

На правах рукописи

Чжу Хан

**МЕТОДЫ И АЛГОРИТМЫ ТЕОРИИ ИГР И ОБУЧЕНИЯ С
ПОДКРЕПЛЕНИЕМ ДЛЯ УПРАВЛЕНИЯ МУЛЬТИАГЕНТНЫМИ
СИСТЕМАМИ ПРИ ЧРЕЗВЫЧАЙНЫХ СИТУАЦИЯХ**

Специальность 2.3.1. Системный анализ, управление и обработка
информации, статистика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Чжу Хан

Москва – 2026

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана).

Научный руководитель: **Селезнева Мария Сергеевна**, доктор технических наук, доцент, профессор кафедры технологий приборостроения МГТУ им. Н.Э. Баумана.

Официальные оппоненты: **Дарьина Анна Николаевна** – доктор физико-математических наук, доцент кафедры исследования операций федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В. Ломоносова».

Алексейчук Андрей Сергеевич – кандидат физико-математических наук, доцент Образовательного центра Института №8 «Компьютерные науки» федерального государственного автономного образовательного учреждения высшего образования «Московский авиационный институт (национальный исследовательский университет)».

Ведущая организация: **Общество с ограниченной ответственностью "Софт эксперт", г. Москва**

Защита диссертации состоится «08» декабря 2026 г. в 14 час. 30 мин. на заседании диссертационного совета 24.2.331.14 по адресу: 105005, г. Москва, Госпитальный переулок, д. 10, ауд. 613м.

С диссертацией можно ознакомиться в библиотеке МГТУ им. Н.Э. Баумана и на сайте <http://www.bmstu.ru>.

Отзывы на автореферат в двух экземплярах, заверенные печатью, просьба направлять по адресу: 105005, г. Москва, 2-я Бауманская ул., д. 5, стр. 1 на имя ученого секретаря диссертационного совета.

Автореферат разослан «__» _____ 2026 г.

Ученый секретарь
диссертационного совета
к.т.н., доцент

Глазков Виталий Владимирович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. С ростом частоты и сложности природных катастроф традиционные системы чрезвычайного реагирования сталкиваются с серьезными вызовами в части скорости отклика, охвата территории и эффективности кооперации. Кооперативные мультиагентные системы становятся важным направлением технологического инновационного развития в области спасательных операций.

В последние годы данные-ориентированные подходы, в частности мультиагентное обучение с подкреплением (МАОП), стали ключевой технологией для сложных задач кооперации. МАОП обеспечивает обучение агентов оптимальным стратегиям в процессе взаимодействия со средой, основанного на экспериментах и корректировках, не требуя точных моделей. Теория игр как развитая рамка для анализа стратегических взаимодействий обеспечивает надежную математическую базу для изучения кооперации и конкуренции агентов.

Таким образом, проведение исследований в области мультиагентного управления и оптимизации обучения с подкреплением на основе теории игр — через инновационную разработку алгоритмов и построение эффективных рамок обучения — имеет важное теоретическое значение и широкие перспективы практического применения для преодоления существующих технологических ограничений и продвижения мультиагентных систем к более высокому уровню автономной кооперации.

Степень разработанности темы характеризуется результатами предшествующих исследований, изложенных в публикациях: Воронова Е.М., Микрина Е.А., Обносова Б.В. — Стабилизация, наведение, групповое управление и системное моделирование беспилотных летательных аппаратов; Yang Y., Wang J. — Мультиагентное обучение с подкреплением на основе теории игр; Клейменов А.Ф. — Неантагонистические позиционные дифференциальные игры; Khadka S., Majumdar S., Miret S., McAleer S., Tumer K. — Эволюционное обучение с подкреплением для эффективной выборочной мультиагентной координации и др..

Большой вклад в разработку алгоритмов интеграции эволюционной теории игр с мультиагентным обучением с подкреплением внесли К.А. Неусыпин, М.С. Селезнева, Чжан Чэ, Чжай Имин, Пэн Сыкунь, Цзи Юньпэн и др.

Целью диссертационного исследования является повышение точности управления мультиагентными системами с помощью теории игр и обучения с подкреплением в условиях чрезвычайных ситуаций.

Для достижения цели в работе были сформулированы и решены следующие **задачи**:

1. Анализ ключевых трудностей мультиагентного управления при чрезвычайных ситуациях, существующих теоретических рамок и интеграций МАОП с теорией игр.

2. Разработка метода оптимизации коммуникационных стратегий мультиагентных систем при чрезвычайных ситуациях на основе теории игр.

3. Разработка алгоритма кооперативного планирования траекторий для гетерогенных мультиагентных систем при чрезвычайных ситуациях на основе МАОП.

4. Разработка алгоритма оптимизации МАОП с интеграцией теории игр для повышения эффективности и стабильности кооперации масштабных мультиагентных систем.

Методы исследования. При решении сформулированных задач использовались методы теории автоматического управления, глубокого обучения, теории системного анализа, теории игр, математического анализа и программные пакеты MATLAB и Python.

Научная новизна работы заключается в следующем:

1. Разработан метод интеграции эволюционной теории игр (ЭТИ) и МАОП и рассмотрены различные возможные стадии реализации данной интеграции.

2. Разработан метод применения теории неантагонистических игр с типами поведения игроков к оптимизации коммуникационных стратегий между беспилотными спасательными системами при чрезвычайных ситуациях, который позволяет адаптироваться к более сложным рабочим сценариям и обеспечивает более эффективное энергопотребление системы.

3. Разработан модифицированный алгоритм планирования траекторий SAC-II, адаптированный для беспилотных спасательных систем при чрезвычайных ситуациях, который по сравнению с аналогичными решениями улучшает показатели выполнения задач и эффективность кооперации при формировании команд.

4. Разработан алгоритм МАОП-ЭТИ на основе интеграции репликаторной динамики с МАОП, который по сравнению с аналогами позволяет улучшить взаимодействие агентов, снизить риск локального оптимума и обеспечить стабильность масштабных групп, что повышает эффективность работы беспилотных спасательных систем.

