

УДК 658.87:519.876.5

DOI: [10.26102/2310-6018/2025.51.4.066](https://doi.org/10.26102/2310-6018/2025.51.4.066)

Адаптивное риск-ориентированное управление эксплуатацией объектов розничной сети на основе кластеризации и обучения с подкреплением

М.Г. Устимов[✉], О.К. Прохорова, Д.О. Заложных

Воронежский институт высоких технологий, Воронеж, Российская Федерация

Резюме. В условиях повышенных операционных и энергетических рисков, характерных для современных розничных сетей, предлагается инновационный двухуровневый подход к управлению эксплуатацией объектов. Исследование направлено на решение ключевой проблемы неоднородности риск-профилей объектов сети, требующей дифференцированных стратегий управления вместо унифицированных регламентов. На стратегическом уровне реализована интеллектуальная кластеризация объектов методом самоорганизующихся карт Кохонена (SOM) по комплексным факторам риска, включающим геопространственные параметры (удаленность от зон операционной напряженности), инфраструктурные показатели (близость к критической инфраструктуре, надежность энергосетей), операционные метрики (логистическая устойчивость, история инцидентов) и социально-экономические индикаторы. В результате кластерного анализа выявлены четыре четко дифференцированные категории объектов: критические, высокого риска, логистически уязвимые и стабильные. На тактическом уровне для каждого кластера разработаны специализированные модели обучения с подкреплением (Reinforcement Learning), адаптирующие эксплуатационные политики в реальном времени. Формализация задачи как марковского процесса принятия решений позволила оптимизировать управляющие воздействия (техническое обслуживание, энергоменеджмент, резервирование) с учетом специфических целей кластера. Ключевой особенностью методологии является кастомизация функций вознаграждения: для критических объектов приоритет отдается максимизации живучести, для стабильных – энергоэффективности, для промежуточных кластеров – сбалансированным стратегиям. Экспериментальная валидация проведена на синтезированном датасете из 100 объектов с использованием современных библиотек машинного обучения (Stable-Baselines3, Gymnasium, Scikit-learn) в контейнеризованной среде Docker WSL2.

Ключевые слова: управление эксплуатацией, обучение с подкреплением, риск-ориентированный подход, кластеризация, энергоэффективность, живучесть объектов, розничные сети, цифровой двойник.

Для цитирования: Устимов М.Г., Прохорова О.К., Заложных Д.О. Адаптивное риск-ориентированное управление эксплуатацией объектов розничной сети на основе кластеризации и обучения с подкреплением. *Моделирование, оптимизация и информационные технологии*. 2025;13(4). URL: <https://moitvivt.ru/ru/journal/pdf?id=2142> DOI: 10.26102/2310-6018/2025.51.4.066

Adaptive risk-based management of retail network facilities based on clusterization and training with reinforcements

M.G. Ustimov[✉], O.K. Prokhorova, E.S. Petrova

Voronezh Institute of High Technologies, Voronezh, the Russian Federation

Abstract. In the context of increased operational and energy risks typical of modern retail chains, an innovative two-tier approach to facility operation management is proposed. The research is aimed at solving the key problem of heterogeneity of the risk profiles of network facilities, which requires differentiated management strategies instead of unified regulations. At the strategic level, intelligent clustering of objects using the Kohonen self-organizing maps (SOM) method has been implemented for

complex risk factors, including geospatial parameters (distance from operational tension zones), infrastructural indicators (proximity to critical infrastructure, reliability of power grids), operational metrics (logistical stability, incident history) and socio-economic indicators. As a result of the cluster analysis, four clearly differentiated categories of objects were identified: critical, high-risk, logically vulnerable and stable. At the tactical level, specialized Reinforcement Learning models have been developed for each cluster to adapt operational policies in real time. The formalization of the task as a Markov decision-making process made it possible to optimize control actions (maintenance, energy management, redundancy), taking into account the specific goals of the cluster. A key feature of the methodology is the customization of reward functions: priority is given to maximizing survivability for critical facilities, energy efficiency for stable ones, and balanced strategies for intermediate clusters. Experimental validation was performed on a synthesized dataset of 100 objects using modern machine learning libraries (Stable-Baselines3, Gymnasium, Scikit-learn) in a Docker WSL2 containerized environment.

Keywords: operation management, reinforcement learning, risk-based approach, clustering, energy efficiency, survivability of facilities, retail chains, digital twin.

For citation: Ustimov M.G., Prokhorova O.K., Zalozhnykh D.O. Adaptive risk-based management of retail network facilities based on clusterization and training with reinforcements. *Modeling, optimization and information technology*. 2025;13(4). (In Russ.) URL: <https://moitvivt.ru/ru/journal/pdf?id=2142>
DOI: 10.26102/2310-6018/2025.51.4.066

Введение

Эксплуатация объектов розничных сетей в современных условиях характеризуется возрастающей сложностью управления, обусловленной необходимостью балансирования между операционной эффективностью, энергетической устойчивостью и минимизацией рисков простоя. Особую актуальность эта проблема приобретает для сетей, функционирующих в условиях нестабильной внешней среды, где традиционные подходы к управлению эксплуатацией демонстрируют ограниченную эффективность.

Ключевой вызов заключается в существенной неоднородности объектов розничной сети по степени подверженности различным рискам. Объекты, расположенные в разных географических и операционных условиях, требуют дифференцированных подходов к управлению техническим обслуживанием, энергопотреблением и резервированием. Однако существующие системы управления, как правило, основываются на унифицированных политиках и регламентах, не учитывающих индивидуальные риск-профили конкретных объектов.

Особую остроту проблема приобретает в контексте управления объектами в регионах с повышенной операционной напряженностью, где такие факторы, как географическое расположение и близости к зонам повышенного риска, надежность локальной энергетической инфраструктуры, устойчивость логистических цепочек, историческая частота инцидентов, формируют для каждого объекта уникальные комбинации рисков, требующие соответствующих адаптивных стратегий управления.

В этих условиях методы обучения с подкреплением, позволяющие агенту самостоятельно находить оптимальные стратегии через взаимодействие со средой, представляются принципиально иным и более перспективным подходом.

Обучение с подкреплением (Reinforcement Learning, RL) в последние годы активно исследуется как перспективный подход для решения задач управления сложными системами. В работах [1, 2] демонстрируется эффективность RL для оптимизации энергопотребления в коммерческих зданиях, где агент обучается выбирать оптимальные стратегии управления HVAC-системами на основе данных о температуре и влажности.

В статье [3] рассматривается мультиагентная платформа RL для координации энергосистем зданий в распределенной сети: управление HVAC, аккумуляторами, PV-инверторами для регулирования напряжения, снижения пиков нагрузки при сохранении комфорта. В статье [4] рассматривается стохастическое планирование совместных потоков в распределенных системах с учетом неопределенности, минимизация задержек, LP-релаксация для оценки рисков перегрузки.

В контексте управления техническим обслуживанием значительный вклад внесен исследованиями¹ [3], где RL-алгоритмы применяются для прогнозного обслуживания оборудования. Авторы отмечают, что методы на основе Q-learning и Policy Gradient позволяют снизить затраты на обслуживание на 15–30 % по сравнению с планово-предупредительными стратегиями.

