ANOMALY DETECTION IN CRYPTOCURRENCY TRANSACTIONS WITH ACTIVE LEARNING

Leandro L. Cunha and Miguel A. Brito
Centro Algoritmi
University of Minho, Portugal

ABSTRACT

Cryptocurrencies have gained tremendous popularity in recent years, with the rise of Bitcoin and other altcoins. However, this surge in popularity has also attracted fraudulent activities, such as scams, phishing, and money laundering. Particularly, machine learning (ML) algorithms have the potential to detect these fraudulent patterns. However, since in the fraud detection (FD) domain labels are scarce and most times very hard to get, traditional supervised ML models cannot be applied. Additionally, traditional unsupervised anomaly detection (AD) algorithms, generally, lead to high false positive rates. Therefore, this study is intended to explore the feasibility of using AD and active learning (AL) algorithms to uncover new fraudulent patterns in cryptocurrency transactions, assuming minimal access to labels.

KEYWORDS

Anomaly Detection, Active Learning, Fraud Detection, Unsupervised Learning, Cryptocurrencies, Machine Learning

1. INTRODUCTION

Computing technology has revolutionized ML, making it accessible for identifying opportunities and managing risks through data pattern recognition in various industries. The need for effective AD systems in the fraud detection domain remains a growing concern, particularly in the cryptocurrency market where digital currencies have increased the risk of fraudulent activities. The absence of regulation in this market makes it more vulnerable to fraud, resulting in significant losses for companies. Thus, companies must adopt efficient strategies for detecting and preventing fraud to avoid bearing the burden of substantial losses.

Training high-performance supervised classifiers for FD is crucial, but the lack of labeled data makes it challenging to train such ML models (Labanca *et al.*, 2022). The process of labeling an entire dataset requires a vast sample of manually reviewed transactions, which is often unpractical due to the high costs involved and the limited investigative time and budget available for companies (Barata *et al.*, 2021). Thus, an efficient mechanism is needed to help analysts focus their investigation on the most relevant transactions with a higher probability of being fraudulent and to efficiently uncover increasingly complex and sophisticated fraud patterns, given that fraudsters are constantly finding new ways to commit fraudulent activities.

The absence of labeled data to train supervised classifiers highlights the immense value of combining AD and AL techniques to detect anomalies in data, as potentially fraudulent transactions, and iteratively collect labels for training supervised ML classifiers. AD algorithms have the potential to alleviate the issue of labeled data by flagging unusual or suspicious behaviors and patterns within the data. However, being unsupervised, they may result in false positives, given the high-class imbalance and growing sophistication of fraudulent behaviors that are becoming more similar to normal patterns. AL has the potential to combine the benefits of unsupervised (i.e., exploration) and supervised (i.e., exploitation) methods by iteratively selecting the most valuable data to be reviewed and labeled by an oracle from an initial unlabeled pool. In other words, the most anomalous transactions are presented to an analyst for review according to an anomalous ranking provided by the anomaly detector and after the feedback, the labeled data can be used to train a supervised classifier that learns fraud. Thus, the usage of AD in conjunction with AL allows for more efficient utilization of the labeled data.

Is intended with this work to assess the feasibility of using AD and AL algorithms to uncover new fraudulent patterns in cryptocurrency transactions, assuming minimal access to labels. Thus, the main objective is to benchmark different AL setups that use AD algorithms and gain insights on how to improve the performance of an anomaly detector, incorporated in an AL process that iteratively gathers labels that would allow a supervised model to be trained with only the most informative transactions, thereby, reducing the necessary labeling effort. We organize the remainder of the paper as follows. Section 2, presents the main concepts related to our work, providing a clear understanding of the topic. Section 3 details the existing literature on this subject matter to gain a more comprehensive view of the problem. Finally, in Section 4, the main conclusions are discussed.