Теоретическая значимость результатов исследования. Разработан метод интеграции ЭТИ с МАОП, позволяющий учитывать различные стадии реализации данной интеграции. Применена неантагонистическая теория игр с типами поведения для оптимизации стратегии коммуникаций мультиагентных систем. Доказана оптимизирующая роль эволюционной динамики в процессе обучения МАОП.

Практическая значимость результатов исследования. Разработанные алгоритмы использованы для управления беспилотными спасательными системами. С помощью алгоритма SAC-II улучшены показатели выполнения задач и кооперации при формировании команд. С помощью алгоритма МАОП-ЭТИ улучшены взаимодействие агентов, снижен риск локальной оптимизации и обеспечена стабильность работы мультиагентной системы.

Результаты работы использованы в учебном процессе кафедры систем автоматического управления МГТУ им. Н.Э. Баумана и в научно-исследовательских проектах ООО «Сучжоу Иновэнс Технолоджи» (КНР).

Достоверность и обоснованность полученных теоретических и практических результатов подтверждаются результатами математического моделирования, а также согласованностью результатов с известными данными в этой области.

Основные положения диссертационной работы, выносимые на защиту.

1. Метод интеграции ЭТИ с МАОП и различные возможные стадии их интеграции.

2. Метод применения теории неантагонистических игр с типами поведения игроков, к оптимизации коммуникационных стратегий между гетерогенными системами при чрезвычайных ситуациях.

3. Алгоритм SAC-II, который адаптирован для задач планирования траекторий гетерогенных мультиагентных систем при чрезвычайных ситуациях и демонстрирует высокую гибкость и масштабируемость.

4. Алгоритм МАОП-ЭТИ, улучшающий кооперацию мультиагентной системы, позволяющий снизить риск возникновения локальных оптимумов и сохранять стабильность при масштабировании системы.

Апробация работы. Основные результаты диссертационной работы докладывались и обсуждались на следующих конференциях: XLIX Академические Чтения по космонавтике (Королевские Чтения, 2025 г.), Международная научно-техническая конференция «Автоматизация» (Сочи, 2025г.), Теория и реализация информационных систем Международная научная конференция (Москва, 2025г.), VI Межвузовской конференции аспирантов, соискателей и молодых ученых. (Москва, 2024г.).

Публикация. По теме диссертации опубликовано 9 научных работ, из них 3 статьи в журналах из Перечня рецензируемых научных изданий – ВАК РФ и 2 статьи в Перечне международных научных изданий в системе Web Science и Scopus, общим объемом 7,16 п.л./3,58 п.л.

Структура и объём работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы из 103 наименований. Текст диссертации изложен на 120 страницах, содержит 19 рисунков и 5

таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы, определены цели, задачи и методы исследования, отмечены научная новизна и практическая значимость, приведены сведения об апробации, внедрении результатов, публикациях, структуре и объему диссертации.

В первой главе рассмотрено применение беспилотных спасательных систем при чрезвычайных ситуациях. Проанализированы преимущества беспилотного летательного судна (БЛС) и беспилотного наземного транспортного средства (БНТС) при выполнении спасательных задач. Представлен обзор применения МАОП и теории игр в управлении мультиагентными системами. Отмечено, что мультиагентное обучение с подкреплением стало перспективной тенденцией развития, при этом теория игр обеспечивает всестороннюю теоретическую поддержку для МАОП. На рисунке 1 представлены ключевые стадии интеграции ЭТИ с МАОП.



Рисунок 1 – Ключевые стадии интеграции ЭТИ с МАОП

ЭТИ открывает многообещающий путь для решения основных проблем мультиагентного управления. Это позволяет не только эффективно решать проблемы неустойчивости, с которыми сталкивается МАОП, но и повышать надежность и адаптивность системы, внедряя идею биологической эволюции. На основе проведенного анализа сформулированы цель и задачи диссертационного исследования.

Во второй главе рассмотрены современное состояние исследований и основные проблемы коммуникации в беспилотных спасательных системах при чрезвычайных ситуациях. Проанализированы дилеммы, с которыми столкнулась система коммуникации во время катастрофы. Исследовано применение теории игр для оптимизации коммуникаций в целях экономии энергии в чрезвычайных ситуациях.

В современных исследованиях БНТС используются в качестве базовой станции, передающей сигнал, а БЛС в качестве приёмников сигнала как показано на рисунке 2.