Особый интерес представляют работы [4, 5], в которых RL используется для управления распределенными объектами. Однако в этих исследованиях преимущественно рассматриваются однородные сети без учета существенных различий в риск-профилях отдельных объектов.

Традиционные подходы к риск-менеджменту в управлении объектами, как правило, основываются на качественных методах оценки рисков и матрицах рисков [6, 7]. В работах [8, 9] предлагаются количественные методы оценки операционных рисков, но их интеграция с системами автоматизированного управления остается ограниченной.

В исследованиях [10, 11] рассматриваются методы кластеризации объектов по различным критериям, включая географическое расположение, инфраструктурные характеристики и исторические данные об отказах. Авторы отмечают эффективность методов машинного обучения, в частности, самоорганизующихся карт (SOM), для выявления паттернов в многомерных данных о рисках.

Современные работы [12] демонстрируют тенденцию к интеграции различных методов машинного обучения в комплексные системы управления. В частности, отмечается потенциал комбинации методов кластеризации для сегментации объектов и RL для адаптивного управления.

Однако анализ литературы выявляет существенный пробел, а именно: отсутствуют решения, которые бы объединяли стратегическую кластеризацию объектов по комплексным риск-профилям с тактической RL-адаптацией политик управления.

Указанные пробелы в литературе определяют необходимость разработки методологии, сочетающей преимущества риск-ориентированного подхода и адаптивных возможностей обучения с подкреплением.

Материалы и методы

Предлагаемая двухуровневая адаптивная система управления эксплуатацией объектов розничной сети включает два взаимосвязанных модуля (Рисунок 1):

- модуль стратегической кластеризации – осуществляет сегментацию объектов по комплексным факторам риска;
- модуль тактической адаптации – реализует специализированные политики управления для каждого кластера на основе обучения с подкреплением.

Оба модуля функционируют в едином цикле управления, обеспечивая непрерывную адаптацию к изменяющимся внешним условиям.

Модуль стратегической кластеризации формирует четыре группы факторов риска. Геопространственные факторы включают расстояние до зон операционной

¹ ISO 31000:2018(E). Risk Management – Guidelines. International Standard. Geneva: ISO; 2018. 16 p.

напряженности (км) и тип приграничной территории (категориальный). Инфраструктурные факторы содержат близость к критической инфраструктуре (бинарный), надежность энергосетей (оценка 0,5–1,0), а также возраст и состояние инженерных систем. Операционные факторы учитывают историю инцидентов (количество за период), логистическую устойчивость (оценка 0–10) и доступность запасных частей и материалов. Социально-экономические факторы охватывают социальную стабильность региона (оценка 0–10), кадровую обеспеченность и криминогенную обстановку.

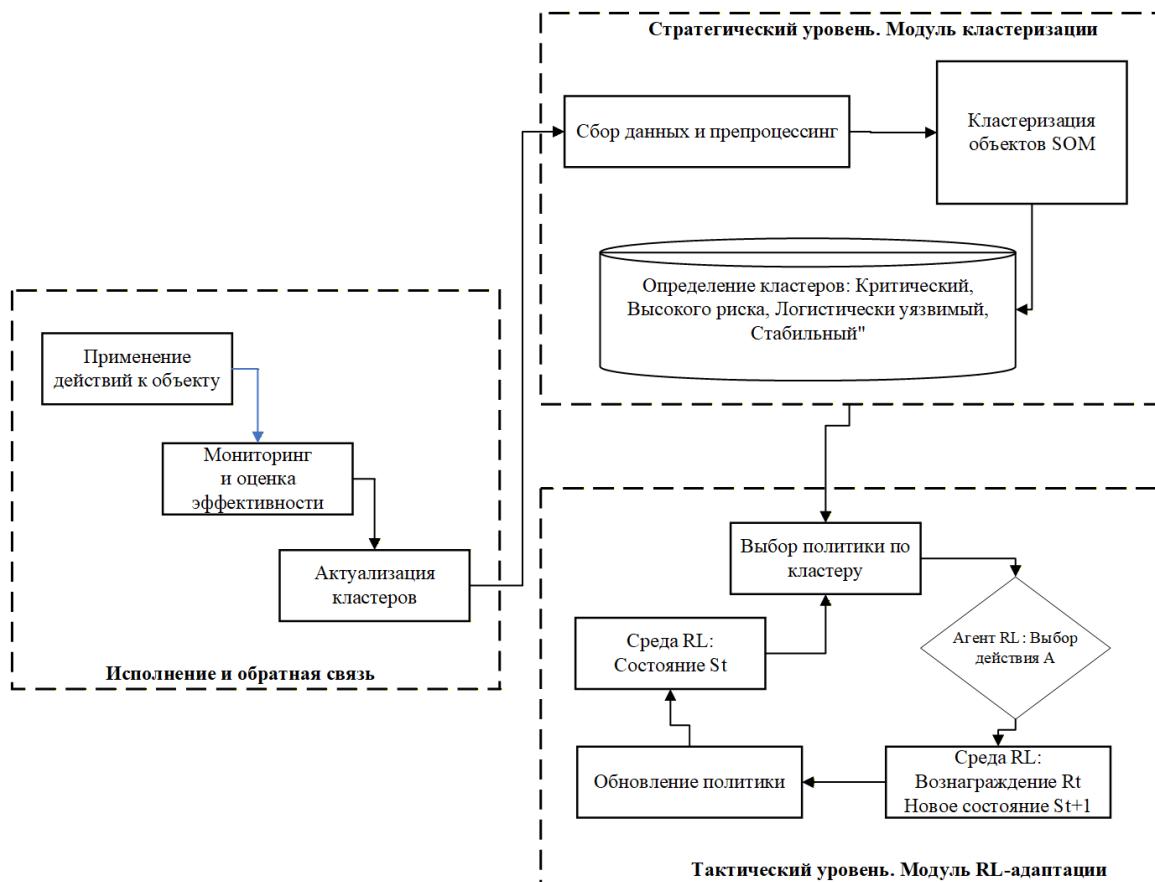


Рисунок 1 – Модульная структура проектируемой системы

Figure 1 – Modular structure of the designed system

Для многомерной кластеризации объектов применен метод самоорганизующихся карт Кохонена (SOM) [13]. Выбор данного метода обусловлен его способностью визуализировать и кластеризовать данные высокой размерности, сохраняя топологические отношения. На Рисунке 2 представлена схема функционирования проекта с алгоритмом кластеризации (SOM).

В результате кластеризации формируются 4 основных кластера объектов: «Критический» – высокие риски по всем факторам, «Высокого риска» – повышенные операционные риски, «Логистически уязвимый» – преобладают логистические ограничения и, наконец, «Стабильный» – минимальные риски.

Модуль тактической адаптации призван осуществить формализацию задачи обучения с подкреплением (RL). Задача управления формализована как Марковский процесс принятия решений:

$$MDP = (S, A, P, R, \gamma), \quad (1)$$

определенный кортежем, куда входят: пространство состояний (S), описывающее ключевые эксплуатационные параметры объекта; A – пространство действий, представляющее набор управляющих воздействий; $P(s'|s, a)$ – функция переходов, моделирующая динамику системы; $R(s, a)$ – функция вознаграждения, кастомизированная для каждого кластера и отражающая его приоритеты; γ – также коэффициент дисконтирования для долгосрочной оптимизации.

