ПРОЗОРІСТЬ РІШЕНЬ ГРАФОВИХ НЕЙРОННИХ МЕРЕЖ У ЗАДАЧАХ ВИЯВЛЕННЯ ТРАНЗАКЦІЙНИХ ЗАГРОЗ У БЛОКЧЕЙН-СИСТЕМАХ

Просолов В.В.
Харківський національний університет радіоелектроніки, Харків, Україна

Сучасні блокчейн-системи характеризуються високою інтенсивністю транзакцій, псевдоанонімністю учасників та складними механізмами обфускації фінансових потоків. Це створює сприятливе середовище для реалізації схем відмивання коштів, ransomware-активності, мікшування та інших транзакційних аномалій.

Класичні методи виявлення загроз, що базуються на статичних евристиках або табличних алгоритмах машинного навчання, не дозволяють повноцінно враховувати топологічний контекст блокчейн-мережі. У зв'язку з цим графові нейронні мережі, зокрема GATv2, є більш придатними для аналізу ончейн-даних, однак їх практичне застосування стримується проблемою «чорної скриньки». У сферах кібербезпеки, фінансового моніторингу та цифрової криміналістики недостатньо лише отримати ризик-скоринг — необхідно також пояснити, які саме ознаки та зв'язки в графі стали підставою для класифікації транзакції як підозрілої [1, 2].

Метою роботи є підвищення прозорості та доказовості рішень нейромережевих класифікаторів у задачах виявлення транзакційних загроз у блокчейн-системах шляхом синергетичного поєднання локальних адитивних пояснень SHAP із ваговими коефіцієнтами шарів графової уваги GATv2 для формування структурованого вектора доказів, придатного до використання в аудиті, AML/CFT-процедурах і цифрових форензик-розслідуваннях.

У роботі як математичну основу використано апарат теорії графів, методи глибинного навчання, графові нейронні мережі з механізмами динамічної уваги, алгоритми балансування даних у латентному просторі та інструменти пояснювального штучного інтелекту. Запропонований підхід ґрунтується на аналізі транзакційного графа Bitcoin, у якому вузли описуються розширеним простором ознак, що включає 94 локальні мікроознаки та 72 топологічні макроознаки [3]. Такий 166-вимірний простір дозволяє моделі враховувати як фінансові параметри транзакцій, так і їх структурне оточення.

Ключовим результатом є удосконалення механізму математичної інтерпретації рішень графового класифікатора. На відміну від традиційних XAI-підходів, де пояснення формуються переважно на рівні окремих ознак, у запропонованому методі пояснення будуються одночасно у двох площинах. Перша площина — атрибутивна, у якій SHAP визначає внесок кожної окремої ознаки у фінальне рішення моделі. Друга площина — топологічна, у якій коефіцієнти уваги GATv2 відображають вагомість конкретних сусідів і ребер у процесі агрегації інформації [4, 5]. Поєднання цих двох механізмів дозволяє трансформувати фінальний ризик-скоринг моделі у структурований evidence vector, де кожен елемент пояснення має як числову, так і структурну інтерпретацію.

Практичне значення такого підходу полягає в тому, що аналітик безпеки отримує не лише оцінку ризику, а й пояснення причин, чому певний вузол або підграф було віднесено до класу підозрілих. Це особливо важливо для цифрової криміналістики, де потрібно аргументовано показати, які саме фінансові атрибути, часові характеристики або топологічні зв'язки стали тригерами класифікації.

Наприклад, SHAP може виділити найбільш впливові ознаки транзакції, а ваги уваги — вказати на підозрілі fan-in, peel chains чи інші аномальні маршрути руху активів у графі.

Таким чином забезпечується підвищення довіри до системи, спрощується аудит рішень та зменшується бар'єр до впровадження графових нейромереж у реальні SOC- та AML/CFT-середовища.

Додатково у роботі показано, що пояснюваність не суперечить високій якості класифікації. Запропонований метод, реалізований у складі програмного комплексу «AI Crypto Scanner», демонструє F1-score 0.8948 та Recall 0.9105 на еталонному Elliptic Dataset, що перевищує результати базових моделей. Отже, інтеграція XAI у графову архітектуру не лише зберігає високу ефективність виявлення аномалій, а й підвищує практичну цінність моделі за рахунок доказовості її рішень.

Запропонований підхід дозволяє підвищити прозорість рішень графових нейронних мереж у задачах виявлення транзакційних загроз у блокчейн-системах.

Поєднання SHAP-пояснень із ваговими коефіцієнтами графової уваги забезпечує формування структурованого вектора доказів, придатного для аудиту, цифрової криміналістики та систем фінансового моніторингу. Це підвищує практичну цінність моделі та розширює можливості її застосування в реальних умовах.

Список літератури

1. Veličković P., Cucurull G., Casanova A., Romero A., Liò P., Bengio Y. Graph Attention Networks // *International Conference on Learning Representations (ICLR).* 2018.

2. Brody S., Alon U., Yahav E. How Attentive are Graph Attention Networks? // *International Conference on Learning Representations (ICLR).* 2022.

3. Kushnerov, O., Prosolov, V., Dudykevych, V., Yevseiev, S., Ivanchenko, Y., Gorbulyk, V., ... & Sukhoteplyi, V. (2026). DEVELOPMENT OF A METHOD FOR IMPROVING THE EFFICIENCY OF TRANSACTION CLASSIFICATION IN THE BITCOIN NETWORK USING AN ATTENTION MECHANISM IN GRAPH NEURAL NETWORKS. *Eastern-European Journal of Enterprise Technologies*, №9. – P. 6-19. DOI: 10.15587/1729-4061.2026.351685

4. Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // *Advances in Neural Information Processing Systems.* 2017. Vol. 30.

5. Zhao T., Zhang X., Wang S., et al. GraphSMOTE: Imbalanced Node Classification on Graphs with Graph Neural Networks // *Proceedings of the 14th ACM International Conference on Web Search and Data Mining.* 2021.