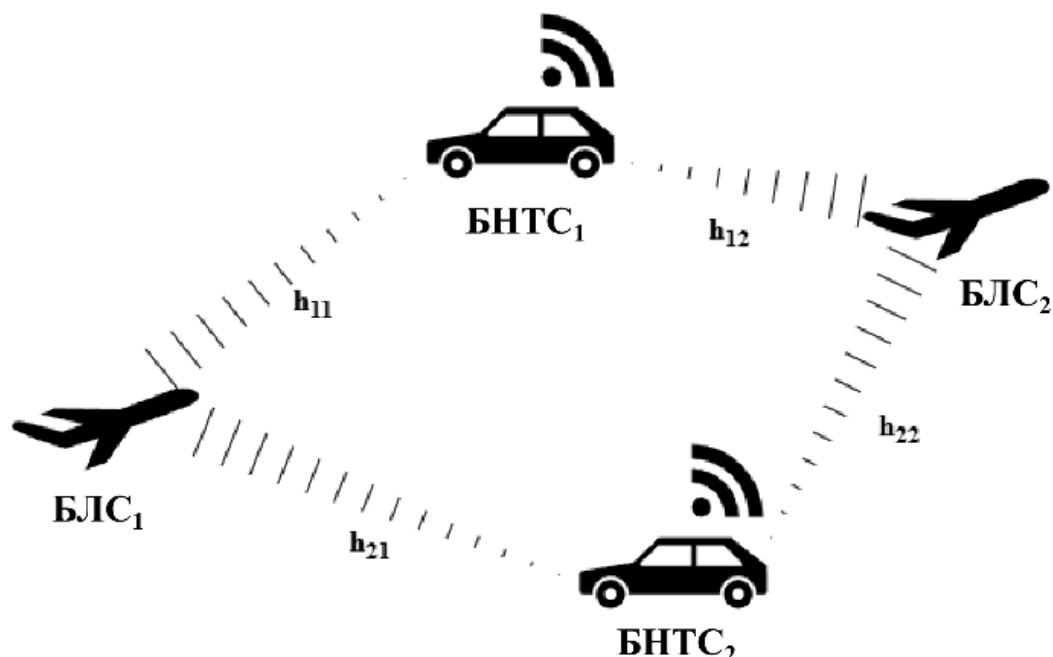


Рисунок 2 – Схема коммуникации беспилотных спасательных систем

Пропускная способность информации является ключевым параметром для оценки производительности коммуникационных систем. Чем выше пропускная способность, тем быстрее система может передавать данные. В нашей системе пропускная способность зависит от мощности передачи сигнала всех базовых станций и расстояния между БЛС и БНТС и т.д. Для k -ой базовой станции с мощностью s_k её пропускная способность информации $t(s_k)$ формулируется следующим образом:

$$t(s_k) = (1 - \exp\{-\gamma_k(s)\})^L \quad (1)$$

где s — мощность взаимодействующих базовых станций в рассматриваемой задаче, $\gamma_k(s)$ — отношение сигнал-помеха-плюс-шум; L — количество информационных битов в пакете

$$\gamma_k(s) = \frac{\Gamma h_k s_k}{\sigma^2 + h_{\setminus k} s_{\setminus k}} \quad (2)$$

где Γ — коэффициент усиления обработки, h — расстояние между БЛС и БНТС, $s_{\setminus k}$ — мощность базовых станций кроме k -ой, σ^2 — мощность аддитивного белого гауссовского шума.

Функция полезности $u_k(s)$ в качестве которой выступает эффективность базовых станциях:

$$u_k(s) = \frac{t_k(s)}{s_k} \quad (3)$$

Представлены базовые концепции и принципы неантагонистических игр с типами поведения игроков. Основные типы поведения: нормальное (nor), альтруистическое (alt), агрессивное (agg) и парадоксальное (par). При выборе различных типов поведения БЛС и БНТС в беспилотных спасательных системах имеют более высокий уровень адаптации к динамически меняющимся условиям катастрофной среды. Например, если каждый агент имеет четыре типа поведения, то в рассматриваемой игре двух лиц с функциями выигрыша I_i и I_j существует 16 возможных комбинаций, показанных в таблице 1.

Таблица 1 – Комбинация поведений двух игроков

Игрок j Игрок i	nor	alt	agg	par
nor	(I_i, I_j)	(I_i, I_i)	$(I_i, -I_i)$	$(I_i, -I_j)$
alt	(I_j, I_j)	(I_j, I_i)	$(I_j, -I_i)$	$(I_j, -I_j)$
agg	$(-I_j, I_j)$	$(-I_j, I_i)$	$(-I_j, -I_i)$	$(-I_j, -I_j)$
par	$(-I_i, I_j)$	$(-I_i, I_i)$	$(-I_i, -I_i)$	$(-I_i, -I_j)$

Процесс решения проблемы неантагонистической игрой с типами поведения игроков формализуем следующим образом:

Шаг 1. Инициализируя параметры: $h_{11} = 1.0, h_{21} = 0.75, h_{12} = 0.5, h_{22} = 1.0$; коэффициент усиления обработки $\Gamma = 4$, отношение максимальной выходной мощности базовой станции к аддитивному белому гауссовскому шуму $p/\sigma^2 = 5$. Типы поведения игроков включают: нормальное и альтруистическое. Игрок выбирает тип поведения и вычисляет точку равновесия Нэша – стратегию s^* .

Шаг 2. В множестве достижимости системы, найти набор стратегий z , который дает каждому участнику тот же выигрыш, что и равновесие Нэша, т.е. стратегии, в которых выполняются равенства:

$$I_i(z) = I_i(s^*), i \in K \quad (4)$$

Шаг 3. Выбирая поведение игроков, находим соответствующую стратегию из набора стратегий z . В данной ситуации система является нормальной и альтруистической, поэтому существуют следующие 4 возможных комбинации ситуаций и выигрышей как показано на рисунке 3. Чтобы было более наглядно, изобразим границу Парето для данной проблемы и найдем функции выигрыша для этих четырёх стратегий.

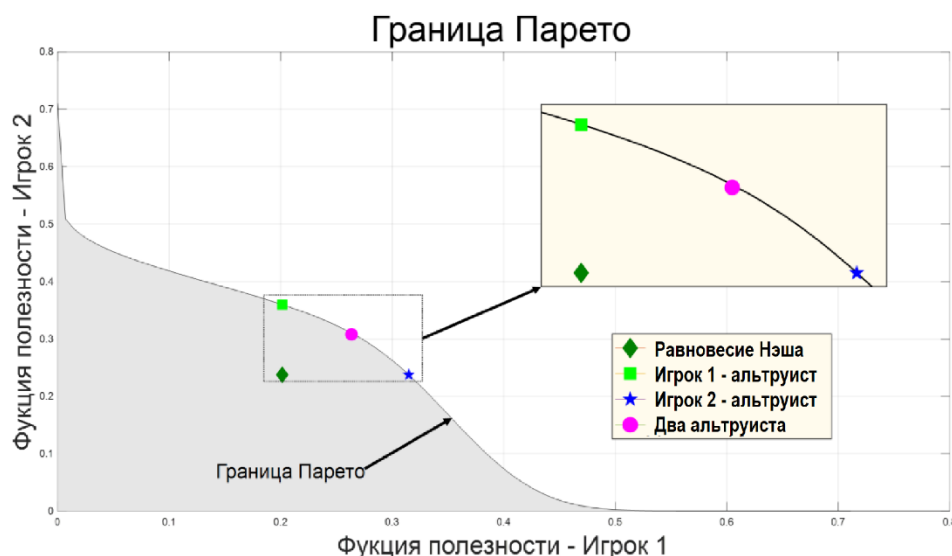


Рисунок 3 – Сравнение с границей Парето

В таблице 2, где сравниваются выигрыши игрока с типами поведения, а в скобках указаны улучшения по сравнению с равновесием Нэша, видно, что после выбора вариантов поведения выигрыши игрока улучшаются по сравнению с равновесием Нэша.