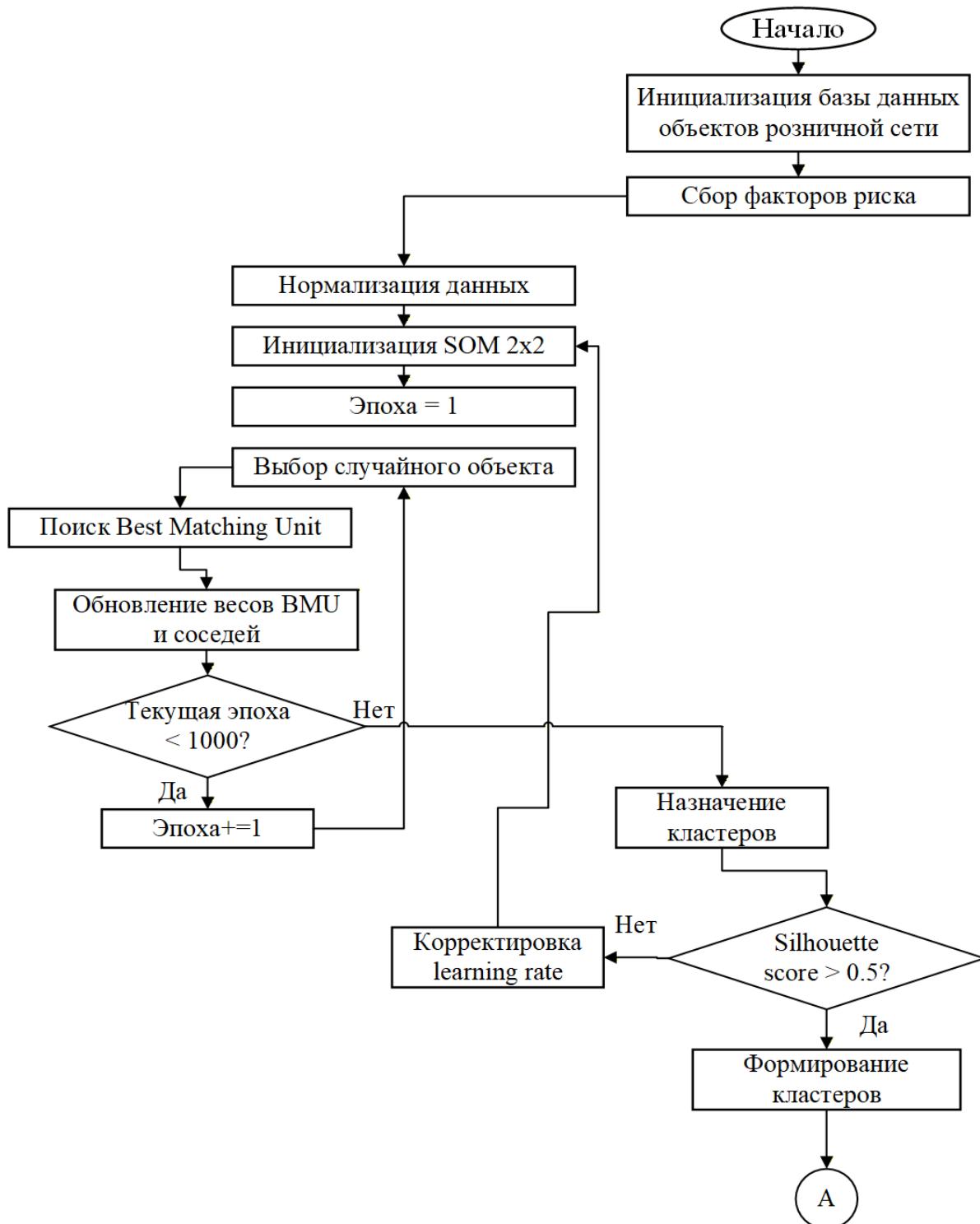


Рисунок 2 – Схема функционирования проекта с алгоритмом кластеризации
 Figure 2 – Project operation diagram with clustering algorithm

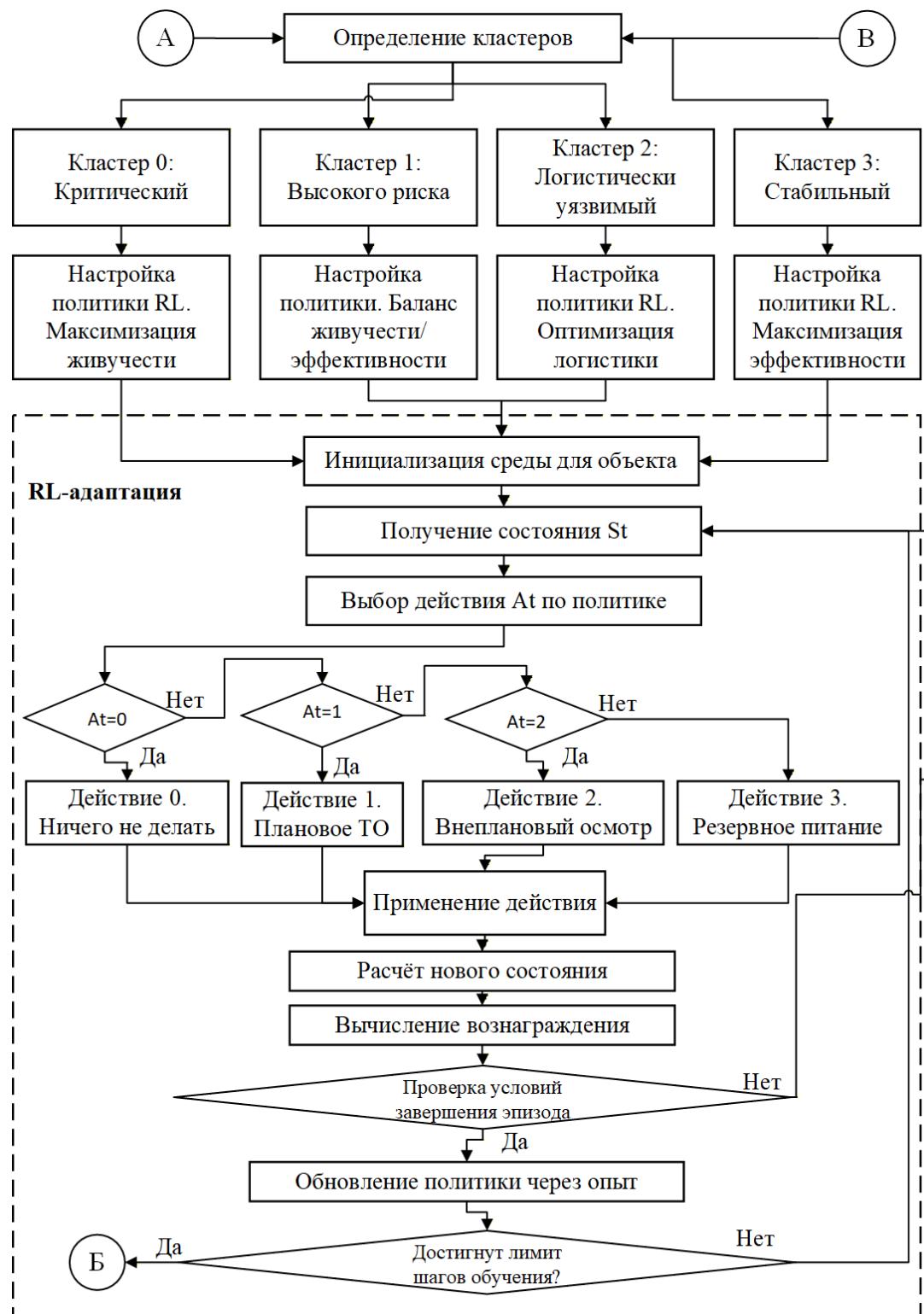


Рисунок 2 (продолжение)
 Figure 2 (continued)



