

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ



УТВЕРЖДАЮ
Декан РТФ

УТВЕРЖДАЮ /А.Н. Дедов/
(Ф.И.О. декана (директора института))

13.02.2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б.1.2.5 Математические основы обучения с подкреплением

(код и наименование дисциплины по учебному плану)

Направление подготовки
(специальность)

09.03.02 Информационные системы и технологии

Квалификация выпускника

Бакалавр

(бакалавр/магистр/специалист)

Направленность

Интеллектуальные информационные системы и
технологии

Курс 3
Семестр 5

Распределение учебного времени

Трудоемкость по учебному плану	144 / 4	часов/зачетных единиц
Лекции	36	часов
Лабораторные работы	-	часов
Практические занятия	54	часов
Иная контактная работа	-	часов
Всего контактной работы (без учета экз.)	90	часов
Контактная работа по экзамену	-	часов
Курсовой проект (работа)	-	семестр
Самостоятельная работа обучающихся (без учета экз.)	54	часов
Самостоятельная работа по подготовке к экзамену	-	часов
Экзамен	-	семестр
Зачет	-	семестр
БРК, ДЗ	5	семестр

(год)

Программа составлена в соответствии с требованиями ФГОС ВО направления подготовки (специальности) 09.03.02 Информационные системы и технологии

Программу составили:

старший преподаватель	ИТРС	СОГЛАСОВАНО	Н.А. Конкин
(должность)	(кафедра)		(И.О. Фамилия)

РАССМОТРЕНА и ОДОБРЕНА на заседании кафедры, за которой закреплена дисциплина
Кафедра информационных технологий, радиотехники и связи

(наименование кафедры)		
13.01.2025	протокол №	2
(дата)		

Заведующий кафедрой	СОГЛАСОВАНО	Н.В. Рябова
		(И.О. Фамилия)

Рабочая программа СОГЛАСОВАНА с факультетом (институтом), выпускающей(ими) кафедрой(ами).

СООТВЕТСТВУЕТ действующей ОП.

Заведующий кафедрой	СОГЛАСОВАНО	Н.В. Рябова
		(И.О. Фамилия)

Председатель методической комиссии факультета (института), в который входит выпускающая кафедра

СОГЛАСОВАНО	А.Н. Дедов
	(И.О. Фамилия)

Эксперт(ы): Дудин Александр Николаевич , Исполнительный директор ООО «Трэвел Лайн Системс»

Рабочая программа проверена и зарегистрирована в УМЦ 17.02.2025 г.

Специалист учебно-методического центра СОГЛАСОВАНО /Т.А. Смирнова/

Раздел 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является достижение планируемых результатов обучения, соответствующих установленным в ОПОП индикаторам достижения компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения
1. MF-1 Способен применять современную теоретическую математику для разработки новых алгоритмов и формулирования перспективных задач искусственного интеллекта	MF-1.2. Применяет аппарат теории вероятностей, математической статистики и теории информации для формулирования и анализа задач искусственного интеллекта.	знания: Знает ключевые концепции статистического машинного обучения, различия между основными подходами (генеративные/дискриминативные модели, линейные/нелинейные методы, параметрические/непараметрические алгоритмы), а также принципы выбора модели с учётом структуры данных, шумов, размерности и объёма выборки. умения: Умеет анализировать свойства данных, подбирать подходящие методы статистического обучения под конкретную задачу, обосновывать выбор алгоритма, интерпретировать различия между моделями и их влиянием на качество, устойчивость и обобщающую способность. навыки: Владеет навыками практического применения статистических методов: построением, обучением и валидацией моделей; сравнением подходов по метрикам и экспериментальным протоколам; корректной интерпретацией результатов; выбором оптимального метода под ограничения задачи и данных.

