

**Санкт-Петербургский политехнический университет
Петра Великого
Институт компьютерных наук и технологий**

На правах рукописи

Ханафи Мохамед Яссин

**Нейросетевые системы управления на основе моделей глубокого
обучения**

Направление подготовки

09.06.01 «Информатика и вычислительная техника»

Направленность

05.13.01 «Системный анализ, управление и обработка информации (по отраслям)»

Код и наименование

НАУЧНЫЙ ДОКЛАД

об основных результатах научно-квалификационной работы (диссертации)

Автор работы:

Ханафи Мохамед Яссин

Научный руководитель:

д.т.н., проф.

Шкодырев Вячеслав

Петрович

Санкт Петербург – 2021

Научно-квалификационная работа выполнена в ВШ/на кафедре Института компьютерных наук и технологий федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого»

Директор ВШ/зав. кафедрой:

*Шкодырев Вячеслав Петрович,
доктор технических наук,
профессор.*

Научный руководитель:

*Шкодырев Вячеслав Петрович,
доктор технических наук,
профессор.*

Рецензент:

*Смирнов Юрий Михайлович
доктор технических наук,
профессор.*

С научным докладом можно ознакомиться в библиотеке ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого» и на сайте Электронной библиотеки СПбПУ по адресу: <http://elib.spbstu.ru>

Актуальность проблемы: Теория управления анализирует свойства управляемых систем, т. е. динамических систем, на которые можно воздействовать с помощью команды (или управления). Задача состоит в том, чтобы перевести систему из заданного начального состояния в определенное конечное состояние, возможно, с соблюдением определенных критериев. Исследуемые системы многочисленны: дифференциальные системы, дискретные системы, системы с шумом, с задержкой... Их истоки очень разнообразны: механика, электричество, электроника, биология, химия, экономика... Целью может быть стабилизация системы, чтобы сделать ее нечувствительной к некоторым возмущениям (стабилизация), или определение оптимальных решений для определенного критерия оптимизации (оптимальное управление). В современных отраслях промышленности, где преобладает понятие эффективности, роль инженера по автоматизации заключается в проектировании, внедрении и оптимизации или, по крайней мере, улучшении существующих методов. Таким образом, области применения многочисленны: аэрокосмическая, автомобильная, робототехника, авионавтика, интернет, связь в целом, а также медицинская и химическая отрасли, машиностроение и др. С математической точки зрения, система управления — это динамическая система, зависящая от динамического параметра, называемого управлением. Для ее моделирования можно использовать дифференциальные, интегральные, функциональные, конечные разности, частные производные, стохастические и т. д. уравнения. По этой причине теория управления находится во взаимосвязи многих математических полей. Управление — это функции или параметры, обычно подверженные ограничениям.

Современный оптимальный контроль и приложения: считается, что современная теория оптимального управления началась в 1950-х годах с формулировки принципа максимума Понтрягина, который обобщает уравнения Эйлера-Лагранжа при исчислении вариаций. С тех пор теория, как и множество ее применений, значительно выросла. В настоящее время автоматизированные системы являются неотъемлемой частью нашей повседневной жизни (мы часто об этом не знаем), для улучшения качества нашей жизни и облегчения определенных задач: ABS тормозные системы, помощь при вождении, серводвигателях, термостатах, контроль влажности, холодильные цепи, управление потоками и т.д. холодильные цепи, управление автомобильными, железнодорожными, воздушными, биржевыми и речными потоками, цифровая фотография, фильтрация и реконструкция изображений, компьютерные сети, поисковые системы в Интернете, электрические и электронные цепи, телекоммуникационные системы управления химическими

процессами, нефтепереработка, промышленные сборочные линии, миротворческие и другие автоматизированные медицинские системы, лазерные операции, робототехника, спутники, аэрокосмическое наведение, биореакторы, дистилляция, ... Список бесконечен, приложениями являются любые системы, на которых мы можем иметь действие, с понятием оптимальной эффективности.

Цель диссертационной работы является разработка метода иерархической Парето-оптимизации для системы оптимального управления динамической сложной иерархической структурой нефтепереработки.

Для реализации цели были поставлены следующие **задачи**:

1. Определение иерархическую сложную структуру динамической системы (нефтепереработка).
2. Определение проблему и определите требуемые знания.
3. Определение набор факторов/целей. Каждый набор верхнего уровня используется для сравнения с набором на уровне, расположенном непосредственно ниже относительно него.
4. Использование приоритеты, полученные в результате сравнения, для определения приоритетов на следующем более низком уровне.
5. Анализ каждый процесс сложной структуры иерархии с помощью нейронной сети для определения границ и поведения цели по управляющим факторам.
6. Решение многоцелевую оптимизацию для каждого процесса и получить для него фронт Парето.
7. Привязка каждый процесс к агенту, который управляет процессом, используя знания, полученные в процессе обучения, и консенсус с другими агентами для получения оптимального решения для иерархической сложной системы (нефтепереработка).

Цель исследования: Иерархическая Парето-оптимизация для интеллектуальной системы управления на основе глубокого обучения с усилением для сложной иерархической динамической системы нефтепереработки.

Предмет исследования: процесс повышения целей (качества и количества) системы нефтепереработки на основе оптимальности каждого процесса этой иерархически сложной системы.

Методы исследования: в работе использована структура Крипке для упрощения сложной структуры иерархии, алгоритмы оптимальности Парето и

фронта Парето, нейронная сеть для анализа, мультиагентная система управления, динамическая рекуррентная нейронная сеть.

Научные положения, выносимые на защиту: на защиту выносятся следующие положения:

1. Представление иерархической комплексной системы нефтепереработки с матрицей целей для основного процесса и упрощение иерархической комплексной структуры.
2. Представление результатов аналитической фазы процессов и обсуждение поведения каждого процесса.
3. Определение многоцелевой задачи оптимизации нефтепереработки и представление решения, предложенного для ее решения.
4. Описание метода иерархической Парето-оптимизации и его применение к динамической сложной иерархической структуре нефтепереработки
5. Описать мультиагентную систему управления и Парето-оптимальность процессов.
6. Представление окончательных результатов, полученных с помощью нашего метода.

Научная новизна работы заключается в следующем:

1. Модель иерархической Парето-оптимизации для оптимальной системы управления.
2. Адаптивная интеллектуальная система управления сложной иерархической динамической системой.
3. Модель многоагентной системы управления для сложной иерархической динамической системы.
4. Динамическая рекуррентная нейросетевая модель на основе глубокого обучения с усилением для многоагентной системы управления.

Практическая ценность работы будет заключаться в следующем:

1. Эффективная система управления, способная адаптироваться к окружающей среде и ее изменениям.
2. Прогнозирование неопределенного сценария и действие в зависимости от обстоятельств окружающей среды.
3. Самонастройка системы управления в зависимости от изменений в окружающей среде.
4. Учиться на опыте, особенно на нежелательных сценариях.
5. Избегание рисков и нежелательных результатов.

6. Периодическое обновление целей мультиагентной системы управления для достижения лучших результатов.

Внедрение результатов работы: Результаты диссертационной работы внедрены с использованием:

реальных данных по нефтепереработке (сгенерированных данных на основе реальных данных) для достижения наилучшей конфигурации системы нефтепереработки и сравнения результатов с нормами.

Структура и объем диссертации.

Диссертация содержит введение, пять глав, заключение, список литературы, 55 рисунков, 18 таблицы. Общий объем работы составляет 203 страниц и 269 библиографических наименования.

Личный вклад автора: Основные научные положения, математические модели, алгоритмы и их программная реализация, содержащиеся в диссертационной работе, получены автором самостоятельно.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении: обоснована актуальность исследования, поставлены цели и задачи исследования, сформулированы положения, выносимые на защиту, раскрыты методы исследования, научная новизна, теоретическая и практическая значимость исследования

мы дали общий обзор эволюции теории оптимального управления. Список ссылок, приведенных выше, не является исчерпывающим. Очевидно, что, несмотря на значительное развитие этой дисциплины, многие проблемы все еще остаются открытыми, что открывает горизонты для новых тем исследований.

В первой главе мы дали общий обзор эволюции теории оптимального управления. Список ссылок, приведенных выше, не является исчерпывающим. Очевидно, что, несмотря на значительное развитие этой дисциплины, многие проблемы все еще остаются открытыми, что открывает горизонты для новых тем исследований.

мы представляем историю создания системы оптимального управления и проблемы в этой области. Мы также иллюстрируем основную структуру нефтепереработки и кратко описываем каждый основной процесс в ней. Мы освещаем теории оптимальной системы управления, методы и связанные с ними работы, уделяя особое внимание классическим и современным системам

управления, а также описываем тезис и то, как мы будем использовать систему управления в нефтепереработке.