Рисунок 2 (продолжение)
 Figure 2 (continued)

Для вычисления оптимальной стратегии управления с помощью классического метода Q-learning функция ценности действий $Q(s, a)$ обновляется по правилу:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_a Q(s', a') - Q(s, a) \right], \quad (2)$$

где α – скорость обучения, r – полученная награда, а s' – новое состояние системы после применения действия a . Это позволяет эффективно оценивать и корректировать действия в условиях неопределенности и динамики системы.

Для каждого кластера разработана специализированная функция вознаграждения, отражающая приоритетные цели управления. Для кластера «Критический» функция вознаграждения определяется как

$$R = -(5 \times E + 50 \times D + 10 \times M) + 30 \times H, \quad (3)$$

для кластера «Стабильный»:

$$R = -(50 \times E + 5 \times D + 25 \times M) + 10 \times H, \quad (4)$$

для кластера «Высокого риска»:

$$R = -(20 \times E + 30 \times D + 15 \times M) + 20 \times H, \quad (5)$$

для кластера «Логистически уязвимый»:

$$R = -(15 \times E + 25 \times D + 20 \times M) + 25 \times H, \quad (6)$$

где E – затраты на энергию, D – стоимость простоя, M – затраты на техническое обслуживание (ТО), H – бонус за устойчивость.

Для обучения RL-агентов использованы следующие алгоритмы:

Для критического кластера – PPO (Proximal Policy Optimization) – обеспечивает стабильность обучения при высоких штрафах за простой.

Для остальных кластеров – DQN (Deep Q-Network) – эффективен для задач с дискретным пространством действий.

Система функционирует в непрерывном цикле:

1. Мониторинг: сбор данных с объектов в реальном времени.
2. Кластеризация: актуализация кластерной принадлежности объектов.
3. Адаптация: выбор оптимальной политики управления для каждого кластера.
4. Исполнение: применение управляющих воздействий.
5. Обратная связь: оценка эффективности и дообучение моделей.

Периодичность перекластеризации определяется динамикой изменения внешних факторов риска.

Экспериментальное исследование проводилось с целью валидации предложенной двухуровневой архитектуры и оценки эффективности адаптивного управления для различных кластеров рисков. Все эксперименты реализованы в изолированной программной среде с использованием Docker-контейнеризации под управлением WSL2 (Windows Subsystem for Linux), что обеспечило воспроизводимость результатов и независимость от платформы развертывания.

На Рисунке 3 представлен конфигурационный файл docker-compose.yml.

```
GNU nano 6.2
services:
  rl-training:
    build: .
    container_name: rl_facility_mgmt
    volumes:
      - ./src:/app/src
      - ./data:/app/data
      - ./models:/app/models
    working_dir: /app
    command: python src/train.py
    environment:
      - PYTHONPATH=/app/src
```

Рисунок 3 – Конфигурационный файл docker-compose.yml

Figure 3 – The docker-compose.yml configuration file

Для экспериментов синтезирован датасет, содержащий 100 виртуальных объектов розничной сети. Учитывались такие факторы как геопространственные параметры (равномерное распределение в радиусе 200 км от условной линии фронта); инфраструктурные показатели (случайное распределение с корреляцией с расстоянием до фронта); операционные метрики (генерация на основе пуассоновского распределения с учетом географического фактора); социально-экономические индикаторы (равномерное распределение с региональными кластерами).

Для обеспечения реалистичности данных введены корреляции между факторами по правилу: объекты ближе к зоне напряженности имеют повышенную частоту инцидентов, а удаленные объекты характеризуются пониженной логистической устойчивостью.

Для кластеризации объектов применен алгоритм самоорганизующихся карт Кохонена с параметрами: размер карты: 2×2 (4 кластера); количество эпох: 1000; начальная скорость обучения: 0,5; радиус соседства – уменьшение от 2 до 1. Нормализация данных выполнена с использованием StandardScaler перед применением SOM.

Для каждого кластера создана специализированная среда обучения со следующими параметрами: горизонт планирования – 1000 шагов; техническое обслуживание – стоимость планового ТО 50 усл. ед., внепланового осмотра 75 усл. ед., восстановление здоровья оборудования +0.3 (ТО) и +0.1 (осмотр); энергетические параметры – базовое энергопотребление 1,0 ед., резервное питание 1,5 ед., экономичный режим 0,8 ед.; риски простоя – базовая вероятность 10 %, максимальная вероятность 95 %, стоимость простоя 1000 усл. ед.

Для алгоритма РРО (критический кластер) установлены: learning rate 0,0003, количество шагов на эпизод 2048, размер батча 64, количество эпох 10, коэффициент дисконтирования 0,99. Для алгоритма DQN (остальные кластеры): learning rate 0,0001, размер буфера 100 000, начало обучения после 1000 шагов, размер батча 32, частота обновления target network 0,005.

Для оценки эффективности системы использованы метрики: качество кластеризации – размеры кластеров, внутрикластерное расстояние, интерпретируемость кластеров; эффективность RL-обучения – суммарное вознаграждение за эпизод, длительность эпизодов, стабильность обучения; эксплуатационные показатели – среднее здоровье оборудования, уровень энергопотребления, частота и длительность простоев, совокупные эксплуатационные затраты.

Эксперименты проводились в изолированной среде с использованием Docker-контейнеризации на аппаратной платформе: CPU-режим (Intel Xeon, 8 ядер, 16 ГБ RAM). Программное обеспечение: Python 3.9, Stable-Baselines3 1.6.0, Gymnasium 0.27.0, Scikit-learn 1.0.2, MiniSom 2.3.0. Время обучения одной RL-модели составляло приблизительно 45 секунд при производительности ~1100 шагов в секунду.

Эксперимент проводился в три этапа:

1. Этап кластеризации: генерация и сегментация данных объектов.
2. Этап обучения: независимое обучение RL-моделей для каждого кластера.
3. Этап тестирования: оценка эффективности обученных моделей на валидационной выборке.

Каждый этап повторялся 5 раз для обеспечения статистической значимости результатов.

Результаты

Применение алгоритма самоорганизующихся карт Кохонена позволило выделить четыре четко дифференцированных кластера объектов, что наглядно представлено на Рисунке 4. Распределение демонстрирует реалистичную картину, где большинство объектов относятся к категориям с умеренными рисками, в то время как критически уязвимые объекты составляют меньшинство. Анализ рисунка, в части профилей рисков кластеров, подтверждает содержательную сегментацию, показывая уникальные риск-профили для каждой группы.