2. MAIN CONCEPTS

2.1 Anomaly Detection

ML algorithms play a fundamental role in FD by using transactional data to identify fraudulent patterns through high-level pattern recognition.

AD is the process of identifying patterns in data that deviate from the expected normal behavior (Injadat *et al.*, 2018). Anomalies are rare occurrences with little to no common characteristics and are not consistent with the majority of observations. The primary objective of AD is to create a profile of normal behavior and classify a data instance as anomalous if it deviates from this norm (Xuan *et al.*, 2018). Particularly, in transactional data, anomalies can detect fraudulent activity by identifying abnormal transactional patterns (Hilal *et al*., 2022).

Essentially, anomalies can be categorized in three ways: point anomaly, contextual anomaly, and collective anomaly. A point anomaly is when a single data instance deviates from the normal pattern. A contextual anomaly is when data instances behave anomalously in a specific context. A collective anomaly is when a group of similar instances act anomalously when compared to the whole dataset (Ahmed *et al*., 2016). Another issue to consider is how anomalies are represented in the output. This can be in the form of a score indicating the likelihood of a data instance being anomalous or a label assigning each data instance as either normal or anomalous (Ahmed *et al*., 2016).

Unlike imbalanced classification, AD is an unsupervised learning technique that aims to identify outliers in a dataset. Generally, unsupervised AD methods can be divided into four different categories: nearest-neighbor, clustering, subspace, and tree-based techniques. Although the primary focus of this work is on unsupervised AD techniques, it should be noted that certain approaches incorporate semi-supervised discriminators trained to acquire knowledge about the normal instances' boundary. In such scenarios, instances that lie outside of the established boundary are considered outliers (Lorenz *et al*., 2020).

2.2 Active Learning

AL is a method that improves ML performance by strategically selecting data for training. We may put queries, based on certain heuristics, in form of unlabeled data to be labeled by someone that understands the nature of the problem (Settles, 2012). This reduces the number of labels needed for training a supervised classifier by iteratively sampling the most informative samples for review from an initial unlabeled pool (Lorenz *et al.*, 2020). Thus, AL can be considered a specific instance of semi-supervised learning.

The AL process begins with an initial pool of unlabeled data or only a few labeled instances. The most informative samples, from the unlabeled pool, are then selected and labeled by an oracle based on a specific AL querying strategy. Further, the labeled samples are moved to the labeled pool and used to train and evaluate high-performance supervised ML models. This loop is repeated until a certain number of labels are collected (i.e., given a specific budget) or the performance of a supervised ML model trained on a fully labeled dataset is achieved. If the performance is still not satisfactory, the querying process continues to incrementally expand the labeled pool. It is important to note that the AL process is applied only to the training set of the data. The training set is divided into two pools: the unlabeled pool (containing the unlabeled transactions) and the labeled pool (containing instances that have already been queried and reviewed by an analyst). The supervised ML model is trained on the labeled pool and evaluated on the test set to assess its effectiveness. Therefore, the advantage of

AL is that we can have a human loop that reviews a set of data instances identified by the anomaly detector (e.g., the top 50 most anomalous transactions), and assigns them to the correct class label. So, with less effort, we can have a system with similar performance to a supervised baseline.

3. LITERATURE REVIEW

3.1 Traditional Approaches for Fraud Detection

Most of the cryptocurrency FD research often uses ML models applied to the Elliptic dataset, one of the largest labeled datasets for any cryptocurrency. The dataset is a graph network of Bitcoin transactions and is highly valuable for financial FD and research communities. Supervised machine learning models have proven effective in detecting fraudulent transactions. Weber *et al*. (2019) applied several binary classification methods to the Elliptic Dataset and found that Random Forest (RF) was the best-performing model, despite the graph structure information suggesting an advantage for graph-based approaches like Graph Convolutional Networks (GCN). Alarab *et al*. (2020) also achieved successful results using an ensemble of supervised learning techniques on the Elliptic dataset, outperforming Weber *et al*. (2019). Particularly, RF is a commonly used supervised ML model in cryptocurrency FD research. Studies have shown that RF performs well in classifying Ethereum transactions as fraudulent or not (Ostapowicz and Żbikowski, 2019; Ibrahim *et al*., 2021). A recent study by (Melo-Acosta *et al*., 2017) proposed a new approach using a BRF (Balanced Random Forest), which involves iteratively selecting samples from the minority class and constructing a tree-based classifier. This approach outperforms traditional supervised methods in both supervised and semi-supervised learning through co-training.