2. MF-4 Способен применять статистические методы для анализа данных, валидации моделей машинного обучения и проведения экспериментов в области искусственного интеллекта	MF-4.1. Применяет статистические методы анализа и машинного обучения для решения задач анализа данных и проведения экспериментов на данных.	знания: Знает основные концепции обучения с подкреплением, структуру MDP, свойства марковских процессов и их связь с динамическими системами; понимает принципы моделирования переходных вероятностей, стохастических политик и функции ценности; знает методы динамического программирования, Монте-Карло, TD-обучения и современные нейросетевые подходы (DQN, Actor-Critic, PPO) для представления сложных многомерных временных процессов. умения: Умеет анализировать многомерные временные ряды и поля, выделять состояния, действия и зависимости; строить формальные модели MDP для динамических систем; применять методы DP, MC, TD, Q-learning, Actor-Critic и нейросеточную аппроксимацию для оценки функций ценности и оптимизации политики; корректно выбирать стратегии исследования и эксплуатации; использовать современные RL-алгоритмы для моделирования и прогноза динамики сложных систем. навыки: Владеет практическими навыками построения и обучения RL-агентов для задач моделирования и управления многомерными процессами; инструментами аппроксимации (линейные модели, нейросети), экспериментальной валидации, A/B-тестирования и статистического анализа; методами интерпретации поведения агента и свойств динамической модели; полным циклом разрабатываемой системы — от формализации среды до оценки устойчивости и качества работы обученного агента.
---	--	--

MF-4.2. Способен применять статистические методы для построения предсказательных моделей, включая методы для анализа и прогнозирования временных рядов, а также моделирования нестационарных случайных процессов.	знания: Знает основные принципы формализации динамических систем в терминах MDP, свойства марковских процессов, вероятностных переходов и стохастических политик; понимает, как временная зависимость и многомерность данных влияют на структуру состояния, динамику переходов и функции ценности. Осознаёт подходы аппроксимации динамики — от классического динамического программирования до нейросетевых моделей в RL. умения: Умеет строить модель динамической системы на основе многомерных временных рядов и полей: выделять состояния, определять действия и структуру вознаграждений, оценивать переходные вероятности; применять методы DP, Monte-Carlo и TD-обучения для получения оценок функций ценности; использовать Q-learning, SARSA и Actor-Critic для восстановления динамики в условиях неопределённости. Корректно подбирать методы аппроксимации (линейные или нейросетевые), учитывая размерность и стохастичность данных. навыки: Владеет практическими техниками моделирования сложных динамических процессов: тренировкой RL-агентов (DQN, Double DQN, PPO, DDPG, A3C) для восстановления скрытой динамики систем; настройкой стратегий исследования–эксплуатации; анализом сходимости и устойчивости оценок; валидацией моделей через A/B-тестирование, доверительные интервалы и байесовские подходы. Уверенно использует методы RL для построения, тестирования и интерпретации моделей многомерных временных рядов и полей.
--	--

MF-4.3. Способен применять статистические методы для оценки качества моделей ИИ, включая метрики и критерии для регрессии, классификации и кластеризации, а также для проведения статистических тестов для сравнения моделей.	знания: Знает основные статистические критерии для сравнения моделей и алгоритмов (классические, модифицированные и байесовские подходы), принципы А/В-тестирования и способы построения доверительных интервалов в условиях стохастичности RL-процессов. Понимает ограничения классических тестов в задачах обучения с подкреплением, влияние коррелированных траекторий, стохастических политик и вариативности вознаграждений. умения: Умеет применять модифицированные критерии для сравнения RL-алгоритмов (коррекция на зависимость данных, стратификация по эпизодам, бутстрэп и байесовские методы), проводить корректные А/В-тесты для политик, вычислять и интерпретировать доверительные интервалы для метрик RL (value, return, regret), оценивать статистическую значимость различий между моделями на основе траекторных данных. навыки: Владеет практическими техниками статистической валидации RL-агентов: реализацией экспериментальных протоколов (fixed seeds, параллельные траектории, randomization tests), использованием байесовских тестов (Bayesian comparison, ROPE, probability of superiority), построением модифицированных доверительных интервалов для стохастических процессов и интеграцией этих методов в анализ результатов А/В-тестов в обучении с подкреплением.
--	--