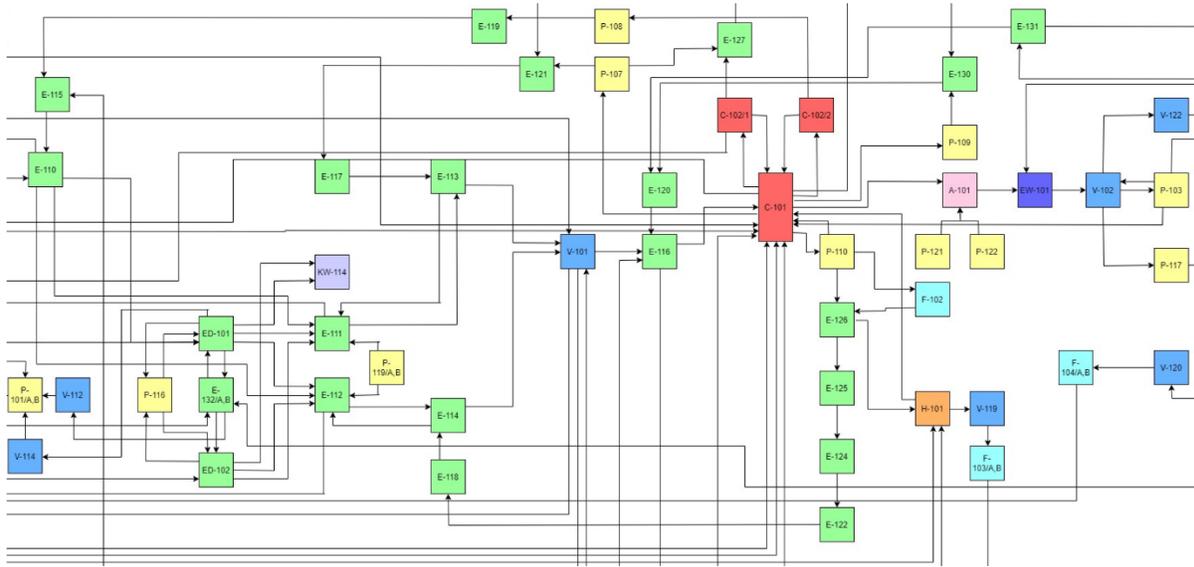


Рисунок 1 часть главной схемы нефтепереработки

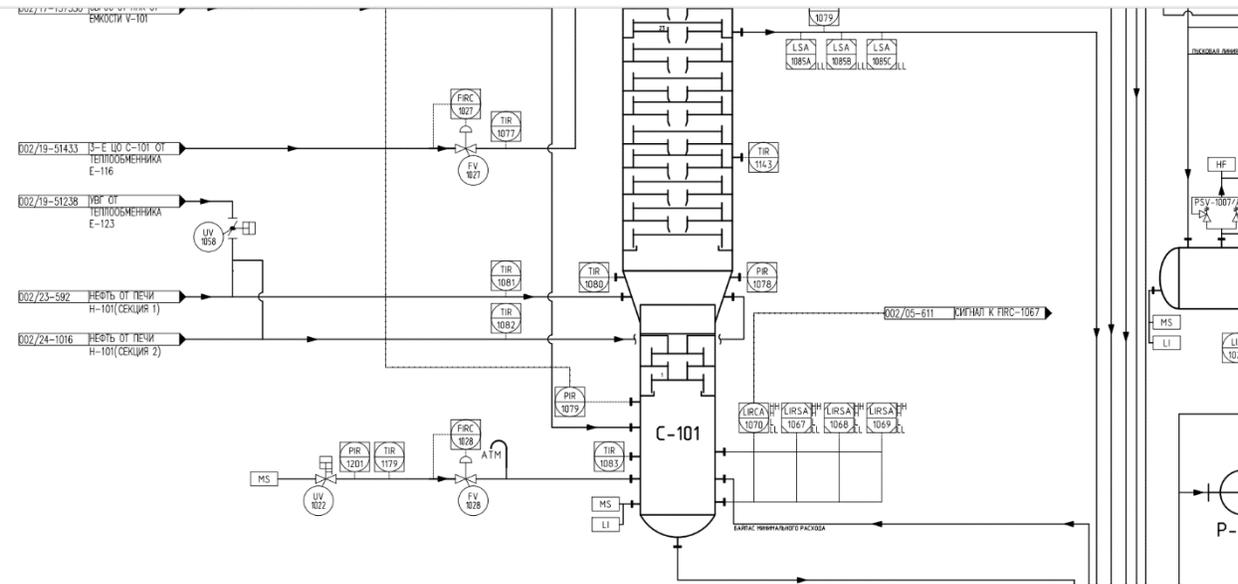


Рисунок 2 часть реальной основной схемы нефтепереработки

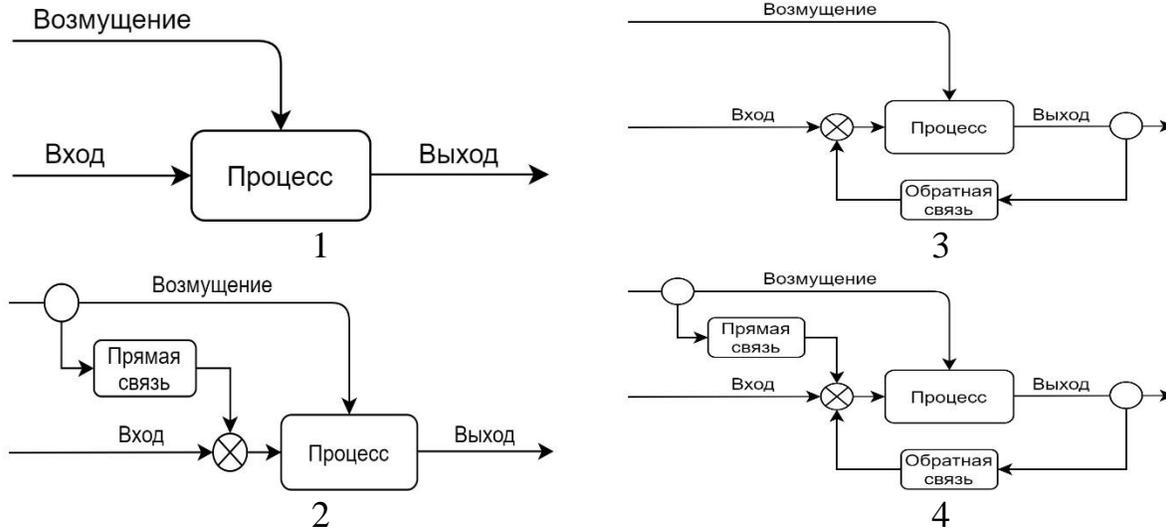


Рисунок 3 Контур систем управления

Во второй главе мы сосредоточились на подходах и теориях многоцелевой оптимизации. Мы представляем современное состояние этих подходов и методы сравнения между ними. Мы также осветили наш новый метод иерархической Парето-оптимизации с кратким описанием.

Поскольку оптимальность системы зависит от оптимальности подсистемы, отметим, что S — это наша система, где: $S \equiv \langle S_1, S_2, \dots, S_n \rangle$, где n - количество подсистем, где:

$$S_i \Leftrightarrow G_i(u_i) = [g_{i1}(u_i), g_{i2}(u_i), \dots, g_{im}(u_i)]$$

И каждая подсистема S_i имеет свои цели $G_i(u)$ (где m - количество целей) и вектор управляющих факторов u_i

$$u_i^* \equiv \underset{u \in \mathbb{R}}{\partial} \left(\|G_i^p(u_i) - G_i^t(u_i)\| \right) \Rightarrow \min$$

Для решения иерархии Парето оптимальности в системе, все функции цели подсистемы и ограничения с вектором переменной решения будут подняты в верхнюю систему.