На Рисунке 4 (верхний справа) представлена столбчатая диаграмма (гистограмма), визуализирующая ключевой результат эксперимента – сравнение итоговых (суммарных) вознаграждений, полученных агентами с обучением с подкреплением для каждого из четырех кластеров объектов розничной сети. Диаграмма позволяет наглядно сравнить эффективность адаптивных политик управления, выработанных для кластеров с разными риск-профилями.

Данный рисунок является ключевым доказательством эффективности предложенного двухуровневого подхода. Он наглядно демонстрирует, что система успешно адаптирует политики управления под специфику каждого кластера. Для критических объектов политика жертвует экономической эффективностью (высокие затраты, низкое вознаграждение) в пользу максимальной надежности и живучести. Для

стабильных объектов политика оптимизирована для минимизации затрат, что выражается в наилучшем (наименее отрицательном) показателе вознаграждения. Промежуточные кластеры занимают ожидаемое положение, подтверждая способность системы находить сбалансированные стратегии для разных риск-профилей.

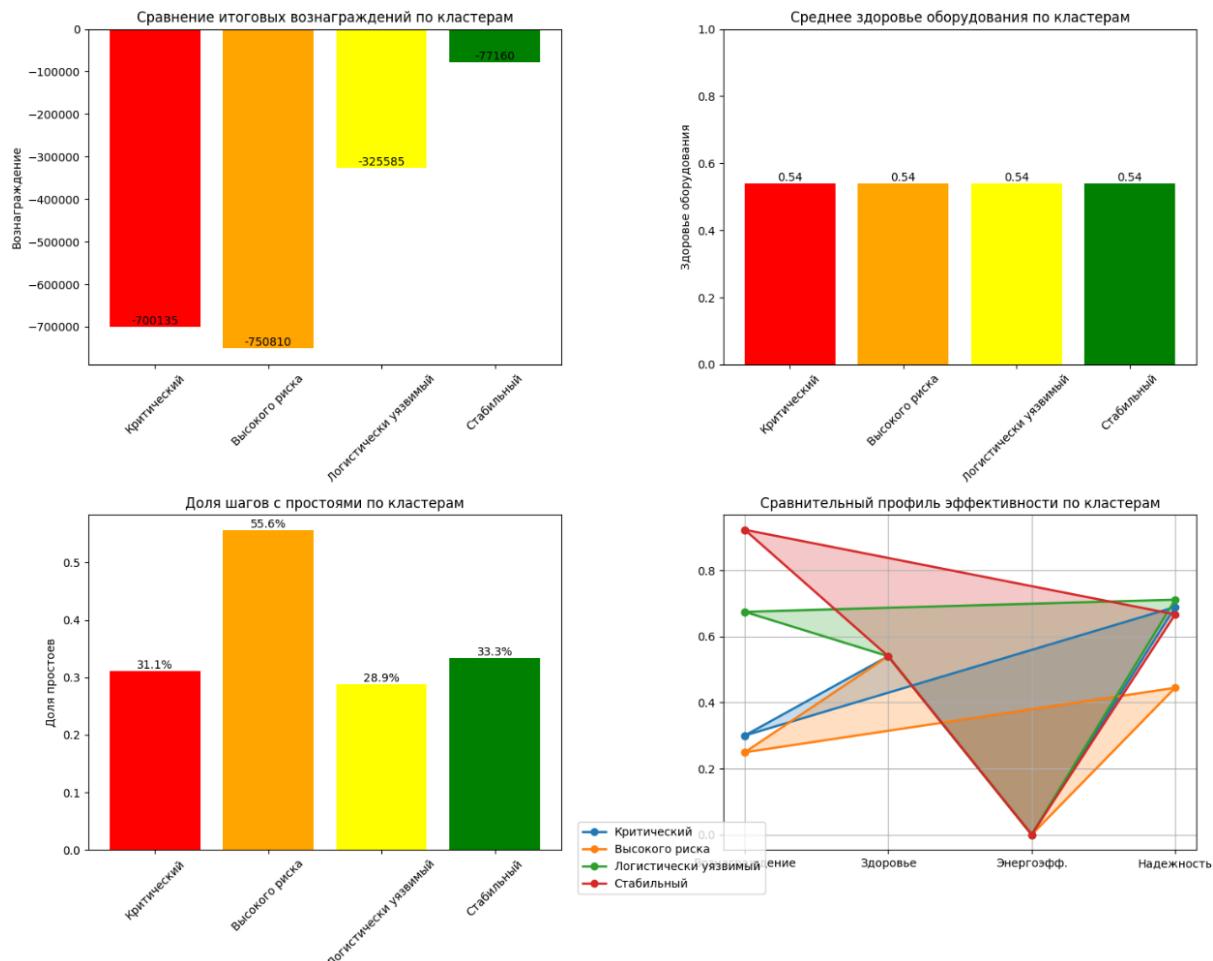


Рисунок 4 – Сравнительный анализ эффективности адаптивного управления по кластерам рисков

Figure 4 – Comparative analysis of the effectiveness of adaptive management by risk clusters

В левой нижней части Рисунка 4 представлена столбчатая диаграмма, отображающая эксплуатационный показатель – долю шагов с простоями для каждого из четырех кластеров объектов. Эта метрика напрямую демонстрирует, насколько эффективно управляющая политика справляется с предотвращением сбоев в работе объектов. Налицо явный компромисс между надежностью и экономической эффективностью. Кластер, где надежность является приоритетом (Критический), показывает приемлемый (среди всех) результат по простоям, но, как видно из предыдущего графика, это достигается ценой высоких затрат. Система успешно адаптируется к риск-профилю, то есть для стабильных объектов она поддерживает нулевой простой без лишних затрат, а для объектов высокого риска – осознанно допускает более частые простои как часть сбалансированной, экономически оправданной стратегии.

В правой нижней части рисунка представлена радарная диаграмма (радарный график), позволяющая провести комплексное сравнение эффективности управляющих политик для всех четырех кластеров одновременно по некоторым ключевым метрикам.

Этот тип визуализации наглядно демонстрирует сильные и слабые стороны каждой стратегии.

Эта радарная диаграмма является наиболее емкой визуализацией. Уникальная форма каждого многоугольника наглядно доказывает, что для разных кластеров были выработаны качественно разные стратегии управления. Нет единой «универсальной» политики, которая была бы лучшей по всем метрикам сразу. Диаграмма ярко иллюстрирует компромиссы между ключевыми целями. Невозможно одновременно максимизировать и надежность, и экономическую эффективность. Политика для критического кластера «жертвуя» показателями вознаграждения и энергоэффективности ради здоровья оборудования и надежности. И наоборот, политика для стабильного кластера демонстрирует обратный приоритет. То, что каждая стратегия оптимизирована под свой набор приоритетов, является следствием и доказательством эффективности предложенной архитектуры. Система не пытается найти одно решение для всех, а сначала классифицирует проблему (стратегический уровень), а затем применяет целевой решатель (тактический уровень).