Additionally, a study conducted by Niu *et al*. (2019) compared supervised and unsupervised ML algorithms using an imbalanced credit card dataset. The results, as expected, showed that supervised models outperformed unsupervised ones.

Unsupervised techniques have shown promise in detecting fraudulent transactions, with AD methods yielding optimistic and encouraging results and achieving high detection rates. Several studies have highlighted the commendable processing speed and detection accuracy of these approaches in the field of financial fraud detection (Zengan, 2009; Monamo *et al*., 2017). As with supervised classifiers, several authors focused on benchmarking the performance of several unsupervised AD techniques for FD (Domingues *et al.*, 2018; Ounacer *et al.*, 2018). The results revealed that the IF (Isolation Forest) algorithm outperforms all other techniques, proving to be an exceptional method for identifying outliers, even on large datasets. Additionally, some studies have even demonstrated the efficacy of AD techniques in uncovering fraudulent transaction patterns that were previously undetected or missed by conventional rule-based systems (Zengan, 2009).

While AD algorithms have shown promising results, their real-world effectiveness is still in question. Testing on a real-world Bitcoin dataset Lorenz *et al*. (2020) showed high false positive rates, highlighting the limitations of relying on synthetic anomalous data. Criminals can imitate normal behavior, making it harder to identify fraud. Therefore, solely relying on AD algorithms without human analyst review can lead to high false positive rates and may not be practical in real-world applications.

3.2 Anomaly Detection with Active Learning

Studies on financial systems are limited, but valuable insights have been provided. Particularly, combining unsupervised AD and supervised query strategies is advocated as a means of improving the AL process. AD algorithms have limitations in detecting fraudulent transactions, with slow identification and high false positive rates. Common supervised AL policies include uncertainty sampling, query-by-committee, and expected model change. Uncertainty sampling selects samples with the highest uncertainty based on the classifier's confidence in its prediction (i.e., the transaction with the predicted probability closest to 0.5) (Sharma and Bilgic, 2017). Query-by-committee employs a committee of multiple learners and leverages the disagreement among classifiers to make a decision (Kee *et al*., 2018). Finally, expected model change chooses the instances that would result in the maximum change of the current model when queried (Cai *et al*., 2017).

Labanca *et al*. (2022) proposed a framework that ranks anomalous data using unsupervised AD methods and trains supervised models on domain experts' feedback using two AL strategies. The authors emphasized the importance of aggregating transactions for capturing correlations and found that IF and RF were the best-performing ML models with high detection rates. Other studies have also contributed to the literature by designing multi-stage AL labeling policies (Lorenz *et al.*, 2020; Barata *et al.*, 2021), which initially apply unsupervised AD algorithms to rank the most anomalous transactions for review before switching to supervised AL policies. Barata *et al*. (2021) proposed an intermediate stage using ODAL as a warm-up learner, which efficiently alleviated the cold start scenario with high-class imbalance. Both studies found that switching to supervised learners improved the models' performance, with Lorenz *et al*. (2020) achieving promising results by matching the performance of a supervised baseline using just 5% of the labels. Lastly, Carcillo *et al*. (2017) developed strategies to handle high-class imbalance in streaming credit card FD scenarios. They evaluated different AL strategies and found that combining the baseline strategy of only querying high-risk transactions with Stochastic Semi-supervised learning resulted in higher fraud detection rates. This being said, an effective AL strategy for AD requires proper tuning.