Пусть S_n - подсистема, наследуемая от системы S .

$$S \equiv \langle S_1, S_2, \dots, S_n \rangle$$

Где решение многоцелевой оптимизации для каждой подсистемы S_n :

$$S_n \equiv \begin{cases} \min/ \max f_{nm}(x) & m = 1, 2, \dots, M; \\ g_{nj}(x) \geq 0 & j = 1, 2, \dots, J; \\ h_{nk}(x) = 0 & k = 1, 2, \dots, K; \\ x_{ni}^L \leq x_{ni} \leq x_{ni}^U & i = 1, 2, \dots, l; \end{cases}$$

И

$$F \equiv \langle f_{1m}(x), f_{2m}(x), \dots, f_{nm}(x) \rangle$$

$$G \equiv \langle g_{1j}(x), g_{2j}(x), \dots, g_{nj}(x) \rangle$$

$$H \equiv \langle h_{1k}(x), h_{2k}(x), \dots, h_{nk}(x) \rangle$$

$$X_i^L \equiv \text{lower} \langle x_{1i}^L, x_{2i}^L, \dots, x_{li}^L \rangle$$

$$X_i^U \equiv \text{Upper} \langle x_{1i}^U, x_{2i}^U, \dots, x_{li}^U \rangle$$

Для всех уравнений, приведенных выше, мы получим:

$$S \equiv \begin{cases} \min/ \max F_m(x) & m = 1, 2, \dots, M; \\ G_j(x) \geq 0 & j = 1, 2, \dots, J; \\ H_k(x) = 0 & k = 1, 2, \dots, K; \\ X_{ni}^L \leq X_{ni} \leq X_{ni}^U & i = 1, 2, \dots, l; \end{cases}$$

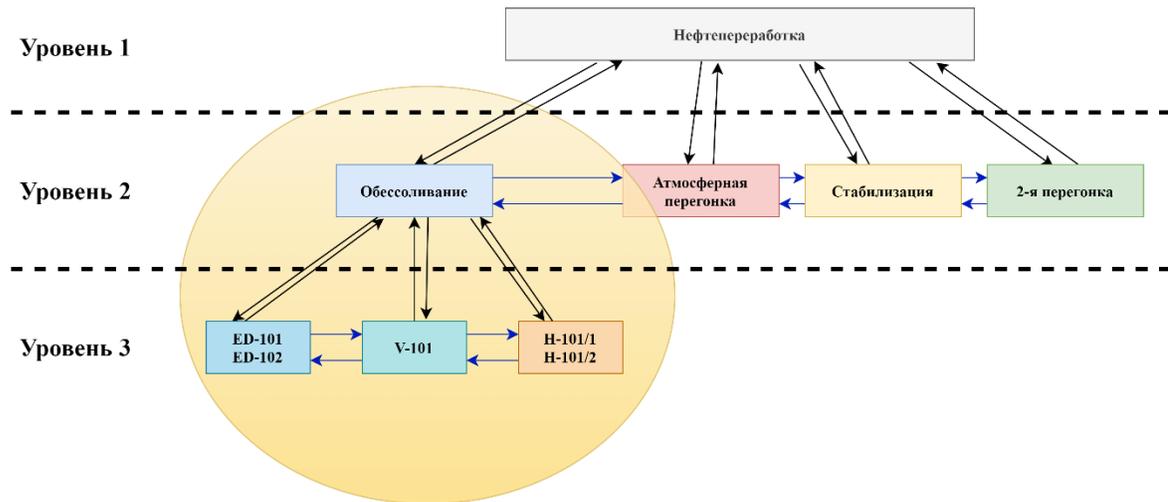


Рисунок 4 Иерархическая структура нефтепереработки

В третьей главе мы описали основы машинного обучения и кратко представили обучение с подкреплением и глубокое обучение с подкреплением, а также то, как работает агент. Мы сосредоточились на

подходах Q-обучения и стратегического обучения и на том, какой из них лучше подходит для наших нужд в иерархической динамической системе нефтепереработки.

мы сосредоточимся на определенном семействе алгоритмов обучения с усилением, которые используют методы градиента стратегий. Эти методы оптимизируют цель эффективности (обычно ожидаемое кумулятивное вознаграждение) путем нахождения хорошей стратегии (например, нейросетевой параметризованной стратегии) благодаря вариантам стохастического градиентного восхождения относительно параметров стратегии. Отметим, что градиентные методы стратегий принадлежат к более широкому классу методов, основанных на стратегиях, которые включают, в частности, эволюционные стратегии. Эти методы используют обучающий сигнал, полученный в результате выборки инстанций параметров политики, и набор политик развивается в направлении политик, которые достигают лучшей доходности. В этом обзоре мы рассматриваем случай агента обучения с усилением, целью которого является поиск стратегии $\pi(s, a) \in \Pi$ для оптимизации ожидаемой доходности $V\pi(s) : S \rightarrow R$ (также называемой функцией V-значения), такой, что:

$$V^\pi(s) = E \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} \mid s_t = s, \pi \right]$$

Где:

- $r_t = E_{a \sim \pi(s_t, \cdot)} R(s_t, a, s_{t+1})$
- $P(s_{t+1} \mid s_t, a_t) = T(s_t, a_t, s_{t+1})$ с $a_t \sim \pi(s_t, \cdot)$

Исходя из определения ожидаемой доходности, оптимальная ожидаемая доходность может быть определена как:

$$V^*(s) = \max_{\pi \in \Pi} V^\pi(s)$$

В дополнение к функции V-значения можно ввести еще несколько функций, представляющих интерес. Функция Q-значения $Q\pi(s, a) : S \times A \rightarrow R$ определяется следующим образом:

$$Q^\pi(s, a) = E \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} \mid s_t = s, a_t = a, \pi \right]$$

Это уравнение можно переписать рекурсивно, используя уравнение Беллмана

$$Q^\pi(s, a) = \sum_{s' \in S} T(s, a, s') (R(s, a, s') + \gamma Q^\pi(s', a = \pi(s')))$$

Аналогично функции V-значения можно определить и оптимальную функцию Q-значения $Q^*(s, a)$:

$$Q^*(s, a) = \max_{\pi \in \Pi} Q^\pi(s, a)$$

Особенностью функции Q-значения по сравнению с функцией V-значения является то, что оптимальная политика может быть получена непосредственно из $Q^*(s, a)$:

$$\pi^*(s) = \arg \max_{a \in A} Q^*(s, a)$$

где $R'(s, a) = \int_{s' \in S} T(s, a, s') R(s, a, s')$, а $\rho^\pi(s)$ - дисконтированное распределение состояния, определяемое как:

$$\rho^\pi(s) = \sum_{t=0}^{\infty} \gamma^t Pr\{s_t = s | s_0, \pi\}$$

Для дифференцируемой стратегии π_w фундаментальным результатом, лежащим в основе этих алгоритмов, является теорема о градиенте стратегии:

$$\nabla_w V^{\pi_w}(s_0) = \int_S \rho^{\pi_w(s)} \int_A \nabla_w \pi_w(s, a) Q^{\pi_w}(s, a) da ds$$

Этот результат позволяет нам адаптировать параметры стратегии w : $\Delta w \propto \nabla_w V^{\pi_w}(s_0)$ из опыта. Этот результат особенно интересен, поскольку градиент стратегии не зависит от градиента распределения состояния (хотя можно было бы ожидать, что он будет зависеть).

На основе этого исследования мы использовали архитектуру "актёр-критик", которая состоит из двух частей:

- Актёр относится к стратегии.
- Критика к оценке функции ценности (например, функции ценности Q).

В глубоком обучении с усилением и актёр, и критик могут быть представлены аппроксиматорами нелинейных функций нейронной сети.

Актёр использует градиенты, полученные из теоремы о градиентах стратегии и корректирует параметры стратегии w .

Критик, параметризованный θ , оценивает приближительную функцию стоимости для текущей стратегии π : $Q(s, a, \theta) \approx Q\pi(s, a)$.

В четвертой главе мы описали основную цель нашего исследования, которая заключается в создании интеллектуальной системы управления. Мы разделили наше исследование на 2 раздела, а именно: Раздел анализа и Раздел решения. Система обычно определяется как совокупность взаимосвязанных процессов, которые принимают вход и производят выход. Теория управления изучает свойства управляемых систем, которые представляют собой динамические системы, на которые можно воздействовать с помощью команды (или управления). Цель состоит в том, чтобы перевести систему из заданного начального состояния в определенное конечное состояние при возможном соблюдении определенных критериев. Целью может быть стабилизация системы таким образом, чтобы она была нечувствительна к определенным возмущениям (стабилизация) или поиск оптимальных решений для определенного критерия оптимизации (оптимальное управление). Таким образом, наша реализация основана на реальных данных, но эти данные не завершены, поэтому мы решили использовать только те участки, на которых мы можем реализовать наши подходы. Теория управления изучает возможность воздействия на динамическую систему в зависимости от переменной времени таким образом, чтобы привести состояние этой системы к заданному состоянию (S_k) в заданное время (t), но также в зависимости от коэффициента управления (u) самой системы. Целью может быть поддержание системы таким образом, чтобы она была устойчива к таким неустойчивостям (стабилизация) или нахождение оптимальных решений для определенного фактора оптимизации (оптимальное управление).

$$S_{k+1} = f(S_k, u, t)$$

$$y = f(S, u)$$

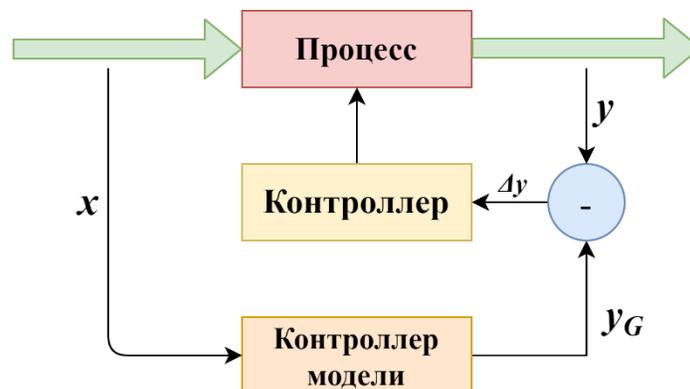


Рисунок 5 Модель системы управления

Выход (y) системы изменяется в зависимости от состояния (S) и управляющего фактора (u) системы в данный момент времени.