Таблица 2. – Выигрыши игрока с типами поведения

Поведения Выигрыш	nor,nor	alt,nor	nor,alt	alt,alt
Игрок 1	0.2013	0.2013	0.3147(56.33%)	0.2635(30.87%)
Игрок 2	0.2376	0.3597(51.42%)	0.2376	0.3080(29.64%)

Продемонстрирована эффективность применения неантагонистических игр для оптимизации коммуникационных стратегий в беспилотных спасательных системах при чрезвычайных ситуациях. Данный подход обеспечивает более эффективное энергопотребление системы и позволяет адаптироваться к более сложным рабочим ситуациям.

В третьей главе проанализированы основные трудности планирования траекторий для гетерогенных мультиагентных систем при чрезвычайных ситуациях. Предложен алгоритм SAC-II для планирования траекторий гетерогенных систем.

В катастрофной среде присутствует большое количество препятствий. БЛС, благодаря своей превосходной маневренности и проходимости, способны предоставлять БНТС более полную информацию об окружающей среде, а также повышать степень выполнения задач и эффективность работы БНТС за счет поддержания кооперативного движения в группе.

Математические характеристики вероятностного распределения действий агента такие, как температура и энтропия, являются ключевыми параметрами алгоритма SAC, определяющими эффективность функционирования агента в режимах исследования и эксплуатации. Основной акцент в SAC сделан на максимизации вознаграждения с учетом энтропии, что обеспечивает более высокую вариативность поведения агентов.

$$\max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t \left(r(s_t, a_t) + \alpha \mathcal{H}(\pi(\cdot | s_t)) \right) \right] \quad (5)$$

где α — параметр температуры, который управляет компромиссом между исследованием и эксплуатацией; $\mathcal{H}(\pi)$ — Энтропия стратегии демонстрирует степень неопределенности в распределении вероятностей действий, принимаемых агентом:

$$\mathcal{H}(\pi(\cdot | s)) = -\mathbb{E}_{\pi} [\log \pi(a|s)] \quad (6)$$

Структура алгоритма SAC представлена на рисунке 4; она включает в себя сеть «Актер», сеть «Критик» и сеть «Цель критика». Алгоритм SAC демонстрирует значительные преимущества при решении задач управления гетерогенными мультиагентными системами, которые проявляются в адаптации к нелинейности и высокоразмерным различиям, высокой эффективности совместного обучения и адаптивности к динамическим средам.

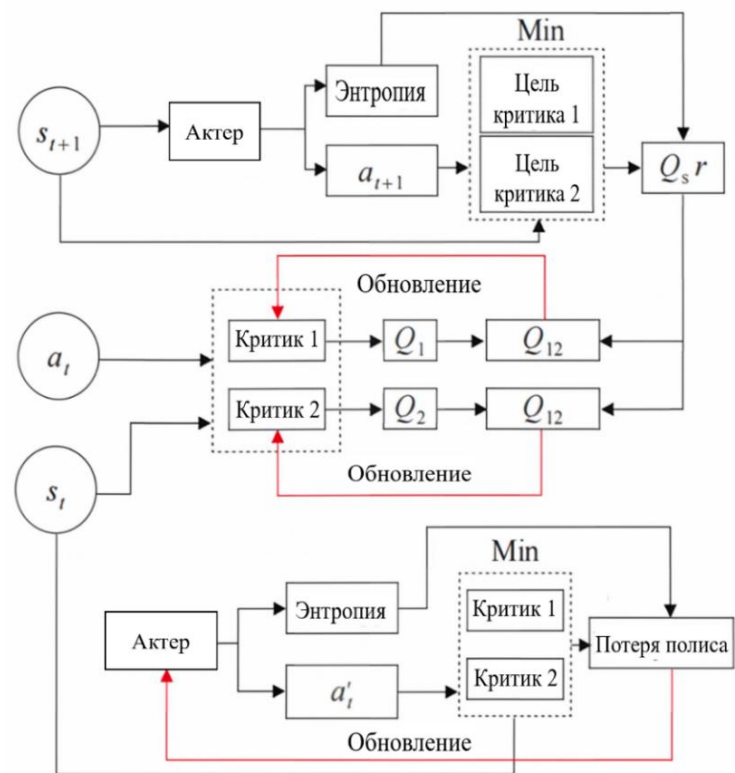


Рисунок 4 – Структура алгоритма SAC

Разработана визуальная среда для совместного планирования траектории БЛС и БНТС. БЛС могут избегать препятствий в окружающей среде и предоставлять БНТС информацию о целевом положении. БНТС, отслеживая положение БЛС, в конечном итоге достигают пункта назначения.

В нашем случае совместная система является частично наблюдаемой, т.е. частично наблюдаемый марковский процесс. Система частично наблюдаема и может быть определена как кортеж $\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \mathcal{Z}, \mathcal{O}, \gamma$, где набор состояний $\mathcal{S}$, вероятность перехода между состояниями $\mathcal{P}$ и функция наблюдения $\mathcal{O}$ являются ненаблюдаемыми для агентов. В то же время набор действий $\mathcal{A}$, набор наблюдений $\mathcal{Z}$, функция вознаграждения $\mathcal{R}$ и коэффициент распада γ являются наблюдаемыми для агентов. Определение некоторых параметров уже было описано, теперь необходимо спроектировать набор наблюдений $\mathcal{Z}$ и функцию вознаграждения $\mathcal{R}$. Разработаны различные механизмы определения функций вознаграждения r для разных типов агентов.

1) Вознаграждения за границу – r_{edge}

$$r_{edge} = \begin{cases} -1, & \text{in the edge,} \\ 0, & \text{else.} \end{cases} \quad (7)$$

Когда агент достигает границы, назначается штраф -1, в противном случае вознаграждение не дается.