	MF-4.4 Обосновывает способы и варианты применения алгоритмов обучения с подкреплением в задачах ИИ, включая их преобразование и адаптацию к специфике задачи	знания: Знает основы марковских процессов и динамических систем, структуру MDP, свойства стохастических переходов, особенности многомерных временных рядов и пространственно-временных полей; понимает, как состояние, действие и вознаграждение описывают динамику сложных систем. умения: Умеет формализовать динамическую систему в виде MDP, выделять признаки состояния из многомерных временных рядов, строить модели переходов и функций ценности; применять методы динамического программирования, Монте-Карло и TD-обучения для анализа эволюции систем. навыки: Владеет практическими навыками построения и анализа моделей динамики: реконструкцией пространства состояний, аппроксимацией многомерных переходных функций, моделированием взаимодействующих процессов и полей; использует RL-методы для прогнозирования поведения и контроля таких систем.
--	---	---

Раздел 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина относится к части, формируемой участниками образовательных отношений ОПОП.

Дисциплина является обязательной

Для продолжения формирования заявленных компетенций необходимы знания предшествующих практик: Учебная практика. Ознакомительная практика (MF-1), Учебная практика. Ознакомительная практика (MF-4)

Изучаемая дисциплина является основой для продолжения формирования указанных компетенций в следующих дисциплинах: Математические основы обучения с подкреплением (MF-1), Математические основы обучения с подкреплением (MF-4); практиках: Производственная практика. Научно-исследовательская работа (MF-1); государственной итоговой аттестации в форме: Выполнение и защита выпускной квалификационной работы (MF-1), Выполнение и защита выпускной квалификационной работы (MF-4)

Раздел 3. ОПИСАНИЕ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

Для формирования заявленных компетенций используются методологические технологии, реализующие деятельностный, личностно-ориентированный, практико-ориентированный подходы.

Основными стратегическими технологиями являются: лекционные занятия, практические занятия

На достижение конкретных целей обучения направлены применяемые тактические технологии: задания, игровое проектирование, информационные, мини-проекты

Раздел 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

5 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
№1 Математические основы и постановка задачи RL	28	MF-1, MF-4
Лекция. Лекция 1 (4ч) Введение в обучение с подкреплением. Постановка задачи и связь с другими подходами ML.	4	
Лекция. Лекция 2 (2ч) Марковские процессы принятия решений (MDP). Состояния, действия, вознаграждения,	4	
Лекция. Лекция 3 (3ч) Свойства марковских процессов, вероятностные переходы, стохастические политики.	3	
Практическое занятие. • Создание среды RL (GridWorld, OpenAI Gymnasium) • Реализация простого MDP • Анализ свойств стационарности и вероятностных	13	
Задания для самостоятельной работы, в том числе выполнение • Выполнить обзор источников по применению RL в интеллектуальных системах (например, робототехника, управление трафиком, финансы). • Ответить письменно на контрольные вопросы: различие между MDP и задачей принятия решений без памяти; роль вознаграждения; свойства стационарности	4	MF-1, MF-4
№2 Динамическое программирование и Монте-Карло	27	
Лекция. Лекция 4 (3ч) Динамическое программирование: итерации по ценности и политике.	3	
Лекция. Лекция 5 (3ч) Методы Монте-Карло. Оценка функции ценности и оптимизация политик.	3	
Лекция. Лекция 6 (3ч) TD (Temporal Difference Learning) - обучение (TD(0), TD(?)). Сходимость, статистические свойства TD-ошибки.	3	
Практическое занятие. • Реализация динамического программирования • Монте-Карло оценка функции ценности • Анализ распределения TD-ошибки и проверка гипотез о сходимости search / random search.	14	
Задания для самостоятельной работы, в том числе выполнение Реализовать итерацию по ценности для заданного MDP (GridWorld 4x4), визуализировать процесс сходимости.	4	MF-1, MF-4
№3 Политики и аппроксимация функций	33	
Лекция. Лекция 7 (2ч) Q-learning и SARSA. Анализ стохастических градиентов и статистических оценок.	2	
Лекция. Лекция 8 (4ч) Исследование–эксплуатация: ϵ -жадные, Softmax и байесовские стратегии.	4	
Лекция. Лекция 9 (3ч) Аппроксимация функции ценности: линейные и нейросетевые подходы.	3	
Практическое занятие. • Q-learning и SARSA в дискретной среде • ϵ -жадная и Softmax стратегия	14	