$$S \equiv \langle S_1, S_2, \dots, S_n \rangle$$

Для каждого состояния динамической системы (S) существует целевая задача (G), которая управляется (u) для получения оптимального результата.

$$S_i \leftrightarrow G_i(u_i) = [g_{i1}(u_i), g_{i2}(u_i), \dots, g_{im}(u_i)]$$

$$u^*_{i \in n} : \partial(\|S_G - S_t\|) \Rightarrow \min$$

$$u_i^* \equiv \partial_{u \in \mathbb{R}} (\|G_i^P(u_i) - G_i^t(u_i)\|) \Rightarrow \min$$

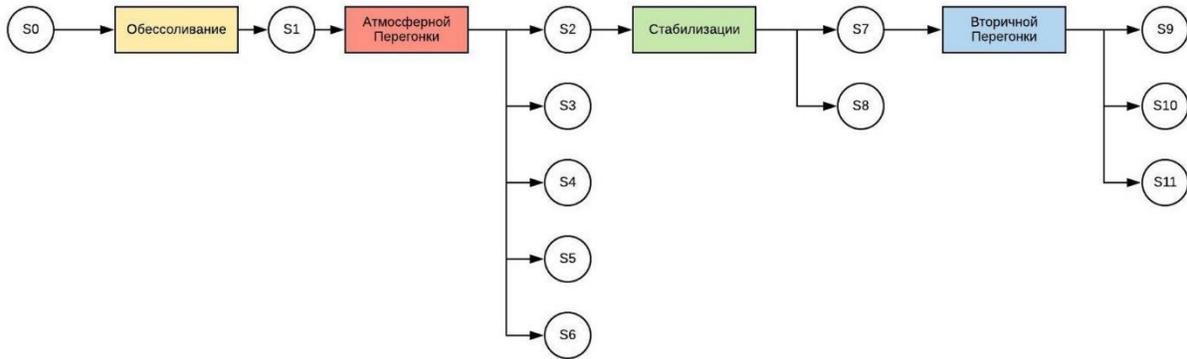


Рисунок 6 Структура нефтепереработки

Таблица 1 Матрица состояния

Состоянии		Процессы	
S ₀	Сырая Нефть	R ₁	Обессоленные
S ₁	Обессоленный Серая Нефти	R ₂	Атмосферной Перегонки
S ₂	Нестабильный Бензин	R ₃	Стабилизации
S ₃	140-240		
S ₄	240-300		
S ₅	300-350		
S ₆	Мазут		
S ₇	Стабильного Бензина	R ₄	Вторичной Перегонки
S ₈	Сжиженного Газа		
S ₉	70		

S₁₀	70-100		
S₁₁	100-140		

Целью переработки нефти является разделение сырой нефти на продукты, и все процессы связаны с типом сырой нефти, поскольку они бывают нескольких типов в зависимости от источника сырой нефти, которая в нашем случае делится на 2 типа, показанные в таблице ниже:

Таблица 2 Характеристики сырой нефти

Характеристика сырья	Сырая нефть	
	тяжелая	легкая
Плотность при 20 °С, кг/м ³	852,0	841,0
Содержание серы, % мас	0,48	0,57
Содержание воды, % мас	0,3	0,3
Содержание солей, мг/дм ³	30	30
Содержание мех. примесей, % мас.	0,0078	0,0078

На основе данных, полученных при изучении процессов нефтепереработки, мы определили факторы управления каждым процессом и их объективные показатели, а также поняли взаимосвязь между ними.

Таблица 3 Матрица целевых показателей

Матрица ЦП Блок Обессоливания			
Процесс	Факторы Влияющие/Управления	Целевые Показатели Эффективности	
		Производительность	Качестве
ED-101	Температура		
	Давление		
ED-102	Температура		
	Давление		
V-101	Температура		
	Давление		
H-101/1	Температура		
	Давление		
H-101/2	Температура		
	Давление		

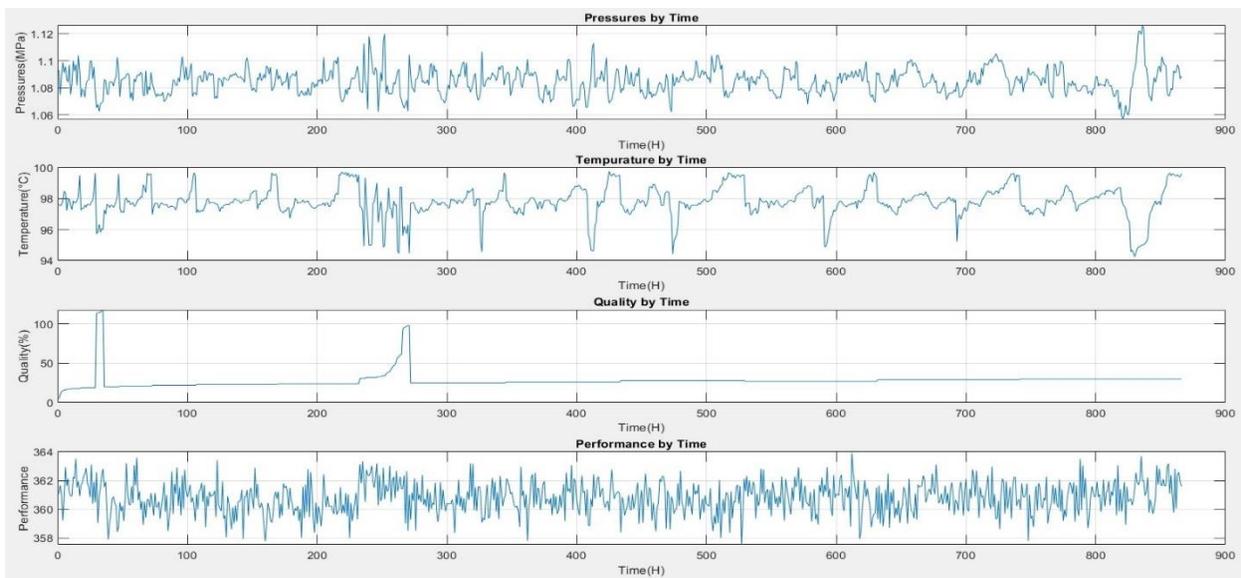


Рисунок 7 Факторы управления - по целевому показателю

Как описано выше, факторы управления оказывают большое влияние на цели, поскольку они могут увеличить или уменьшить результаты процессов, но поскольку у нас есть более одного фактора управления, нам нужно понять, как они влияют на целевой показатель, когда они оба находятся на процессах, и для этого мы использовали простую аппроксимирующую нейронную сеть, чтобы понять поведение факторов управления.

Аппроксимация функции — это метод оценки неизвестной базовой функции с использованием исторических или доступных наблюдений из области. В контролируемом обучении набор данных состоит из входов и выходов, и алгоритм контролируемого обучения учится тому, как наилучшим образом сопоставить примеры входов с примерами выходов. Мы можем считать, что это сопоставление регулируется математической функцией, называемой функцией отображения, и именно эту функцию алгоритм контролируемого обучения стремится наилучшим образом аппроксимировать. Нейронные сети являются примером алгоритма контролируемого обучения и стремятся приблизить функцию, представленную нашими данными. Это достигается путем вычисления ошибки между прогнозируемым и ожидаемым выходом и минимизации этой ошибки в процессе обучения.

Нейронная сеть, используемая в наших процессах, откладывается на основе процесса и реальных данных, используемых для фазы обучения, но из-за огромной сложности процессов мы использовали среднюю конфигурацию модели нейронной сети, чтобы гарантировать, что не произойдет перебор или недобор.