2) Вознаграждения за формирование дистанционного бонуса – r_{follow}

$$r_{follow} = \begin{cases} 1 - 0.001 * d_{ij}, & d_{ij} < 50 \\ -0.001 * d_{ij}, & \text{else.} \end{cases} \quad (8)$$

где d_{ij} – расстояние между i – ым БНТС и i – ым БЛС.

3) Вознаграждения за препятствие – r_{obs}

$$r_{obs} = \begin{cases} 0, & d_o > 2R_o, \\ -2, & R_o \leq d_o \leq 2R_o, \\ -500, & d_o < R_o. \end{cases} \quad (9)$$

где d_o – расстояние между агентом и препятствием.

4) Вознаграждения за выполнение задачи r_{goal}

$$r_{goal_{\text{БНТС}}}^i = \begin{cases} 1000, & d_{ig} < 40 \\ -0.001 * d_{ig}, & \text{else} \end{cases} \quad (10)$$

$$r_{goal_{\text{БЛС}}}^j = \begin{cases} 0.1 - 0.001 * d_{jg}, & d_{jg} < 40 \\ -0.001 * d_{jg}, & \text{else} \end{cases} \quad (11)$$

где $r_{goal_{\text{БНТС}}}^i$ – вознаграждения для i – ого БНТС и d_{ig} – расстояние между i – ым БНТС и целевым пунктом. $r_{goal_{\text{БЛС}}}^j$ – вознаграждения для

j – ого БЛС и d_{jg} – расстояние между j – ым БЛС и целевым пунктом.

После разработки функции вознаграждения мы получаем общее вознаграждение

$$R = \omega_1 * r_{edge} + \omega_2 * r_{follow} + \omega_3 * r_{obs} + \omega_4 * r_{goal} \quad (12)$$

где $[\omega_1, \omega_2, \omega_3, \omega_4]$ — веса для каждого компонента. Для БНТС веса $\omega_1 = \omega_2 = \omega_3 = \omega_4 = 1$, для БЛС разработаны два контекстно-зависимых режима работы:

- 1) Целевое взвешивание: Когда БЛС находится далеко от целевой точки, $\omega_1 = \omega_2 = \omega_3 = 1, \omega_4 = 10$, приоритет отдается быстрой навигации к целевой позиции.
- 2) Взвешивание по формации: Когда БЛС находится далеко от БНТС, $\omega_1 = \omega_3 = \omega_4 = 1, \omega_2 = 10$, взвешивание стимулирует поддержание формации для повышения эффективности сотрудничества.

Затем определим параметры в алгоритме SAC. Сеть «Актер» представляет собой полносвязный слой со структурой 7-256-256-2, функция активации состоит из relu и tanh , начальные веса нейронной сети случайным образом выбираются из нормального распределения со средним значением 0 и дисперсией 0,1, а скорость обучения составляет $1e-3$. Сеть «Критик» принимает в качестве входных данных инкрементальную информацию о состояниях и действиях. Количество узлов во входном слое составляет $7 \times N + 2$, функция активации состоит только из relu , а начальные веса нейронной сети такие же, как у сети «Актер», скорость обучения составляет $3e-3$. Общее количество циклов обучения устанавливается равным 500, тестирование алгоритма показано на рисунке 5.

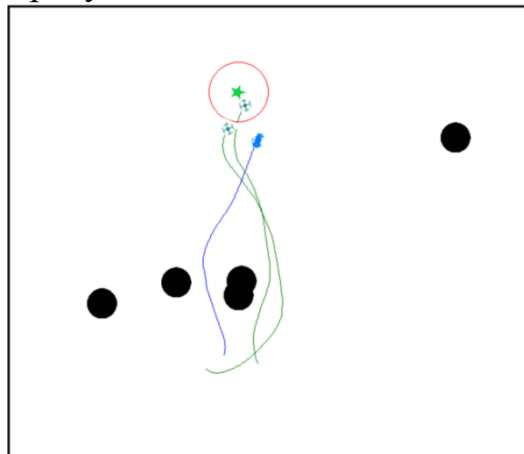


Рисунок 5 – Тестирование алгоритма SAC-II

Устанавливая разное количество препятствий и БЛС, проведено 100 экспериментов Монте-Карло для каждого из различных сценариев. В таблице 3 приведены результаты алгоритма SAC-II, первые данные — это коэффициент выполнения задачи, а данные в скобках — коэффициент

формирования команды. Результаты показывают, что при наличии препятствий, независимо от изменения количества БЛС, коэффициент выполнения задачи планирования траектории составляет более 67%, а коэффициент формирования команды — более 58%.

Таблица 3. – Эффективность алгоритма SAC-II

Кол-во БЛС	Кол-во препятствий		
	1	3	5
1	90%(59%)	77%(66%)	67%(58%)
2	91%(72%)	82%(69%)	68%(74%)
3	93%(75%)	82%(75%)	68%(75%)

Таблица 4. – Сравнение алгоритмов

Алгоритм	Кол-во препятствий		
	1	3	5
Random	5%(6%)	3%(7%)	2%(5%)
MADDPG	62%(12%)	52%(15%)	36%(14%)
SAC-II	91%(72%)	82%(69%)	68%(74%)

По сравнению со случайной стратегией и алгоритмом MADDPG данный подход демонстрирует превосходную производительность на более чем 50% и обеспечивает эффективный синергетический эффект кооперации на более чем 3 раза, как показано в таблице 4.

В четвертой главе представлены распространенные методы МАОП на основе эволюционных алгоритмов. На рисунке 6 показаны четыре распространенных способа комбинирования МАОП и ЭА: а) Оптимизация RL с поддержкой ЭА; б) Оптимизация ЭА с поддержкой RL; в) ЭА и RL решают одну и ту же задачу одновременно и интегрируют стратегии; г) Разделение задачи на подзадачи, подходящие для ЭА, и подзадачи, подходящие для RL.