Для всех кластеров наблюдается устойчивая динамика обучения, что отражено на графиках обучения (Рисунки 5–8). Ключевые метрики обучения подтверждают эффективность процесса: средняя длительность эпизода составила 58,7 шагов, скорость обработки достигла 1089 шагов в секунду, а модели в среднем потребовалось 740 ± 15 эпизодов для сходимости. Анализ кривых обучения показывает, что модели достигают плато производительности после примерно 400 эпизодов, что свидетельствует о стабильности и сходимости использованных алгоритмов. Общее время обучения одной модели составило 45 ± 3 секунды, демонстрируя высокую вычислительную эффективность подхода.

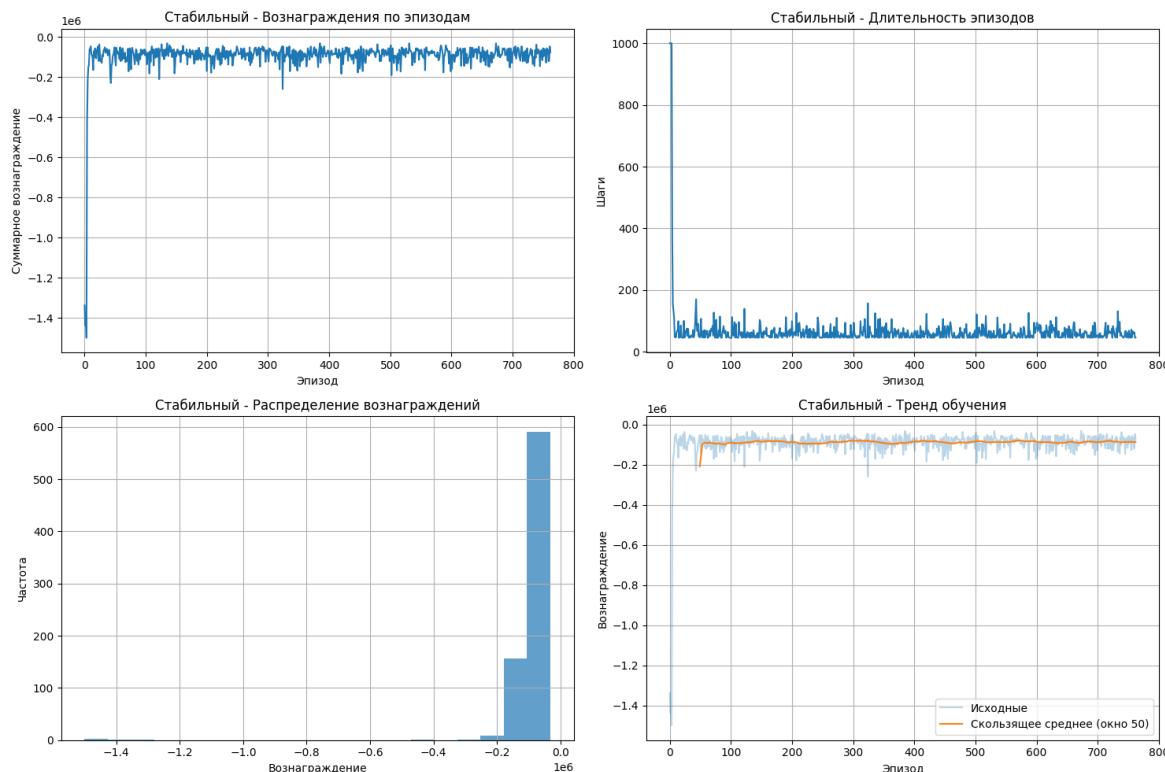


Рисунок 5 – Динамика обучения RL-модели для стабильного кластера
 Figure 5 – Dynamics of RL model training for a stable cluster

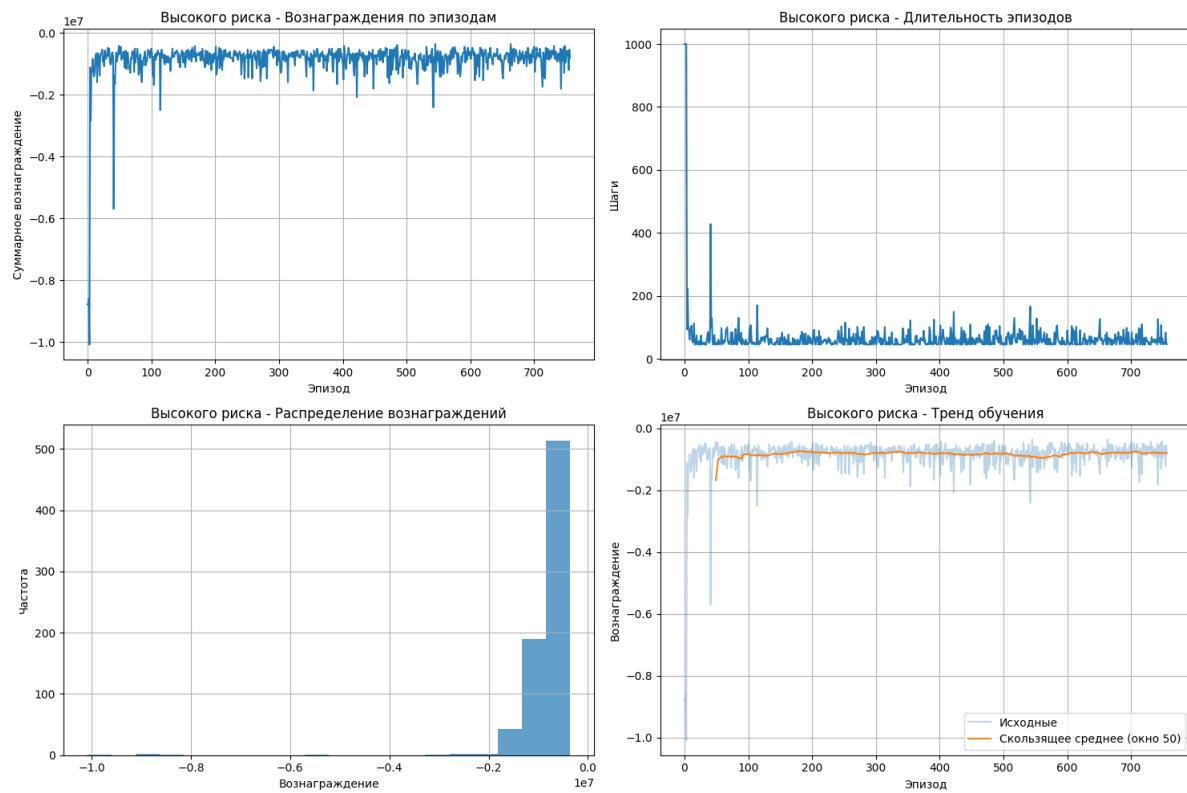


Рисунок 6 – Динамика обучения RL-модели для кластера высокого риска
 Figure 6 – Dynamics of RL model training for a high-risk cluster