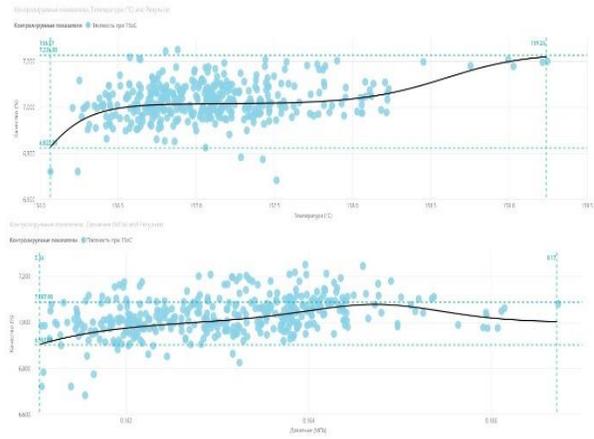
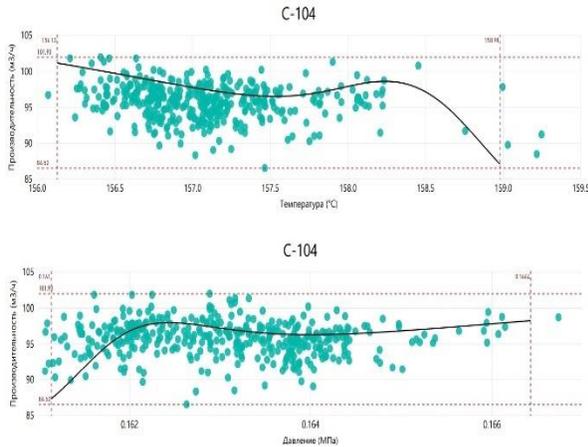


Рисунок 8 Графики функций аппроксимации

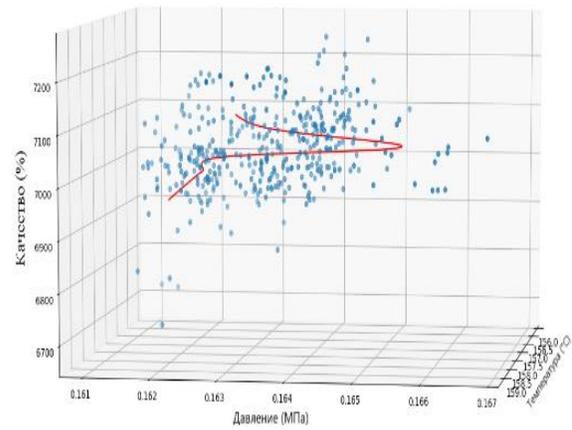
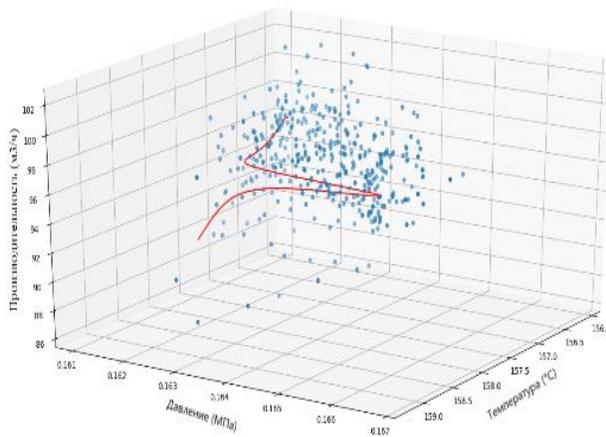


Рисунок 9 Графики функций аппроксимации

Приведенный выше рисунок наглядно иллюстрирует аппроксимационную функцию агрегатов, поскольку синие точки представляют конфигурации, а красная линия - аппроксимационную функцию.

Правая часть рисунка иллюстрирует факторы управления (температура и давление) по целевому показателю (производительность).

Правая часть рисунка иллюстрирует факторы управления (температура и давление) по целевому показателю (качество).

Каждый процесс в нашей структуре связан с интеллектуальным агентом, который наблюдает, обучается и управляет процессом, и в зависимости от самого процесса интеллектуальный агент адаптируется к процессу, что позволяет не только управлять им, но и предотвращать риски, поскольку он может понять поведение процесса. Интеллектуальный агент использует

преимущества иерархического подхода к оптимизации по Парето (см. раздел 2.4.1) и в зависимости от полученных результатов обновляет параметры для быстрого достижения своих целей (целей процесса). Каждый агент — это программное обеспечение, которое предпринимает определенные действия в зависимости от взаимодействия с окружающей средой, и каждый из них обладает стандартными свойствами, к которым относятся:

Действие — это одно из всех возможных действий, которые может совершить агент. Действие почти не требует пояснений, но следует отметить, что агенты обычно выбирают из списка дискретных, возможных действий (в нашем случае манипулирование управляющими факторами процесса).

Окружение: Мир, в котором перемещается агент и который реагирует на его действия. Среда принимает на вход текущее состояние и действие агента и возвращает на выходе вознаграждение агента и его следующее состояние.

Состояние: конкретная и непосредственная ситуация, в которой оказался агент, т. е. конкретное место и момент, мгновенная конфигурация, которая ставит агента относительно других значимых вещей.

Вознаграждение — это обратная связь, с помощью которой мы измеряем успех или неудачу действий агента в каждом состоянии (в нашем случае максимальное значение качества и производительности) (см. главу 3 раздел 3).

Фактор дисконтирования — это множитель. Будущие вознаграждения, обнаруженные агентом, умножаются на этот коэффициент, чтобы ослабить кумулятивный эффект этих вознаграждений на текущий выбор действий агента. В этом заключается суть обучения с подкреплением, т.е. постепенное уменьшение ценности будущих вознаграждений, чтобы недавние действия имели больший вес. Это критически важно для парадигмы, которая работает по принципу "отложенного действия".

Стратегия — это стратегия, которую агент использует для определения следующего действия на основе текущего состояния. Она отображает состояния на действия, действия, которые обещают наибольшее вознаграждение, в нашем случае мы использовали метод Актёр-Критик (см. главу 3, раздел 6.3).

Q-ценность (ценность действия): Q- значение похож на значение, за исключением того, что он принимает дополнительный параметр - текущее действие. Она относится к долгосрочной отдаче от действия, предпринимающего конкретное действие в рамках конкретной стратегии из текущего состояния.

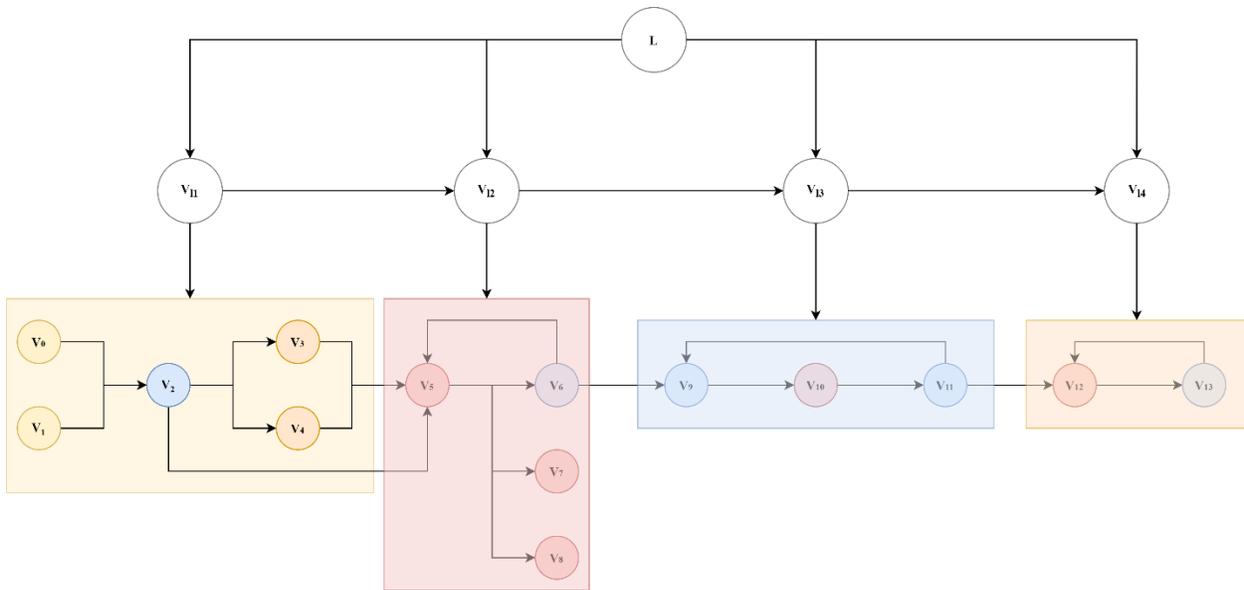


Рисунок 10 Агентская схема структуры нефтепереработки

Таблица 4 Связь агента с процессами

Агент	Процесс	Агент	Процесс
V ₀	ED-101	V ₇	C-102/1
V ₁	ED-102	V ₈	C-102/2
V ₂	V-101	V ₉	V-103
V ₃	H-101/1	V ₁₀	C-103
V ₄	H-101/2	V ₁₁	V-104
V ₅	C-101	V ₁₂	C-104
V ₆	V-102	V ₁₃	V-105

Мы изучали методы глубокого обучения с подкреплением для многоагентных систем (см. главу 3). Q-обучение сталкивается с проблемой присущей окружающей среде динамики, в то время как градиент политики страдает от дисперсии, которая увеличивается с ростом числа агентов.

Мультиагентная система может рассматриваться как сеть, состоящая из группы агентов, каждый агент называется узлом в сети.