4. DISCUSSION

Most literature shows that AD methods often underperform in the AL process, leading to suboptimal performance. The slow identification of fraudulent transactions in the unlabeled pool is the primary reason for this. One solution is to switch to a supervised hot learner after a threshold or combine unsupervised and supervised techniques to select the most informative transactions for querying. However, it's unclear if enhancing the anomaly detector's performance in an AL setup would lead to better results. If it can detect more fraudulent transactions in earlier iterations, it may improve the overall model performance.

We aim to improve the anomaly detectors' performance by benchmarking different AL setups on the Elliptic Dataset. Similarly, to Lorenz *et al*. (2020) we will use unsupervised and supervised AL strategies depending on the number of illicit transactions in the labeled pool (i.e., when a certain threshold occurs, we either switch from an unsupervised warmup learner to a supervised hot-learner or continue to use the unsupervised learner). The goal is to improve AD algorithm performance in the early stages by considering the intrinsic characteristics of the Bitcoin transaction network. Given the possibility that illicit transactions are often connected to other illicit transactions in the network, it is possible to leverage this information by combining it with the anomaly scores produced by the anomaly detector. We propose to investigate this, by benchmarking each AL setup against a supervised baseline.

5. CONCLUSION

The present article delves into the significant challenges posed by traditional ML approaches in detecting financial fraud using cryptocurrencies. To address these issues, a combination of AD and AL algorithms is proposed. The proposed approach aims to simulate a typical real-life scenario where a limited number of labels can be acquired through manual annotation by experts. By incorporating the intrinsic characteristics of the Bitcoin transaction network and graph information, the article proposes to investigate how the performance of the anomaly detector can be enhanced in an AL setup, leading to better results. The proposed approach has the potential to significantly improve the accuracy of the FD system while minimizing the need for labeled data. Furthermore, it can enable financial institutions to detect fraud in real-time, preventing monetary losses and ensuring customer safety.

ACKNOWLEDGEMENT

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

REFERENCES

Ahmed, M., Naser Mahmood, A. and Hu, J. (2016) 'A survey of network anomaly detection techniques', *Journal of Network and Computer Applications, 60, pp. 19–31*. Available at: https://doi.org/10.1016/J.JNCA.2015.11.016.

Alarab, I., Prakoonwit, S. and Nacer, M.I. (2020) 'Comparative Analysis Using Supervised Learning Methods for Anti-Money Laundering in Bitcoin', *ACM International Conference Proceeding Series, pp. 11–17*. Available at: https://doi.org/10.1145/3409073.3409078.

Barata, R. *et al*. (2021) 'Active learning for imbalanced data under cold start', *ICAIF 2021 - 2nd ACM International Conference on AI in Finance* [Preprint]. Available at: https://doi.org/10.48550/arxiv.2107.07724.

Cai, W., Zhang, M. and Zhang, Y. (2017) 'Batch mode active learning for regression with expected model change', *IEEE Transactions on Neural Networks and Learning Systems, 28(7), pp. 1668–1681*. Available at: https://doi.org/10.1109/TNNLS.2016.2542184.

Carcillo, F. e*t al.* (2017) 'An assessment of streaming active learning strategies for real-Life credit card fraud detection', *Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017, 2018-Janua, pp. 631–639*. Available at: https://doi.org/10.1109/DSAA.2017.10.

Domingues, R. *et al*. (2018) 'A comparative evaluation of outlier detection algorithms: Experiments and analyses', *Pattern Recognition, 74, pp. 406–421*. Available at: https://doi.org/10.1016/J.PATCOG.2017.09.037.

Hilal, W., Gadsden, S.A. and Yawney, J. (2022) 'Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances', *Expert Systems with Applications, 193, p. 116429*. Available at: https://doi.org/10.1016/J.ESWA.2021.116429.