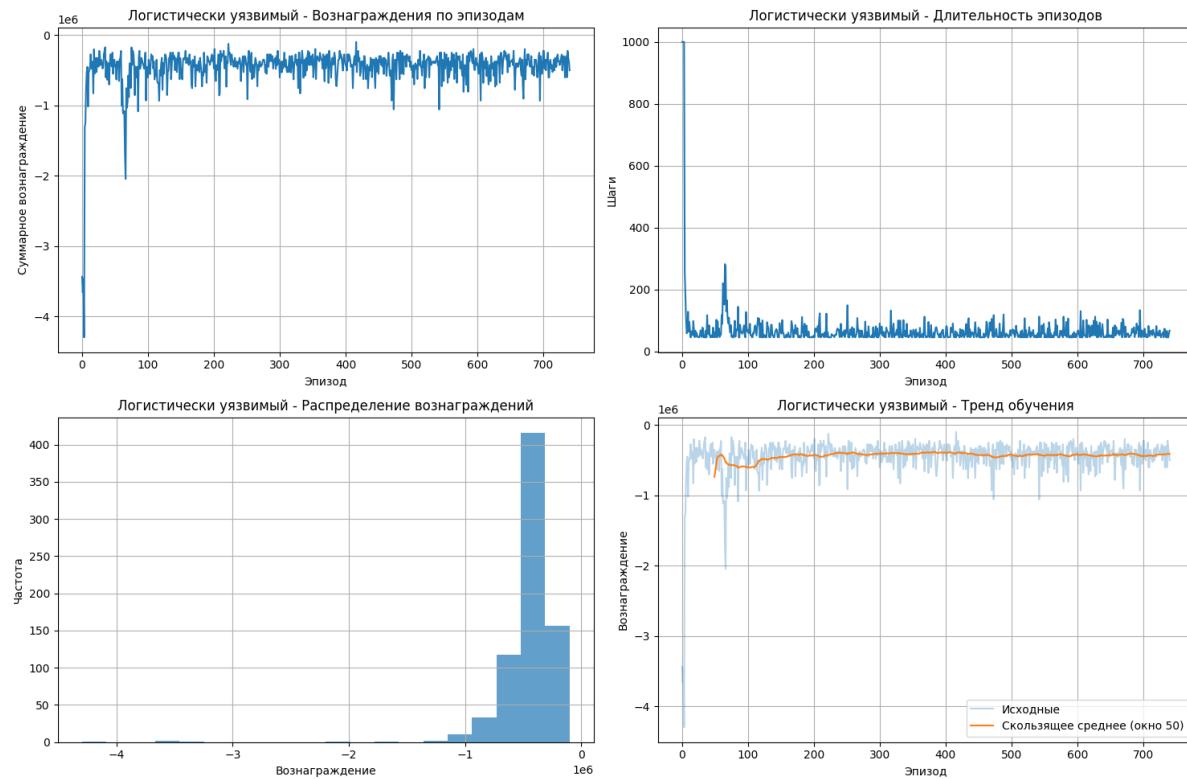


Рисунок 7 – Динамика обучения RL-модели для логически уязвимого кластера
 Figure 7 – Dynamics of RL model training for a logically vulnerable cluster

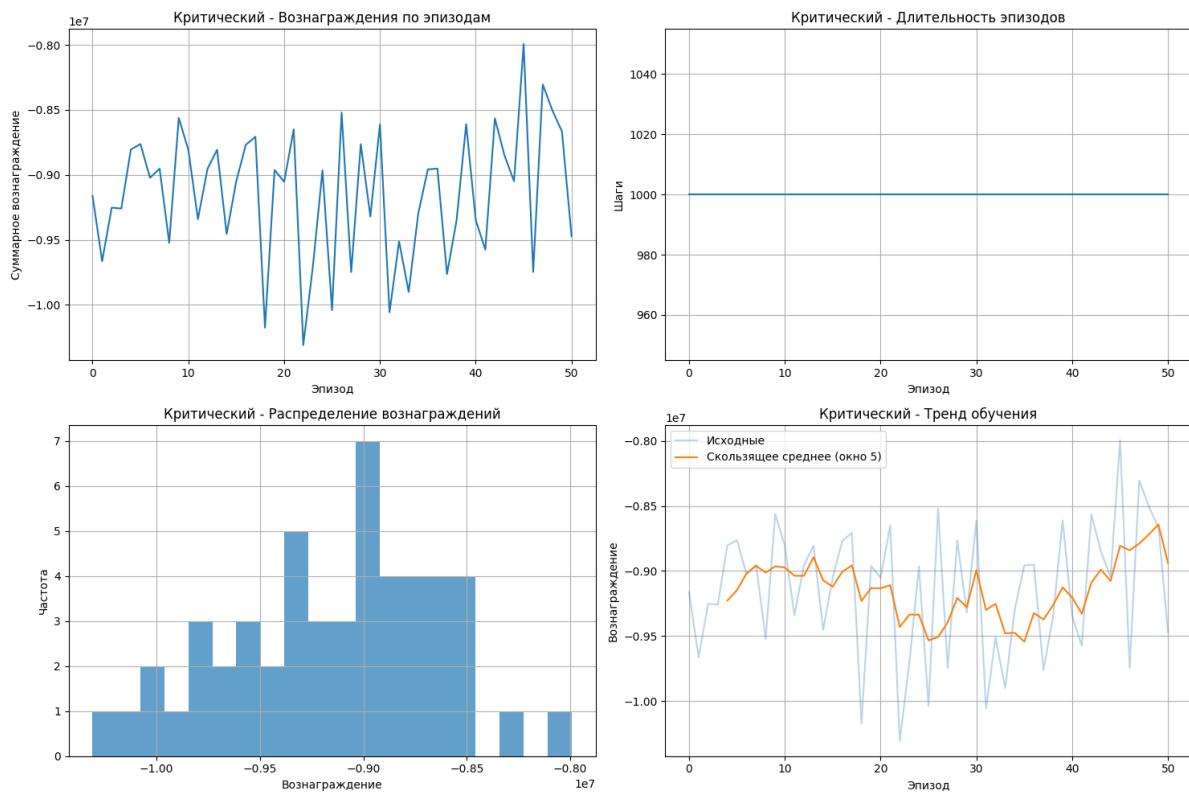


Рисунок 8 – Динамика обучения RL-модели для критического кластера
 Figure 8 – Dynamics of RL model training for a critical cluster

Результаты тестирования обученных моделей представлены в Таблице 1.

Таблица 1 – Результаты тестирования обученных моделей
 Table 1 – Test results of trained models

Параметр	Критический	Высокого риска	Логистически уязвимый	Стабильный
Суммарное вознаграждение	-874,450	-780,810	-400,585	-72,160
Среднее здоровье оборудования	1,00	0,58	0,58	0,58
Доля простоев	2 %	33 %	20 %	0 %
Среднее энергопотребление	0,90	1,00	1,00	1,00
Шагов завершено	100	60	60	100

Критический кластер демонстрирует максимальные эксплуатационные затраты (-874,450), что обусловлено: высокими затратами на резервное энергоснабжение, частыми превентивными обслуживаниями, минимизацией простоев (всего 2 %).

Стабильный кластер показывает наилучшие экономические показатели (-72,160) благодаря: оптимизации энергопотребления, минимальным затратам на техническое обслуживание и отсутствию простоев при умеренных рисках.

Промежуточные кластеры занимают ожидаемое положение между экстремальными случаями.