Пусть $x_i \in R^n$ обозначает состояние узла v_i . Мы называем $G_x = (G, x)$ сетью (или алгебраическим графиком) с состоянием $x \in R^{\bar{N}n}$

Где:

$$x = [x_1^T, \dots, x_N^T]^T.$$

\bar{N} - количество узлов в сети.

Мы говорим, что все узлы достигают консенсуса, если и только если $x_i(k) \rightarrow x_j(k)$ для всех i, j как $k \rightarrow \infty$. в то время как для проблемы консенсуса с лидером v_0 , требуется, чтобы $x_i(k) \rightarrow x_j(k)$ для всех i , как $k \rightarrow \infty$, где $x_0(k)$ является траекторией состояния лидера.

На основе наших исследований и работ в этой области мы определили лучшее уравнение, которое мы можем использовать для обновления значений и улучшения стратегий для каждого агента:

1. Обновление значения:

$$V_i^{s+1}(e_i(k)) = y_i^T(k) Q_{ii} y_i(k) + u_i^{sT}(k) R_{ii} u_i^s(k) + \sum_{j \in N_i} u_j^{sT}(k) R_{ij} u_j^s(k) + V_i^s(e_i(k+1))$$

2. Совершенствование стратегии

$$u_i^{s+1}(k) = \operatorname{argmin}_{u_i(k)} \left(y_i(k)^T Q_{ii} y_i(k) + u_i^s(k)^T R_{ii} u_i^s(k) + \sum_{j \in N_i} u_j^s(k)^T R_{ij} u_j^s(k) \right) + V_i^{s+1}(e_i(k+1))$$

Полученные результаты были удовлетворительными по сравнению с имеющимися у нас данными.

Таблица 5 Характеристики фракции 100–180 °С

Фракция 100–180 °С	
Характеристики	Норма по нормативному документу
Плотность, при 15 °С,	750-760
НК, не менее	105
10, не более	106
50, не более	121
90, не более	154
95, не более	160
КК, не более	168
Сумма Сб, % масс, не более	0.5

В таблице выше представлены характеристики фракции 100–180 °С, полученной с технологической установки С-104.

Синие точки - возможные конфигурации, которые разрабатываются агентом, подключенным к процессу С-104, красная линия - множество Парето, достигнутое агентом после многих итераций и обучения, зеленая линия - цель, поставленная агентом в качестве задачи для достижения.

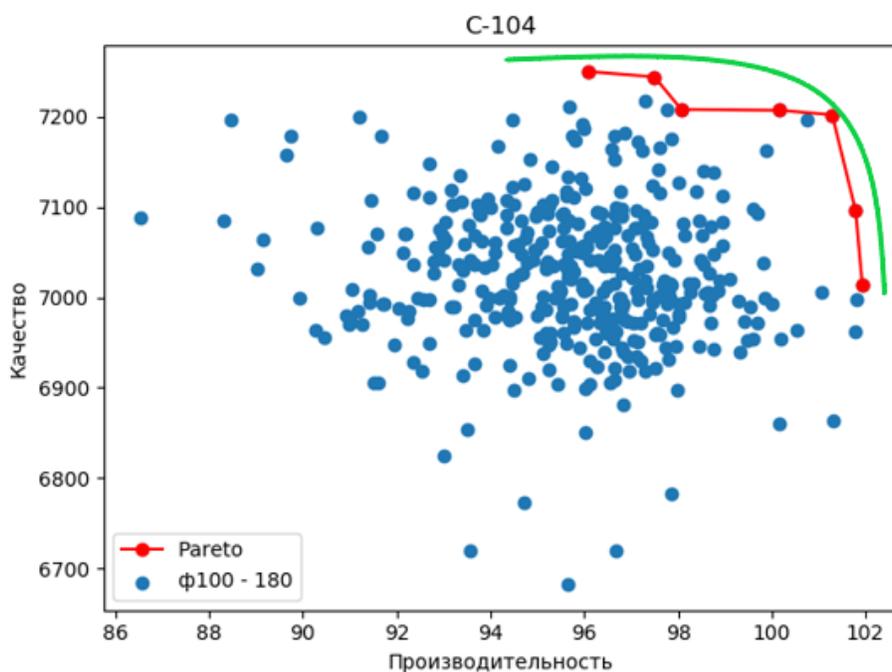


Рисунок 11 Фронт Парето С-104

Таблица 6 Фронт Парето С-104

Парето	Температура(°С)	Давление (МПа)
1	156.884307861328	0.163963586091995
2	156.807205200195	0.164595454931259
3	156.646057128906	0.163790956139565
4	157.444549560547	0.163974747061729
5	157.903137207031	0.163318753242493
6	156.637451171875	0.162251457571983
7	156.409057617188	0.162886694073677

Таблица иллюстрирует результаты множества Парето как оптимальное решение для процесса.

На каждом шаге агент получает информацию о том, как работают предыдущий и последующий процессы, и на основе этой информации он обновляет свою стратегию и значение, чтобы достичь оптимального управления своим процессом.

Агент определяет свою цель, чтобы максимизировать качество и производительность, каждый раз он разрабатывает новую стратегию и определяет новое значение, достижение которого зависит от окружающей среды. Красная линия представляет оптимальное значение, которого достиг агент (множество Парето), а зеленая полоса - новое значение, которого пытается достичь агент (фронт Парето).

В пятой главе мы представили инструменты и некоторые коды некоторых наших подходов, использованных для реализации. Затем мы представили различные интерфейсы, разработанные для нашего решения. Наконец, будут обсуждены полученные результаты.

Наша реализация основана на языке программирования Python, используя преимущества богатых библиотек, которыми он обладает, а именно: Scikit-learn, NumPy, TensorFlow, PyTorch, PyBrain, Keras, Pandas, Matplotlib.

Для реализации интерфейса мы использовали Visual Studio (язык C#), для визуализации происходящей модификации на основе результатов кода python (считывание данных с выходов python), а также использовали многопоточность для проведения нескольких расчетов одновременно.