Задания для самостоятельной работы, в том числе выполнение		
• Реализовать алгоритмы Q-learning и SARSA для одной среды (например, CliffWalking) и сравнить скорости сходимости.		
• Исследовать влияние параметра γ на эффективность γ -жадной стратегии: построить зависимость среднего	10	
№4 Современные методы RL и статистическая валидация	56	MF-1, MF-4
Лекция. Лекция 10 (3ч) Actor-Critic и стохастические градиентные методы.	2	
Лекция. Лекция 11 (3ч) Современные методы RL: DQN, Double DQN, PPO, DDPG, A3C	2	
Лекция. Лекция 12 (3ч) Валидация RL-агентов: A/B тестирование, доверительные интервалы, байесовские тесты.	3	
Практическое занятие. • Реализация Actor-Critic, DQN и PPO • A/B тестирование стратегий, оценка различий моделей	13	
Задания для самостоятельной работы, в том числе выполнение РГР		
Задания для самостоятельной работы, в том числе выполнение Расчётно-графическая работа по задаче от ООО "ОМЕГА СОФТ" на тему «Проектирование и обучение RL-агента для интеллектуального мониторинга и восстановления работоспособности ИС в системе СФИНКС»	36	
Иная контактная работа:	0	

Раздел 5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Подготовка к лабораторным занятиям включает ознакомление с планом занятия, работу с учебной и учебно-методической литературой, научными изданиями и электронными образовательными ресурсами, рекомендованными рабочей программой дисциплины. Содержание самостоятельной работы определяется рабочей программой дисциплины, оценочными и методическими материалами, заданиями и указаниями преподавателя. Самостоятельная работа может осуществляться в аудиторной и внеаудиторной формах. Эффективным средством осуществления самостоятельной работы является электронная информационно-образовательная среда университета, которая обеспечивает доступ к образовательной программе, рабочей программе дисциплины (модуля), к электронным библиотечным системам, профессиональным базам данных и информационным справочным системам.

Контроль и аттестация

Периодичность проведения, формы текущего контроля успеваемости, система оценивания хода освоения дисциплины представлены в рабочей программе.

Формой промежуточной аттестации по дисциплине является БРК.

1. Необходимое программное обеспечение и инструментальные средства

Среды разработки

В качестве среды разработки для языка программирования Python рекомендуется использовать следующие решения:

1. **Visual Studio Code** (<https://code.visualstudio.com/>) — универсальный редактор с богатыми возможностями для Python-разработки
2. **PyCharm Community Edition** (<https://www.jetbrains.com/pycharm/>) — специализированная IDE для Python
3. **Jupyter Notebook/JupyterLab** (<https://jupyter.org/>) — интерактивная среда для исследовательской работы с данными

Облачные платформы для машинного обучения

Для организации удаленного коллективного решения задач машинного обучения рекомендуется использовать следующие платформы:

1. **Google Colaboratory** (<https://colab.research.google.com/>) — облачная среда с бесплатным доступом к GPU/ TPU, интеграцией с Google Drive и поддержкой Jupyter Notebook
2. **Kaggle** (<https://www.kaggle.com/>) — платформа для соревнований по машинному обучению с большим разнообразием датасетов, примеров решений и возможностью работы в облачных ноутбуках с GPU
3. **Kaggle Learn** (<https://www.kaggle.com/learn>) — бесплатные интерактивные курсы по Python, машинному обучению и глубокому обучению

2) Содержание дисциплины включает изучение исследований и разработок, представленных в научных публикациях на конференциях уровня А*, статьях, опубликованных в журналах 1 уровня Белого списка научных журналов Минобрнауки России (УММ на основе публикаций в научных журналах 1 или 2 уровня Белого списка научных журналов Минобрнауки России, публикаций на конференциях уровня А*/А.

Ниже приведен перечень конференций, рекомендуемых для подачи работ, соответствующих уровню подготовки студентов.

1. **Advances in Neural Information Processing Systems (NeurIPS)** // Conference on Neural Information Processing Systems. – URL: <https://neurips.cc/>
2. **International Conference on Learning Representations (ICLR)** // The International Conference on Learning Representations. – URL: <https://iclr.cc/>
3. **International Conference on Machine Learning (ICML)** // International Conference on Machine Learning. – URL: <https://icml.cc/>
4. **International Joint Conference on Artificial Intelligence (IJCAI)** // International Joint Conference on Artificial Intelligence. – URL: <https://www.ijcai.org/> (дата обращения: ...).
5. **AAAI Conference on Artificial Intelligence (AAAI)** // Annual AAAI Conference on Artificial Intelligence. – URL: <https://aaai.org/conference/>
6. **ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)** // Knowledge Discovery and Data Mining Conference. – URL: <https://www.kdd.org/kdd2025/>
7. **Conference on Learning Theory (COLT)** // Conference on Learning Theory. – URL: <https://learningtheory.org/colt2024/>

Предусмотрено участие ИТ-практиков в различных форматах обучения (лекции, семинары с ИТ-практиками, конкурсы, хакатоны, кейс-чемпионаты, мастер-классы, командные проектные задания и др.) и в качестве наставников.

В программе предусмотрено изучение кейсов (кейсы для изучения, практической и проектной работы) и реальных задач (проектов по разработке, внедрению, развитию цифровых технологий, ИТ-разработок и др.) из практического опыта Индустриальных партнеров в объеме не менее 1 кейса на 1 з.е.

Расчётно-графическая работа. Тема: Проектирование и обучение RL-агента для интеллектуального мониторинга и восстановления работоспособности ИС в системе СФИНКС

1. Цель работы

Разработать математическую модель задачи мониторинга/восстановления работоспособности информационной системы в терминах обучения с подкреплением (RL) и построить демонстрационного агента, который принимает решения о запуске тестов, диагностических сценариев и процедур восстановления, минимизируя время простоя и количество инцидентов.

2. Исходные данные

Система СФИНКС осуществляет:

- сбор параметров доступности, задержек, ошибок;
- выполнение тестовых сценариев (имитация действий пользователя);
- генерацию событий (предупреждение/инцидент/нештатная ситуация);
- запуск автоматизированных процедур восстановления;
- получение метрик от SCOM, PRTG, Orion и др.;
- оповещение по каналам (почта, SMS, ServiceDesk OmniTracker);
- формирование отчётов и графиков доступности.

3. Индивидуальное задание студента

3.1. Построение модели динамической системы

1. Описать динамическую модель состояния ИС в виде **Марковского процесса принятия решений (MDP)**.
В рамках модели определить:
 - пространство состояний S (например: «норма», «предупреждение», «инцидент», «восстановление», «недоступна», «тестирование»);
 - множество действий A (выполнить тест, запросить данные мониторинга, запустить сценарий восстановления, уведомить СПП, ничего не делать);
 - функцию переходов $P(s'|s,a)$ — на основе статистических данных или предположений;
 - функцию вознаграждения $R(s,a)$, учитывающую время простоя, количество ошибок, успешность восстановления.
2. Построить **граф состояний и переходов**, демонстрирующий динамику поведения ИС.