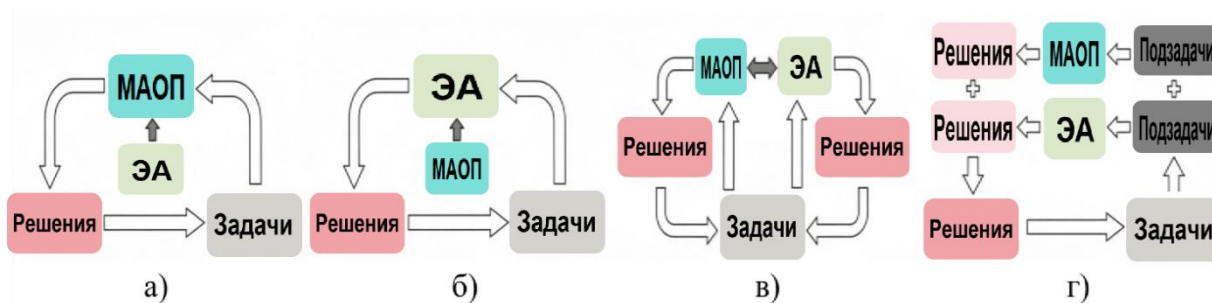


Рисунок 6 – 4 распространенных способа комбинирования МАОП и ЭА

Используется архитектура «Основная сеть и несколько-голов» для представления мультиагентной системы. За счет совместного использования весов основной сети эта архитектура эффективно повышает эффективность обучения для задач МАОП, а её конкретная структура представлена на рисунке 7:

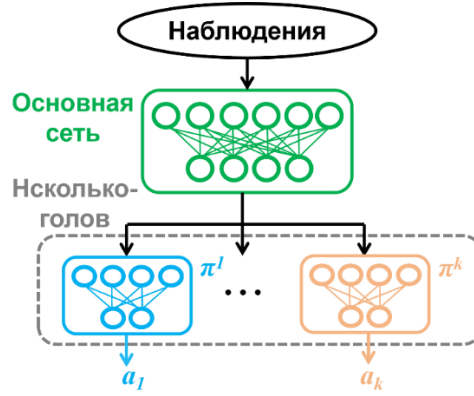


Рисунок 7 – Архитектура системы глубокого обучения

В данной работе предлагается новый алгоритм МАОП-ЭТИ, который сочетает оптимизацию без градиентов из ЭТИ и градиентную оптимизацию из МАОП — как показано на рисунке 8.

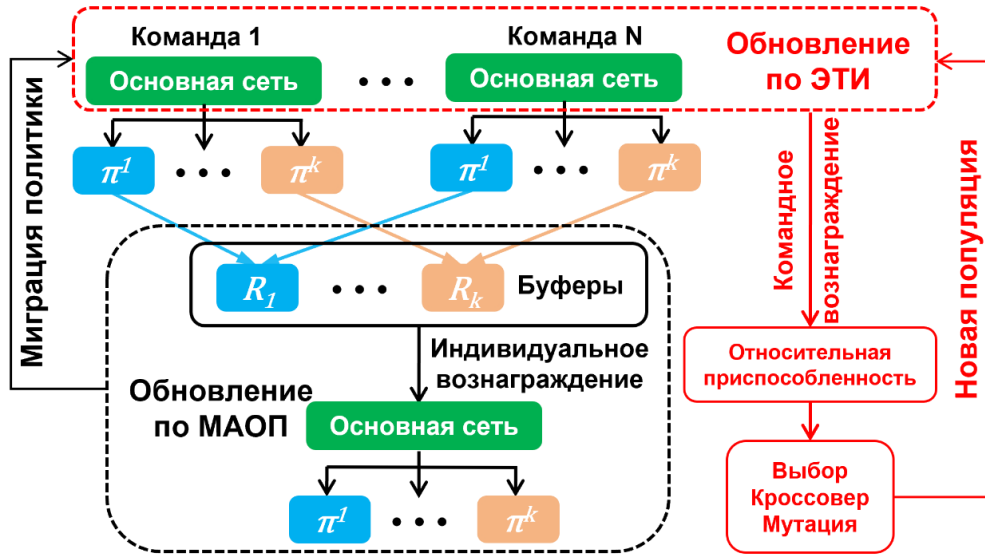


Рисунок 8 – Архитектура алгоритма МАОП-ЭТИ

При обновлении основной сети интегрируются идеи эволюционных динамик, с мотивацией увеличения командного вознаграждения — это способствует возникновению кооперативного поведения между агентами.

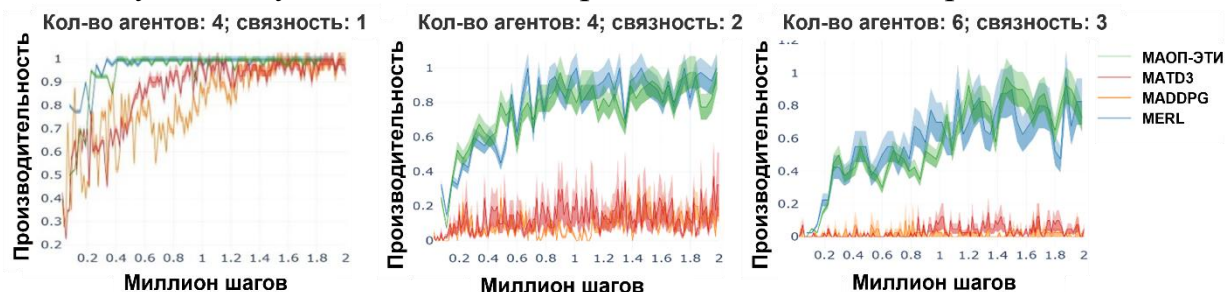
$$\dot{x}_s(t) = \frac{dx_s(t)}{dt} = x_s(t) \cdot [u_s(x) - \bar{u}(x)], i \in \{1, \dots, N\} \quad (13)$$

где $x_s(t)$ — доля стратегии s в популяции, $u_s(x)$ — приспособленность стратегии s , $\bar{u}(x)$ — средняя приспособленность популяции.

Градиентная оптимизация применяется для формирования стратегий агентов: ориентируясь на индивидуальное вознаграждение каждого участника системы, данный подход интенсифицирует его специфические функциональные способности. Градиентные стратегии регулярно добавляются в эволюционную популяцию, направляя её постоянную эволюцию.