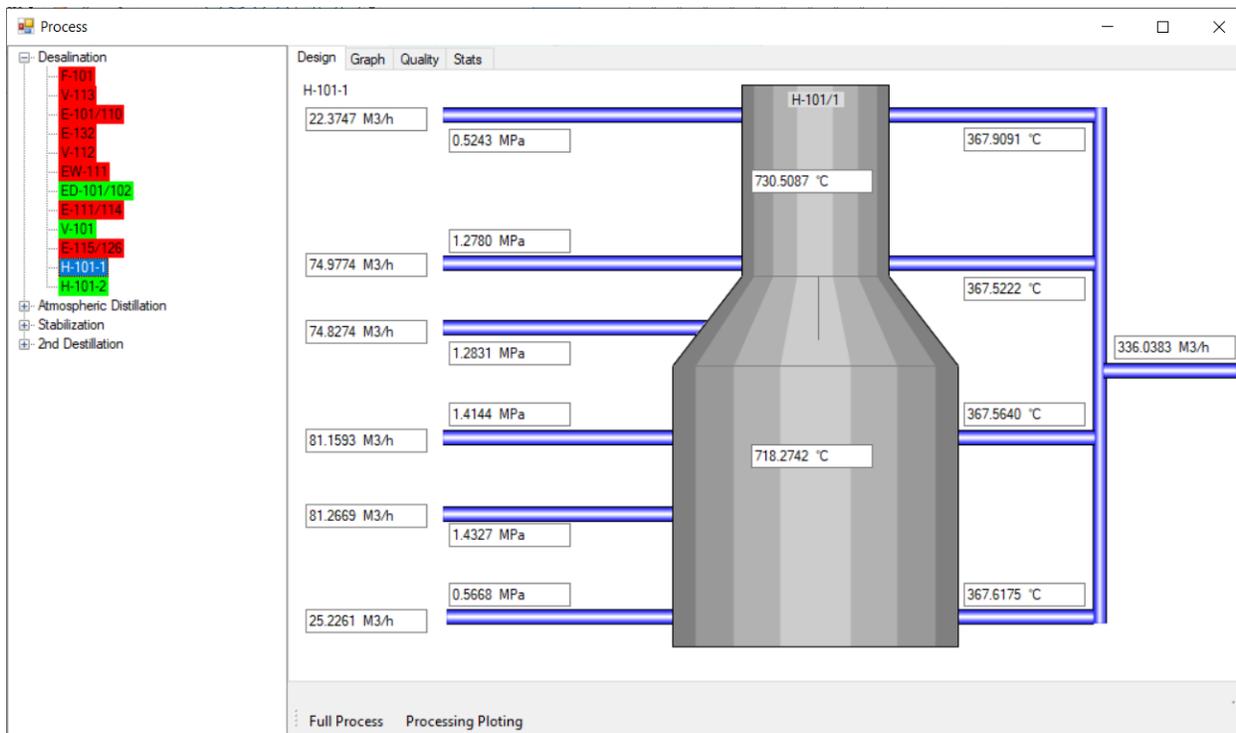


Рисунок 12 Интерфейс процесса H-101/1

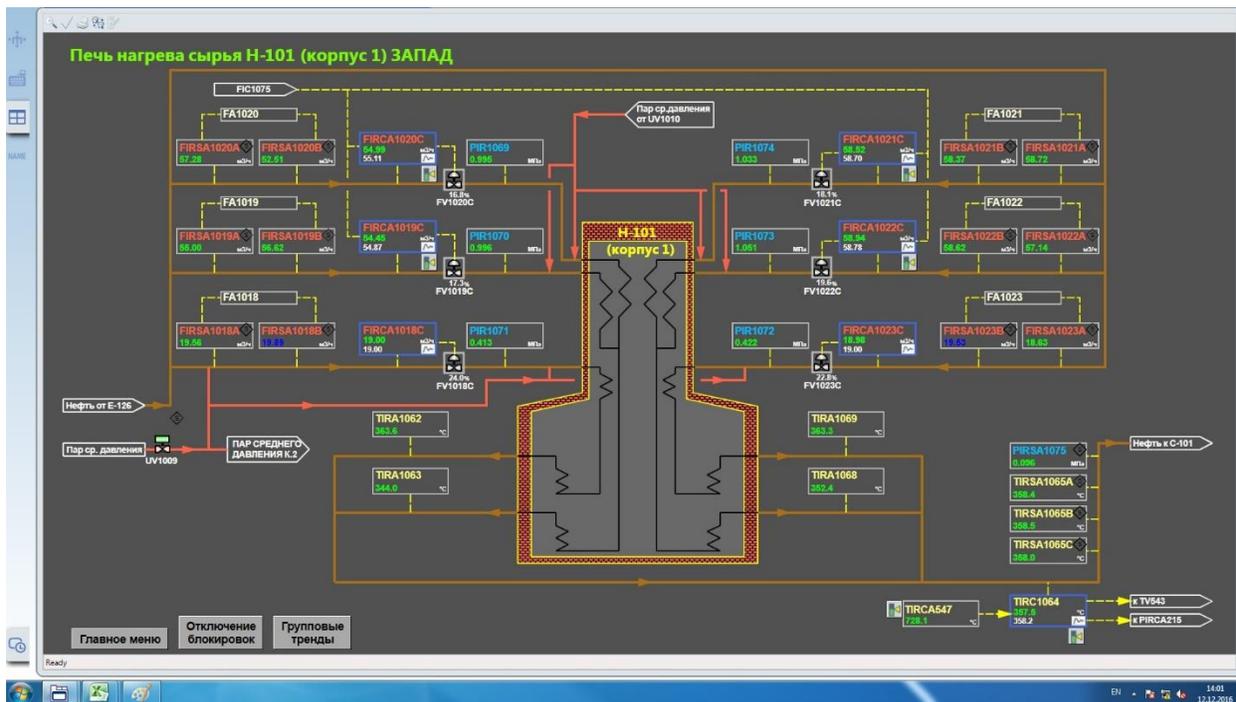


Рисунок 13 Интерфейс SCADA H-101/1

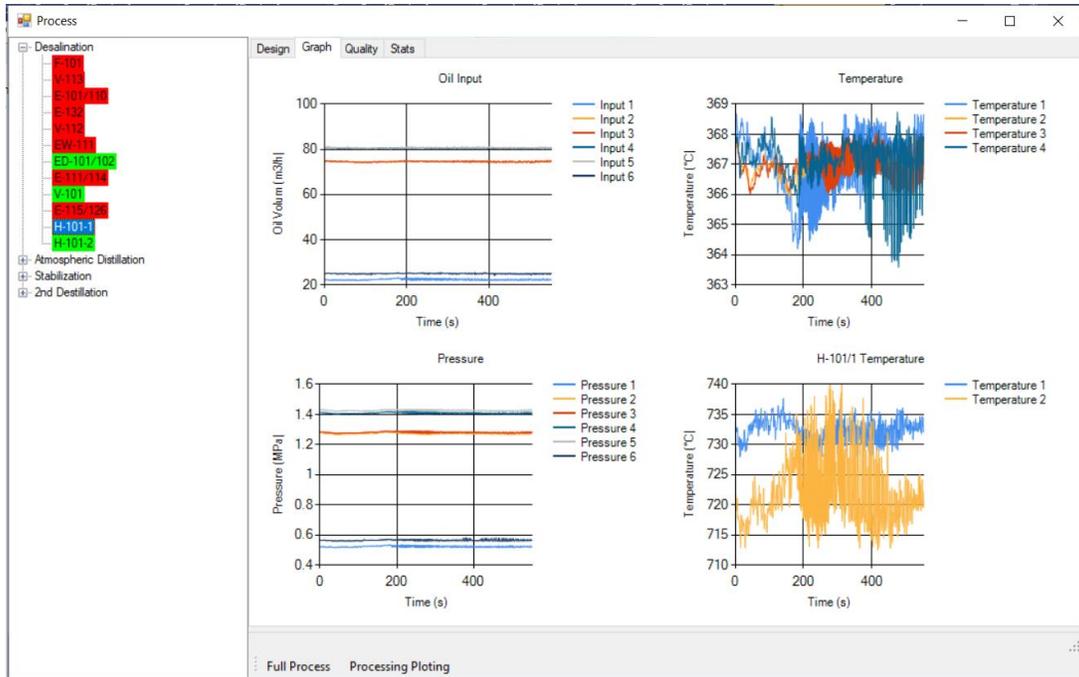


Рисунок 14 Графическая реализация процесса H-101/1

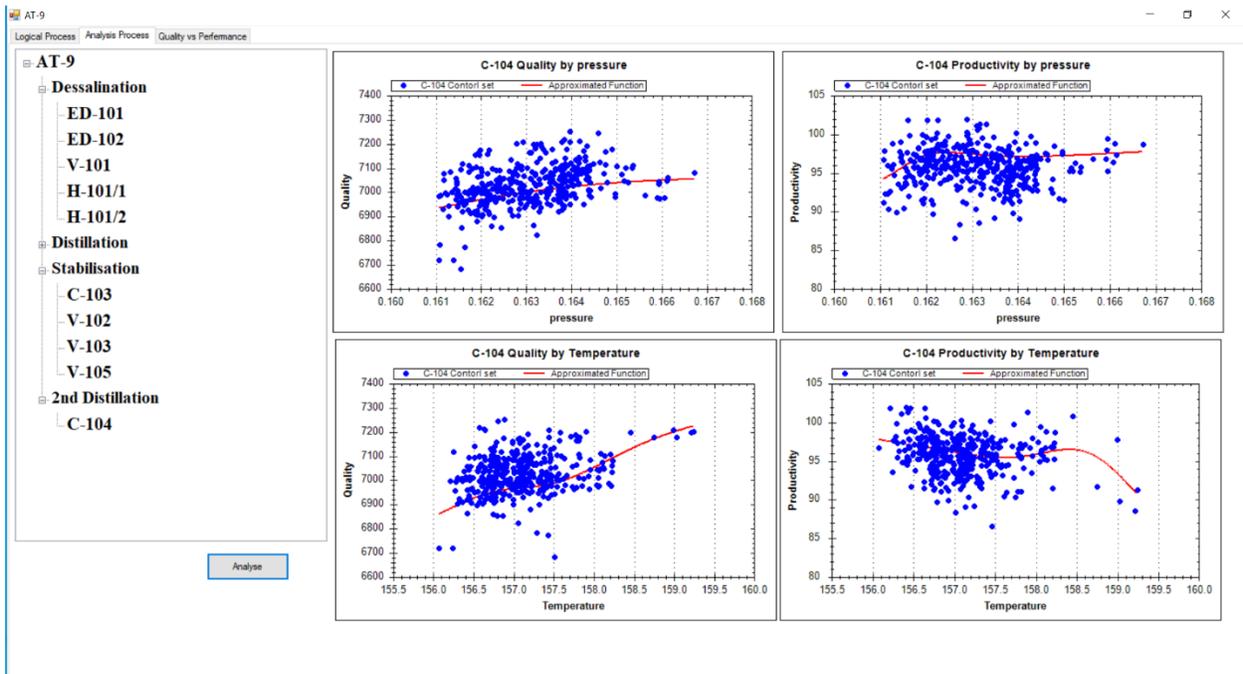


Рисунок 15 Интерфейс функции аппроксимации

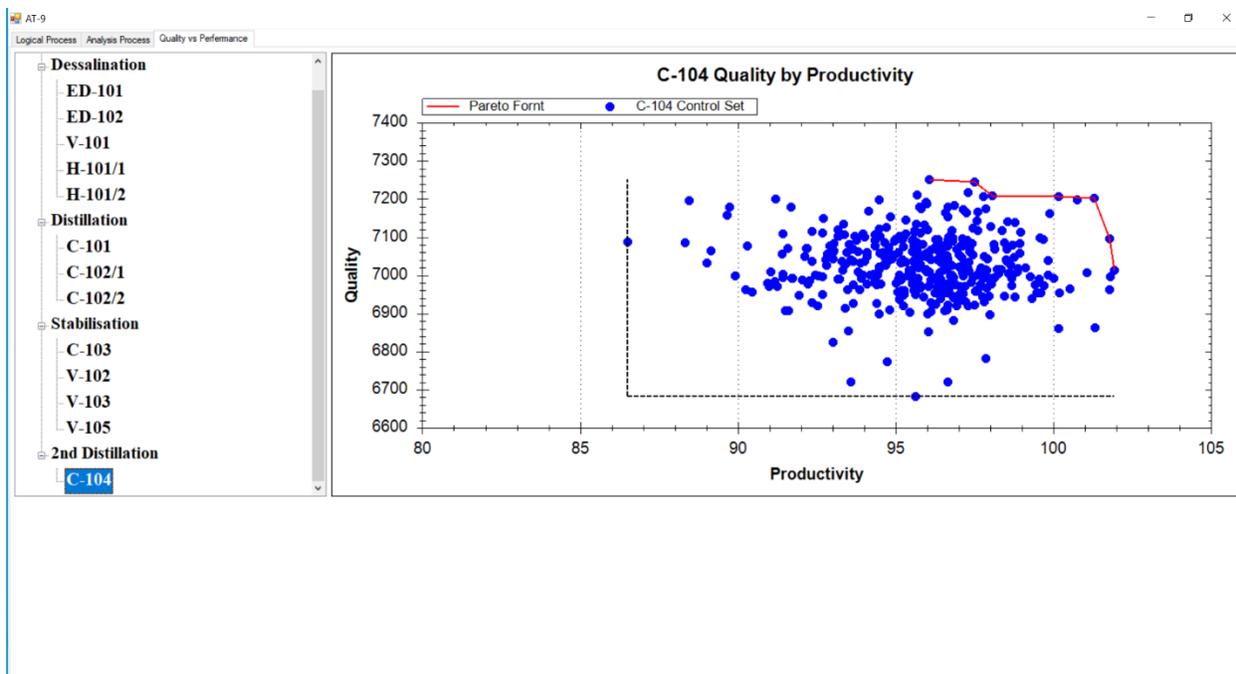


Рисунок 16 интерфейс "Фронт Парето"

Раздел реализации был настолько сложным из-за большого объема данных, которые обучаются, анализируются и генерируются, а также из-за того, что нейросетевая модель обрабатывает большое количество данных не только на одном, но и на нескольких процессах одновременно. Многопоточная стратегия, используемая для реализации модели даже при использовании технологии CUDA в NVIDIA GPU 1060 6 ГБ и Intel® Core™ i7-7700HQ CPU @ 2.80GHz, была сложной.

Из-за этих трудностей мы решили выполнить их по частям (вы можете заметить файл сохранения на кодах) и использовать полученные данные для следующего процесса. Даже несмотря на эти трудности, мы получили удовлетворительный результат.

Заключение

Последовательное принятие решений остается активной областью исследований со многими теоретическими, методологическими и экспериментальными проблемами, которые все еще остаются открытыми. Важные разработки в области глубокого обучения способствовали появлению множества новых направлений, где методы обучения с усилением и глубокого обучения сочетаются. В частности, глубокое обучение принесло важные возможности обобщения, что открывает новые возможности для работы с большими, высокоразмерными пространствами состояний и/или действий.

Есть все основания полагать, что это развитие продолжится в ближайшие годы, с более эффективными алгоритмами и множеством новых приложений.

Многие реальные задачи управления имеют неизвестную динамику, что затрудняет применение стандартных решений в области техники управления. Более того, типичные задачи управления требуют эффективных решений, т.е. таких, которые достигают цели при минимизации затрат, что требует формы оптимальности в управлении, которую трудно вычислить для нелинейных областей задач. Следовательно, оптимальные или близкие к оптимальным решения нелинейных задач управления с неизвестной динамикой остаются открытой проблемой как в области управления, так и в робототехнике и машинном обучении. В данной диссертации предложена схема, мотивированная инженерными соображениями, для поиска решений по управлению системами с динамикой нелинейных задач, которая может быть либо известна априори, либо неизвестна вообще.

Общий дизайн структуры Парето-оптимальности иерархии был представлен в главе 2. С чисто инженерной точки зрения, он был мотивирован использованием абстракций, уменьшением размерности состояния и сокращением количества необходимых контроллеров, при этом стремясь использовать лучшее из двух миров - обучение с усилением и оптимальное управление. Все это естественно предполагает иерархическую архитектуру, которая позволяет реализовать методы оптимизации для глобального решения по управлению пониженной сложности, естественно подходящие для обучения с применением усиления. С инженерной точки зрения исследования Обучение с усилением и Оптимальное управление были мотивированы теорией оптимального управления, открытием моторных примитивов на уровне спинного мозга, существованием оптимальных путей обратной связи — все это указывает на пригодность оптимальных контроллеров обратной связи на низком уровне управления. И в равной степени: символический/иерархический выбор/представление действий и реализация как модельных, так и безмодельных алгоритмов обучения усилению — все это указывает на целесообразность использования агента обучения усилению на высоком уровне иерархии. Поэтому было выбрано иерархическое построение, с использованием высокоуровневого агента обучения усилению и низкоуровневых контроллеров обратной связи.