3.2. Реализация агента

1. Реализовать **одного из трёх агентов** (выбор студента):
 - Monte-Carlo агент;
 - TD-агент (TD(0) или TD(λ));
 - Q-learning или SARSA агент.
2. Агента обучать на имитационной модели:
 - передать параметры: вероятность отказа, задержки, процент инцидентов, эффективность тестов и восстановлений;
 - реализовать сценарии мониторинга как действия агента.
3. Провести обучение агента:
 - построить графики Q-функции или функции ценности;
 - оценить сходимость (граф TD-ошибки, доходы G ?, λ -вознаграждение).

3.3. Исследование стратегии агента

1. Реализовать **ϵ -жадную**, Softmax или байесовскую стратегию исследования–эксплуатации.
2. Сравнить стратегии по:
 - скорости нахождения оптимальной политики,
 - стабильности,
 - итоговому среднему доходу.
3. Построить графики:
 - зависимости дохода от эпизода,
 - зависимости числа инцидентов от эпизода,
 - распределения состояний.

3.4. Аппроксимация функции ценности

1. Реализовать **линейную или нейросетевую аппроксимацию** $V(s)$ или $Q(s,a)$.
2. Сравнить аппроксимацию со табличной моделью.
- 3.

1. Оценить влияние аппроксимации на сходимость и устойчивость.

3.5. Валидация RL-модели

1. Провести **A/B-тест**:
 - A — стратегия фиксированных интервалов тестирования;
 - B — стратегия RL-агента.
2. Построить:
 - доверительные интервалы доступности ИС,
 - распределение времени восстановления,
 - частоту ложных срабатываний.
3. При возможности вычислить:
 - модифицированные доверительные интервалы,
 - байесовские факторы,
 - вероятность превосходства RL-агента.

4. Требования к оформлению

- Использовать Python (NumPy, Gymnasium/Gym, PyTorch или аналогично).
- В отчёте должны быть представлены:
- Приложить исходный код.
- Объём отчёта: **25–35 страниц**.

4. Рекомендованные источники учебно-методической литературы из:

Дополнительные ссылки на источники информации (KPM):

MF-3:

1. Мятлев, В. Д., Панченко, Л. А., Ризниченко, Г. Ю., Терехин, А. Т.

Теория вероятностей и математическая статистика. Математические модели

— 2-е изд., испр. и доп. — Москва : Юрайт, 2025. — 321 с. — ISBN 978-5-534-01698-7

2. Гмурман, В. Е. Теория вероятностей и математическая статистика: учебник и практикум для академического бакалавриата — Москва : Юрайт, 2020. — 384 с. — ISBN 978-5-534-01698-7

3. Брюс, П., Брюс, Э. Практическая статистика для специалистов Data Science / П. Брюс, Э. Брюс. — Санкт-Петербург : БХВ-Петербург, 2018. — 304 с. — ISBN 978-5-9775-3974-6

4. Сукар, Л. Э. Вероятностные графовые модели. Принципы и приложения / Л. Э. Сукар. — Москва : ДМК Пресс, 2020. — 338 с. — ISBN 978-5-97060-874-6

5. Манюкова, Н. В., Гателюк, О. В. Проверка статистических гипотез: учебное пособие для вузов [Электронный ресурс] / Н. В. Манюкова, О. В. Гателюк. — Электрон. дан. — Санкт-Петербург : Лань, 2022. — 112 с. — Режим доступа: <https://www.litres.ru/book/natalya-vladislavovn/proverka-statisticheskikh-gipotez-uchebnoe-posobie-dly-67931435/>

6. Леман, Э. Проверка статистических гипотез / Э. Леман. — Москва : Наука, 1979. — 408 с.