Задачи спасательных операций часто требуют, чтобы спасательная система исследовала несколько точек интереса в окружающей среде, что полностью соответствует среде Rover Domain. Именно в этой среде и проведена верификация алгоритма: k агентов сотрудничают для достижения M точек интереса, и каждая точка интереса считается успешно достигнутой только тогда, когда её наблюдают определенное количество агентов. Требуемое количество наблюдений называется требованием к связности, т.е. связность представляет собой командное сотрудничество. Командное вознаграждение представляет собой процент точек интереса, наблюдаемых в конце выполнения задачи. Проведены тесты при фиксированном количестве точек интереса (4 штуки) и различных количествах агентов и коэффициентах связности, чтобы сравнить производительность различных алгоритмов: два традиционных алгоритма МАОП – MADDPG и MATD3, два алгоритма с интеграцией эволюционного алгоритма – MERL и МАОП-ЭТИ.

Рисунок 9 демонстрирует производительность различных алгоритмов за 2 миллиона шагов обучения при следующих условиях: количество агентов — 4, 4, 6, и коэффициент связности — 1, 2, 3 соответственно. Из экспериментальных результатов видно, что в сценариях, не требующих командного взаимодействия, как традиционные МАОП, так и варианты MARL с интеграцией эволюционных алгоритмов успешно справляются с поставленной задачей. Однако при сравнении эффективности обучения на рис. а) и б) обнаруживается резкое падение показателей у традиционного MARL; на рис. в) при значении коэффициента связности — 3 выполнение задачи становится практически невозможным. А алгоритма MERL и МАОП-ЭТИ, интегрированные с эволюционным алгоритмом, в задачах, требующих командного взаимодействия, демонстрируют значительно лучшие результаты по сравнению с традиционными алгоритмами МАОП. При большом количестве агентов алгоритмы МАОП не могут выполнить командную задачу лишь на основе простой системы вознаграждений.



а) Связность 1

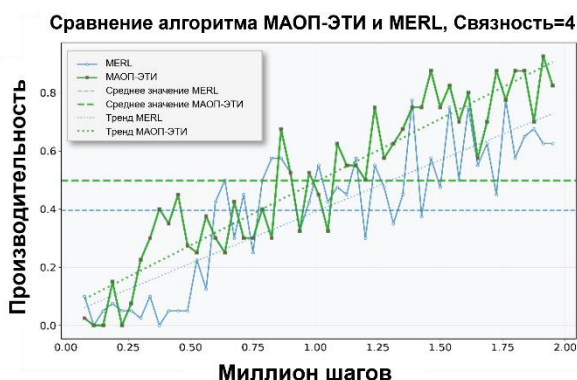
б) Связность 2

в) Связность 3

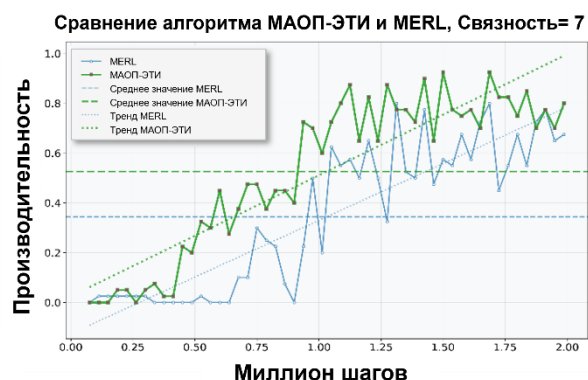
Рисунок 9 – Сравнение производительности различных алгоритмов

Когда количество агентов превышает 8, традиционный МАОП в принципе не работает. На рисунке 10 более подробно сравнивается эффект

обучения алгоритмов MERL и МАОП-ЭТИ. Слева представлен сценарий с 8 агентами и коэффициентом связности, равным 4, а справа – сценарий с 14 агентами и коэффициентом связности, равным 7. За 2 миллиона шагов обучения средняя производительность алгоритма MERL составила 0,395 и 0,344 соответственно, при этом алгоритм начал быстро сходиться на 525 тысячах и 750 тысячах шагов обучения. А средняя производительность алгоритма МАОП-ЭТИ составила 0,498 и 0,526 соответственно, при этом алгоритм начал быстро сходиться на 300 тысячах и 450 тысячах шагов обучения. Средняя производительность алгоритма МАОП-ЭТИ соответственно превышает алгоритм MERL на 26,1% и 52,9% и скорость сходимости при обучении алгоритмов увеличивается на 40%. Сравнение с алгоритмом MERL показало, что при небольшом количестве агентов алгоритм МАОП-ЭТИ обладает сопоставимой скоростью сходимости и производительностью. Однако с увеличением количества агентов и коэффициента связности, т.е. когда требования к способности командного взаимодействия растут, скорость сходимости алгоритма MERL начинает снижаться. В то время как предложенный в данной работе алгоритм МАОП-ЭТИ успешно смягчает негативное влияние роста количества агентов, обеспечивает быструю сходимость и превосходит алгоритм MERL по производительности.



а) Агент 8, Связность 4



б) Агент 14, Связность 7

Рисунок 10 – Сравнение алгоритмов MERL и МАОП-ЭТИ

Рисунок 11 демонстрирует результаты выполнения задачи алгоритмами МАОП-ЭТИ и MERL в среде Rover Domain при количестве агентов, равном 8. Оранжевый цвет обозначает успешно достигнутые точки интереса, а серый — не достигнутые. Как видно из представленных данных, при одинаковом распределении точки интереса в алгоритме МАОП-ЭТИ агенты, достигнув определенной точки интереса, не попадают в локальный оптимум, а сохраняют способность к постоянному исследованию среды, что обеспечивает высокий уровень завершения задачи.

