

УДК 004.891.2:656.1

АРХИТЕКТУРА MAS-XAI ДЛЯ ИНТЕРПРЕТИРУЕМОГО УПРАВЛЕНИЯ ТРАНСПОРТНЫМИ ПОТОКАМИ

Хаджинова Н. В., Михнюк А. И.

*Белорусский государственный университет информатики
и радиоэлектроники*

E-mail: khajynova@bsuir.by

Summary. *This study develops a framework for integrating Multi-Agent System (MAS) with Explainable AI (XAI) to address interpretability challenges in traffic management. The proposed three-layer framework implements local XAI modules (LIME, SHAP) at the agent level, explanation aggregation at the coordinator level, and human-machine interaction at the user level. Framework is expected to enhance operators trust and enable more effective traffic management in complex urban environments.*

Актуальность разработки интеллектуальных систем управления дорожным движением обусловлена растущей сложностью транспортных сетей, где традиционные подходы демонстрируют ограниченную эффективность [1]. Применение мультиагентных систем и алгоритмов искусственного интеллекта сталкивается с проблемой интерпретируемости решений. Научная новизна работы заключается в разработке концепции интеграции MAS-XAI и трехуровневой архитектуры, обеспечивающей сквозную интерпретируемость решений. Практическая значимость состоит в создании прозрачных систем управления, позволяющих операторам анализировать и корректировать работу алгоритмов. Ожидаемые результаты внедрения включают повышение доверия к автоматизированным системам и обеспечение контроля в городской инфраструктуре.

Сложность интерпретируемости в мультиагентных системах управления транспортным потоком обусловлена незапланированным поведением. Глобальные паттерны в системах возникают в результате нелинейных взаимодействий автономных агентов [2]. Это исключает прямую трассировку причинно-следственных связей между решениями и последствиями. Дополнительную сложность вносит конфликт интересов агентов. Когда оптимизация локальных целевых функций приводит к субоптимальным результатам на системном уровне, возникают барьеры для практического внедрения.

В качестве методологического ответа на вызовы интерпретируемости выступает направление объяснимого искусственного интеллекта (XAI). XAI представляет собой совокупность методов, направленных на обеспечение прозрачности решений, принимаемых алгоритмами. В контексте мультиагентных систем особую значимость приобретают методы последующей интерпретации.

К числу наиболее релевантных подходов относится метод LIME (Local Interpretable Model-agnostic Explanations), который аппроксимирует

сложную модель локально для конкретного предсказания. А метод SHAP (SHapley Additive exPlanations) обеспечивает теоретически обоснованный вклад каждого признака на основе теории игр. Каждый агент, принимая решение, использует эти методы для генерации объяснений. Они выделяют ключевые факторы, повлиявшие на выбор.

Предлагаемая архитектура гибридной системы MAS-XAI реализует трехуровневую модель. Базовый уровень составляют интеллектуальные агенты, оснащенные модулями XAI. Координационный уровень выполняет обобщение локальных объяснений и синтез глобальной картины поведения системы. На этом уровне выявляются системные паттерны, вызванные взаимодействием агентов. Пользовательский уровень обеспечивает человеко-машинное взаимодействие через интерфейс визуализации. Оператор получает возможность анализировать визуализацию причинно-следственных связей и влияние факторов на глобальную ситуацию.



Рисунок 1 – Архитектурная схема гибридной системы

Проведенное исследование демонстрирует возможность интеграции мультиагентных систем с методами объяснимого искусственного интеллекта. Предложенная архитектура позволяет обеспечить сквозную интерпретируемость решений. Это открывает возможности для принятия обоснованных решений операторами. Предложенный подход открывает путь к созданию надежных интеллектуальных транспортных систем. Дальнейшие исследования предполагают разработку специализированных протоколов обмена объяснениями между агентами и экспериментальную проверку подхода на комплексных транспортных моделях городов.

Список использованных источников

1. Roess, R. P. Traffic Engineering / R. P. Roess, E. S. Prassas, W. R. McShane. – 5th ed. – London : Pearson, 2021.
2. Belle, V. Principles and Practice of Explainable Machine Learning / V. Belle, I. Pantoni // Frontiers in Big Data. – 2021. – Vol. 4. – Art. 688969.