

П.В. Травничева, А.В. Каторжевский

*Витебский государственный университет им. П.М. Машерова,
г. Витебск, Беларусь*

АЛГОРИТМЫ КЛАСТЕРИЗАЦИИ В БЛОКЧЕЙН

Блокчейн – это децентрализованная база данных, в которой транзакции записываются в виде цепочки блоков [1]. Одной из ключевых особенностей данной технологии является анонимность: пользователи взаимодействуют через публичные адреса, не привязанные напрямую к их личным данным. Однако с развитием методов анализа блокчейна стало возможным идентифицировать группы связанных кошельков, выявлять транзакционные паттерны и даже деанонимизировать пользователей.

Основной инструмент для этого – алгоритмы кластеризации, которые позволяют находить закономерности в блокчейн-данных и группировать объекты с общими характеристиками.

Эвристические методы кластеризации

Эвристические методы основаны на логических предположениях о поведении пользователей. Один из наиболее известных – метод общего входа, который предполагает, что если несколько входов принадлежат одной транзакции, то, вероятно, они контролируются одним владельцем. Однако данный метод уязвим перед такими техниками анонимизации, которые смешивают входы разных пользователей, усложняя их идентификацию.

Кластеризация на основе разбиения

Разбиение данных на группы – один из наиболее распространённых подходов в машинном обучении. К числу наиболее популярных алгоритмов этого типа относятся:

- **K-means** – итеративный метод, который разделяет данные на k групп, минимизируя внутрикластерное расстояние. В контексте блокчейна он может применяться для выявления сегментов пользователей по их транзакционной активности.

- **DBSCAN** – алгоритм плотностной кластеризации, который позволяет обнаруживать аномальные транзакции и выбросы.

Он эффективен при анализе нерегулярных сетевых структур, таких как блокчейн [2].

Графовые методы кластеризации

Блокчейн можно представить в виде графа, где узлы – это адреса кошельков, а рёбра – транзакции между ними. Для анализа таких структур используются:

- **Алгоритм Louvain** – метод разбиения графов, основанный на оптимизации модульности. Применяется для выявления сообществ и групп аффилированных адресов.

- **Label Propagation Algorithm** – алгоритм, который распространяет метки по графу, позволяя находить естественные кластеры без необходимости предварительно задавать их количество.

- **PageRank** – метод, изначально разработанный для ранжирования веб-страниц, но применяемый в блокчейн-анализе для выявления наиболее значимых узлов в сети (например, бирж и крупных кошельков) [3].

Выбор метода кластеризации зависит от множества факторов: структуры данных, наличия шумов и вычислительной сложности. Оценка качества кластеризации в блокчейне затруднена из-за отсутствия размеченных данных. Однако используются следующие метрики:

- **Silhouette Score** – измеряет плотность кластеров и качество разбиения данных.

- **Modularity** – применяется для оценки качества разбиения графов.

- **Adjusted Rand Index** – позволяет оценить схожесть полученных кластеров с эталонными данными, если таковые имеются.

Кластеризация блокчейн-данных – перспективное направление, помогающее в анализе транзакций, борьбе с мошенничеством и соблюдении регуляторных требований. Несмотря на достигнутый прогресс, остаются нерешенные задачи, связанные с повышением точности и эффективности этих методов, особенно в условиях использования техник повышения анонимности. Дальнейшие исследования должны быть направлены на разработку гибридных подходов, объединяющих различные методы анализа данных, а также на адаптацию существующих алгоритмов к специфике блокчейн-технологий.

Литература

1. *Башир, И.* Блокчейн. Исчерпывающее руководство / пер. с англ. С.В. Черникова и В.И. Бахура. М.: ДМК Пресс, 2025. 894 с.: ил.
2. *Лесковец Ю., Ульман Дж.Д., Раджараман А.* Машинное обучение. Анализ данных: учебник / ред. Д.А. Мовчан; пер. А.А. Слинкин. М.: ДМК Пресс, 2016. 498 с.
3. *Рафгарден Т.* Совершенный алгоритм. Графовые алгоритмы и структуры данных. СПб.: Питер, 2019. 256 с.: ил. (Серия «Библиотека программиста»).