ⅳ. 결론 및 향후 연구

본 연구는 비트코인 네트워크에서 생성되는 트랜잭션 데이터를 GAN 을 이용하여 과표본추출하는 방법을 제안하였다. 이를 통해 비대칭 데이터의 불균형 문제를 해결하여 기존의 과표본추출방식 보다 분류 모델의 성능향상이 있을 것으로 기대하였다. 그렇지만 실험결과 유의미한 성능 차이를 확인하지 못하였는데 이는 GAN 모델의 근본적인 문제와 트랜잭션 데이터 자체의 문제로 예상된다. GAN 모델은 구분자와 생성자 두 네트워크를 사용하게 되는데 이때 특정 네트워크의 성능이 지나치게 우수하면 다른 모델의 성능이 오르지 않는 현상이 일어날 수 있다. 또한 구분자가 실제 데이터 분포를 완벽하게 학습하면 GAN 의 기본 원리인 Minmax Game[8]에 모순이 생겨 학습 도중에 특정 데이터 분포에 진동(oscillation)하며 학습이 반복되는 문제가 있다. 그렇기 때문에 다른 기계학습 모델에 비해 학습이 최적의 값에 수렴하는 것이 상대적으로 어렵다. 그리고 수집한 트랜잭션 데이터의 특징 분포를 확인한 결과 불법거래와 정상 거래가 중첩되는 부분이 상당한 것으로 확인되어 클래스 분류 자체가 어렵다는 문제가 있었다. 향후 연구로는 최근 우수한 성능을 보여주고 있는 DCGAN 을 이용하여 모델의 성능을 높이고 데이터의 클래스를 분류에 영향을 줄 수 있는 추가적인 특징들을 수집하여 데이터클래스 분류가 더욱 용이하도록 할 예정이다

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구 임 (No.2018-0-00539)

참 고 문 헌

[1] Nakamoto, S. Bitcoin: A peer-to-peer Electronic Cash System (2008)

[2] Foly, S. Karlsen, J. R., & Putniņ ,T. J. Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies?. The Review of Financial Studies, pp. 1798- 1853. (2019)

[3] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. SMOTE: synthetic minority oversampling technique. Journal of artificial intelligence research. pp. 321-357. (2002)

[4] Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications. pp. 220-239. (2017)

[5] Mullick, S. S., Datta, S., & Das, S. Generative adversarial minority oversampling. In Proceedings of the IEEE International Conference on Computer Vision. pp. 1695-1704. (2019)

[6] Ba, H. Improving Detection of Credit Card Fraudulent Transactions using Generative Adversarial Networks. arXiv preprint arXiv:1907.03355. (2019).

[7] Arjovsky, M., Chintala, S., & Bottou, L. Wasserstein gan. arXiv preprint arXiv:1701.07875. (2017).

[8] Chen, T., & Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785-794. (2016)

[9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S. & Bengio, Y. Generative adversarial nets. In Advances in neural information processing systems. pp. 2672-2680. (2014)

[10] Walletexplorer: smart bitcoin block explorer. Available at https://www.walletexplorer.com/.

[11] Bitcoin.org. Bitcoin core json apis. Available at https://bitcoin.org/en/developer-reference#bitcoin-core-apis.