Ibrahim, R.F., Elian, A.M. and Ababneh, M. (2021) 'Illicit Account Detection in the Ethereum Blockchain Using Machine Learning', *2021 International Conference on Information Technology (ICIT), pp. 488–493*. Available at: https://doi.org/10.1109/ICIT52682.2021.9491653.

Injadat, M. *et al.* (2018) 'Bayesian Optimization with Machine Learning Algorithms Towards Anomaly Detection', *2018 IEEE Global Communications Conference, GLOBECOM 2018 - Proceedings* [Preprint]. Available at: https://doi.org/10.1109/GLOCOM.2018.8647714.

Kee, S., del Castillo, E. and Runger, G. (2018) 'Query-by-committee improvement with diversity and density in batch active learning', *Information Sciences, 454–455, pp. 401–418*. Available at: https://doi.org/10.1016/J.INS.2018.05.014.

Labanca, D. *et al*. (2022) 'Amaretto: An Active Learning Framework for Money Laundering Detection', *IEEE Access, 10, pp. 41720–41739.* Available at: https://doi.org/10.1109/ACCESS.2022.3167699.

Lorenz, J. *et al*. (2020) 'Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity', *ICAIF 2020 - 1st ACM International Conference on AI in Finance* [Preprint]. Available at: https://doi.org/10.1145/3383455.3422549.

Melo-Acosta, G.E., Duitama-Munoz, F. and Arias-Londono, J.D. (2017) 'Fraud detection in big data using supervised and semi-supervised learning techniques', 2017 IEEE Colombian Conference on Communications and Computing, COLCOM 2017 - Proceedings [Preprint]. Available at: https://doi.org/10.1109/COLCOMCON.2017.8088206.

Monamo, P.M., Marivate, V. and Twala, B. (2017) 'A Multifaceted Approach to Bitcoin Fraud Detection: Global and Local Outliers', pp. 188–194. Available at: https://doi.org/10.1109/ICMLA.2016.0039.

Niu, X., Wang, L. and Yang, X. (2019) 'A Comparison Study of Credit Card Fraud Detection: Supervised versus Unsupervised'. Available at: https://doi.org/10.48550/arxiv.1904.10604.

Ostapowicz, M. and Żbikowski, K. (2019) 'Detecting Fraudulent Accounts on Blockchain: A Supervised Approach', ArXiv, *11881 LNCS, pp. 18–31*. Available at: https://doi.org/10.1007/978-3-030-34223-4_2.

Ounacer, S. *et al*. (2018) 'Using Isolation Forest in anomaly detection: The case of credit card transactions', *Periodicals of Engineering and Natural Sciences, 6(2), pp. 394–400.* Available at: https://doi.org/10.21533/PEN.V6I2.533.

Settles, B. (2012) 'Active Learning', *Synthesis Lectures on Artificial Intelligence and Machine Learning, 18, pp. 1–111*. Available at: https://doi.org/10.2200/S00429ED1V01Y201207AIM018.

Sharma, M. and Bilgic, M. (2017) 'Evidence-based uncertainty sampling for active learning', *Data Mining and Knowledge Discovery, 31(1), pp. 164–202*. Available at: https://doi.org/10.1007/S10618-016-0460-3/TABLES/9.

Weber, M. *et al*. (2019) 'Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics'. Available at: https://doi.org/10.48550/arxiv.1908.02591.

Xuan, S. *et al.* (2018) 'Random forest for credit card fraud detection', *ICNSC 2018 - 15th IEEE International Conference on Networking, Sensing and Control, pp. 1–6*. Available at: https://doi.org/10.1109/ICNSC.2018.8361343.

Zengan, G. (2009) 'Application of cluster-based local outlier factor algorithm in anti-money laundering', Proceedings - *International Conference on Management and Service Science, MASS 2009* [Preprint]. Available at: https://doi.org/10.1109/ICMSS.2009.5302396.