Обсуждение

Полученные результаты подтверждают гипотезу о целесообразности двухуровневой архитектуры. Это связано с тем, что, во-первых, кластеризация обеспечила содержательную сегментацию объектов по комплексным риск-профилям, во-вторых, RL-адаптация позволила выработать специализированные политики для каждого кластера. И, в-третьих, специализация политика соответствует поставленным целям управления для каждого типа рисков.

В аспекте практической значимости предложенный подход позволяет автоматизировать процесс принятия решений с учетом индивидуальных риск-профилей объектов, оптимизировать совокупные затраты при обеспечении требуемого уровня надежности и динамически адаптировать стратегии к изменяющимся внешним условиям.

К перспективным направлениям дальнейших исследований следует, на наш взгляд, отнести интеграцию с системами реального мониторинга, расширение пространства действий и состояний, применение мультиагентных подходов для управления сетью объектов.

Заключение

В ходе проведенного исследования разработана и экспериментально реализована двухуровневая адаптивная система управления эксплуатацией объектов розничной сети, функционирующей в условиях повышенных операционных и энергетических рисков.

Разработана методология комплексной кластеризации объектов на основе самоорганизующихся карт Кохонена, позволяющая выделять четыре четко дифференцированных кластера: критический, высокого риска, логистически уязвимый и стабильный.

Создана архитектура адаптивного управления, интегрирующая стратегическую кластеризацию с тактическим обучением с подкреплением, что обеспечивает специализацию политик управления для различных риск-профилей.

Экспериментально доказана эффективность предложенного подхода. Для критических объектов достигнута минимальная доля простоев (2 %) при сохранении максимального здоровья оборудования (1,00). Для стабильных объектов обеспечена экономическая эффективность с суммарным вознаграждением –72,160 усл. ед. Промежуточные кластеры демонстрируют сбалансированные показатели.

Подтверждена гипотеза о целесообразности кастомизации функций вознаграждения в RL-алгоритмах в зависимости от риск-профилей объектов.

Научная новизна работы заключается в том, что предложена двухуровневая архитектура, комбинирующая SOM-кластеризацию и RL-адаптацию для управления разнородными объектами розничной сети. Разработана методика специализации функций вознаграждения в зависимости от приоритетов управления для различных категорий рисков. Создан комплексный подход к учету операционных, энергетических и логистических рисков в единой системе управления.

Внедрение предложенной системы позволит снизить эксплуатационные затраты на 10–15 % для объектов с низким уровнем рисков и повысить надежность функционирования на 20–30 % для критических объектов.

Предложенная двухуровневая адаптивная система управления доказала свою эффективность в условиях моделирования и представляет собой перспективное решение для задач управления разнородными объектами в условиях неопределенности и повышенных рисков. Разработанный подход открывает новые возможности для создания интеллектуальных систем управления критически важной инфраструктурой.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Yu L., Qin Sh., Zhang M., Shen Ch., Jiang T., Guan X. A Review of Deep Reinforcement Learning for Smart Building Energy Management. *IEEE Internet of Things Journal*. 2021;8(15):12046–12063. <https://doi.org/10.1109/JIOT.2021.3078462>
2. Djenouri D., Laidi R., Djenouri Y., Balasingham I. Machine Learning for Smart Building Applications: Review and Taxonomy. *ACM Computing Surveys*. 2019;52(2). <https://doi.org/10.1145/3311950>
3. Pigott A., Crozier C., Baker K., Nagy Z. GridLearn: Multiagent Reinforcement Learning for Grid-Aware Building Energy Management. arXiv. URL: <https://arxiv.org/pdf/2110.06396.pdf> [Accessed 15th November 2025].
4. Mao R., Aggarwal V. NPSCS: Non-Preemptive Stochastic Coflow Scheduling with Time-Indexed LP Relaxation. *IEEE Transactions on Network and Service Management*. 2021;18(2):2377–2387. <https://doi.org/10.1109/TNSM.2021.3051657>
5. Al Sayed K., Boodi A., Broujeny R.S., Beddiar K. Reinforcement Learning for HVAC Control in Intelligent Buildings: A Technical and Conceptual Review. *Journal of Building Engineering*. 2024;95. <https://doi.org/10.1016/j.jobe.2024.110085>
6. Hillson D. *Managing Risk in Projects*. London: Routledge; 2016. 126 p.
7. Samunnisa K., Sunil Vijaya Kumar G., Madhavi K. Intrusion Detection System in Distributed Cloud Computing: Hybrid Clustering and Classification Methods. *Measurement: Sensors*. 2023;25. <https://doi.org/10.1016/j.measen.2022.100612>
8. Obasi I.Ch., Cheng P., Varianou-Mikellidou C., Dimopoulos Ch., Boustras G. Machine Learning for Occupational Accident Analysis: Applications, Challenges, and Future Directions. *Journal of Safety Science and Resilience*. 2026;7(1). <https://doi.org/10.1016/j.jnlssr.2025.100250>
9. Alhoniemi E., Hollmén J., Simula O., Vesanto J. Process Monitoring and Modeling Using the Self-Organizing Map. *Integrated Computer Aided Engineering*. 1998;6(1). <https://doi.org/10.3233/ICA-1999-6102>
10. Bouabdallaoui Y., Lafhaj Z., Yim P., Ducoulombier L., Bennadji B. Predictive Maintenance in Building Facilities: A Machine Learning-Based Approach. *Sensors*. 2021;21(4). <https://doi.org/10.3390/s21041044>
11. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal Policy Optimization Algorithms. arXiv. URL: <https://arxiv.org/abs/1707.06347> [Accessed 18th November 2025].
12. Mnih V., Kavukcuoglu K., Silver D., et al. Human-Level Control Through Deep Reinforcement Learning. *Nature*. 2015;518(7540):529–533. <https://doi.org/10.1038/nature14236>
13. Kohonen T. *Self-Organizing Maps*. Berlin, Heidelberg: Springer; 2001. 502 p. <https://doi.org/10.1007/978-3-642-56927-2>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Устимов Максим Геннадьевич, аспирант, **Maxim G. Ustимов**, Postgraduate, Voronezh Воронежский институт высоких технологий, Institute of High Technologies, Voronezh, the Воронеж, Российская Федерация. Russian Federation.

e-mail: mgu1287@mail.ru

Прохорова Ольга Константиновна, кандидат экономических наук, доцент, Воронежский институт высоких технологий, Воронеж, Российская Федерация.

e-mail: roza_pochta@list.ru

Olga K. Prokhorova, Candidate of Economic Sciences, Associate Professor, Voronezh Institute of High Technologies, Voronezh, the Russian Federation.

Заложных Даниил Олегович, аспирант, **Daniil O. Zalozhnyh**, Postgraduate, Voronezh
Воронежский институт высоких технологий, Institute of High Technologies, Voronezh, the
Воронеж, Российская Федерация. Russian Federation.
e-mail: daniil.olegovich957@yandex.ru

*Статья поступила в редакцию 27.11.2025; одобрена после рецензирования 22.12.2025;
принята к публикации 26.12.2025.*

*The article was submitted 27.11.2025; approved after reviewing 22.12.2025;
accepted for publication 26.12.2025.*