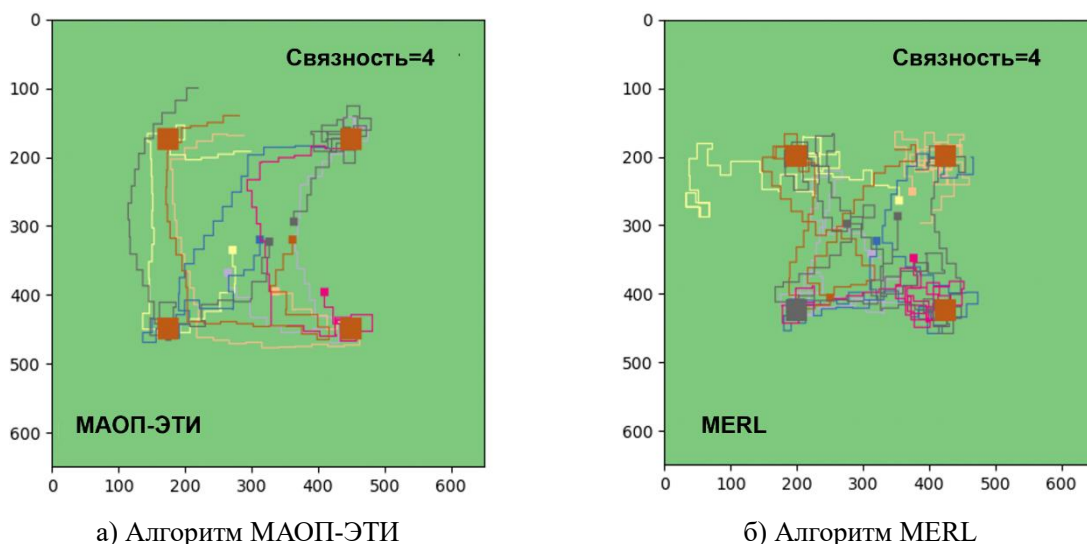


Рисунок 11 – Моделирование в среде Rover Domain

Результаты экспериментов подтверждают, что алгоритм МАОП-ЭТИ обеспечивает кооперацию между агентами, снижает вероятность застревания в локальных оптимумах и сохраняет высокую стабильность даже в сценариях с большим количеством агентов.

В заключении сформулированы основные результаты работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Разработан метод интеграции ЭТИ с МАОП и рассмотрены различные возможные стадии реализации данной интеграции.

2. Разработан метод применения теории неантагонистических игр с учетом типов поведения игроков для оптимизации коммуникационных стратегий между гетерогенными системами в условиях чрезвычайных ситуаций. Данный подход позволяет адаптироваться к более сложным рабочим ситуациям и обеспечивает более эффективное энергопотребление системы.

3. Разработан модифицированный алгоритм планирования траекторий SAC-II, адаптированный для гетерогенных мультиагентных систем при чрезвычайных ситуациях. За счёт более эффективной кооперации при формировании команды данный подход повышает показатели выполнения задач по сравнению с аналогичными решениями на 50%.

4. Разработан алгоритм МАОП-ЭТИ на основе интеграции динамики репликаторов в процесс МАОП. Данный подход помогает улучшить взаимодействие агентов, снизить риск локальной оптимизации и обеспечить стабильность с масштабной командой. За счет совместного использования весов основной сети и интеграции динамики репликаторов скорость сходимости при обучении алгоритмов увеличивается на 40%.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Селезнева М. С., Чжу Х., Чжан Ч. Улучшение мультиагентного обучения с подкреплением на основе эволюционной теории игр // Известия Тульского государственного университета. Технические науки. 2026. № 6. С. 85-93. (1,04п.л./0,44п.л.).
2. Чжу Х., Пэн С., Селезнева М. С. Совместное планирование траекторий динамических объектов на основе обучения с подкреплением // Авиакосмическое приборостроение. 2025. № 8. С. 56-63. (0,92п.л./0,45п.л.).
3. Чжу Х., Селезнева М. С. Обзор исследований по интеграции эволюционной теории игр и мультиагентного обучения с подкреплением // Инженерный вестник Дона. 2026 № 3 (135). С. 15-30. (1,85п.л./0,63п.л.).
4. Оптимизация стратегий коммуникации между БПЛА и беспилотными автомобилями на основе неантагонистической игры / Х. Чжу [и др.]. // Прикладная физика и математика. 2025. № 1. С. 38-44. (0,81п.л./0,4п.л.).
5. Чжу Х., Селезнева М. С. Планирование траекторий беспилотного автомобиля на основе алгоритма мягкий актер-критик // Теория и реализация информационных систем: Сборник статей I международной научной конференции (Москва, 15 мая 2025 г.). М.: Издательство МГТУ им. Н.Э. Баумана, 2025. С. 429-431. (0,46п.л./0,16п.л.).
6. Чжу Х., Селезнева М.С. Система управления квадрокоптером на основе линеаризации обратной связью // XLIX Академические чтения по космонавтике, посвященные памяти академика С.П. Королёва и других выдающихся отечественных ученых - пионеров освоения космического пространства (Москва, 28 - 31 января 2025 г.): Сборник тезисов : в 2 т. Т. 2. М.: Издательство МГТУ им. Н.Э. Баумана, 2025. С. 216-217. (0,23п.л./0,1п.л.).
7. Чжу Х., Селезнева М.С. Система управления квадрокоптером с LQR при возмущении ветра // Наука, технологии и бизнес: VI Межвузовская конференция аспирантов, соискателей и молодых ученых (Москва, 16–18 апреля 2024 года): материалы конференции. М.: Издательство МГТУ им. Н.Э. Баумана, 2024. С. 531-536. (0,69п.л./0,3п.л.).
8. Zhu H., Selezneva M.S., Zhai Y. Trajectory Planning of Heterogeneous Multi-Agent System Based on Soft Actor-Critic Algorithm // 2025 International Russian Automation Conference (RusAutoCon). IEEE. 2025. P. 18-23. (0,58п.л./0,52п.л.).
9. Zhai Y., Selezneva M.S., Zhu H. Automatic Parking Trajectory Optimization: A Hybrid Genetic Algorithm with Q Learning // 2025 International Russian Automation Conference (RusAutoCon). 2025. IEEE. P. 12-17. (0,58п.л./0,58п.л.).