Данная работа находится в рамках оптимального управления иерархической структурой и посвящена управлению их глобальным поведением. Проблема состоит в том, чтобы выяснить, как:

1. Система функционирует со всеми своими подсистемами.

2. Факторы управления вмешиваются в целевой показатель каждой подсистемы

Когда мы понимаем систему, возникает вторая проблема - как ее решить, чтобы достичь оптимального управления, и это путем правильного выбора подходов искусственного интеллекта.

В данной диссертации мы определили нашу иерархическую динамическую структуру нефтепереработки, затем структуру, факторы управления и целевые показатели каждой подсистемы, представленные в главе 01 и подробно описанные в главе 04.

Мы также подробно описали проблемы многоцелевой оптимизации наряду с алгоритмами и предложенным подходом иерархической Парето-оптимизации, представленным в главе 02 и подробно рассмотренным, и примененным к нашей структуре в главе 04.

В главе 03 мы уделили внимание подходам машинного обучения, описали множество подходов к обучению и сосредоточились на этапе обучения мультиагентной системы, использованном в главе 04.

В главе 05 мы иллюстрируем некоторые коды и библиотеки, использованные в нашей реализации, а также некоторые интерфейсы и результаты. Мы обращаем внимание на проблемы, с которыми мы столкнулись, и на аппаратное обеспечение, использованное для реализации.

В заключение, интеллектуальная система нейроруправления, основанная на глубоком обучении с подкреплением для иерархической динамической системы нефтедобычи, является сложной темой, которая имеет много интересного и содержит много теории, учитывая новые и старые подходы, и использует много методов, а именно: Структура Крипке для упрощения структуры, аппроксимационная функция для понимания структуры, оптимальность иерархии по Парето, динамическое программирование, подходы оптимальной системы управления, нейронная сеть, мультиагентная система и т.д.

Несмотря на то, что мы столкнулись со многими трудностями, мы получили удовлетворительные результаты, которые по сравнению с реальными данными просто великолепны.

Методы, использованные в данной диссертации, рассматриваются не только для нефтедобычи, но могут быть использованы в различных областях и отраслях промышленности.

Список работ, опубликованных по теме диссертации

публикации в изданиях, входящих в международные реферативные базы данных и системы цитирования

Журнал 1 квартал

Hanafi Mohamed Yassine and Vyacheslav P. Shkodyrev. 2020. The intelligent control system of optimal oil manufacturing production. In 2020 The 3rd International Conference on Computational Intelligence and Intelligent Systems (CIIS 2020). Association for Computing Machinery, New York, NY, USA, 131–135. DOI: <https://doi.org/10.1145/3440840.3440848>

Журнал 3 квартал

Yassine, H.M., Shkodyrev, V.P.: Optimal Production Manufacturing Based on Intelligent Control System. In: Schaumburg, H., Korablev, V., and Ungvari, L. (eds.) Technological Transformation: A New Role for Human, Machines and Management. pp. 210–220. Springer International Publishing, Cham (2021).

Публикации в изданиях, рекомендованных ВАК РФ

Hanafi M.Ya., Shkodyrev V.P. Reinforcement learning for industrial manufacturing control system. Computing, Telecommunications and Control, 2021, Vol. 14, No. 1, Pp. 60–69. DOI: 10.18721/JCSTCS.14106.