7. Дружинин, Н. К. Логика оценки статистических гипотез / Н. К. Дружинин. — Москва : Наука, 1982. — 256 с.

8. Цейтлин, Н. А.

Из опыта аналитического статистика Москва : Солар, 2007. — 904 с. : ил., табл.; 22 см. — ISBN 5-86567-091-3.

MF-4:

1. Юре, Л. Статистика машинного обучения [Электронный ресурс] / Л. Юре, Р. Ананд, Д. У. Джеффри. — Электрон. дан. — Москва : ДМК Пресс, 2017. — 356 с. — Режим доступа: <https://e.lanbook.com/book/28374>

2. Горлач, Б. А. Методы статистического анализа данных в машинном обучении [Электронный ресурс] / Б. А. Горлач. — Электрон. дан. — Санкт-Петербург : Лань, 2016. — 320 с. — Режим доступа: <https://e.lanbook.com/book/4654>

Изучение современных технологий по материалам конференций:

1. Mnih V., Kavukcuoglu K., Silver D. et al. Human-level control through deep reinforcement learning // Nature. – 2015. – Vol. 518, № 7540. – P. 529–533. – DOI: 10.1038/nature14236. [ResearchGate](#)

Краткое описание: Статья представляет первую архитектуру глубокого Q-обучения (DQN), позволившую агенту достигать уровня человека в ряде игр Atari 2600 при использовании только пиксельных наблюдений и вознаграждения. Работа впервые показала, что нейронные сети могут стабильно аппроксимировать функцию ценности в условиях большой размерности состояния. Авторы предложили две ключевые идеи — буфер воспроизведения (Experience Replay) и фиксированную целевую сеть — которые обеспечили устойчивость обучения. Статья считается фундаментальной для современной deep RL-парадигмы и используется как основа в обучении агентов в сложных динамических средах.

1. Mnih V., Kavukcuoglu K., Silver D. et al. Playing Atari with Deep Reinforcement Learning // arXiv preprint arXiv:1312.5602. – 2013. [arXiv+1](#)

Краткое описание: Классическая работа DeepMind, положившая начало современному Deep RL. Авторы впервые применили сверточную нейронную сеть как аппроксиматор Q-функции (DQN), показав способность агента обучаться от «сырых» пикселей без инженерии признаков. Статья демонстрирует основы стабилизации обучения: experience replay и целевые сети.

2. Van Hasselt H., Guez A., Silver D. Deep Reinforcement Learning with Double Q-Learning // Proceedings of the AAAI Conference on Artificial Intelligence. – 2016. – Vol. 30, № 1. – DOI: 10.1609/aaai.v30i1.10295. <https://arxiv.org/abs/1509.06461>

Краткое описание: Статья представляет улучшение классического Q-learning, направленное на устранение смещения, возникающего при использовании одной и той же функции ценности как для выбора, так и для оценки действия. Авторы вводят алгоритм **Double Q-Learning** в контексте глубоких нейронных сетей, показывая, что разделение этапов выбора действия и оценки его ценности существенно уменьшает переоценку Q-функции. В работе проводится широкий набор экспериментов на Atari 2600 из набора Arcade Learning Environment: Double DQN демонстрирует более стабильное обучение, меньшую вариативность между запусками, лучшую сходимость и заметное превосходство над классическим DQN по итоговым метрикам. Статья заложила основу для последующих улучшенных алгоритмов Deep RL и стала одним из ключевых результатов исследовательской группы DeepMind.

3. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal Policy Optimization Algorithms // arXiv preprint arXiv:1707.06347. – 2017. [arXiv+1](#)

Краткое описание: Статья вводит PPO — один из самых устойчивых и популярных RL-алгоритмов. Основан на ограничении шага обновления политики через clipped objective. Широко используется в промышленности в задачах управления, игр, мультиагентных систем.

4. Dabney W., Rowland M., Bellemare M. G., Munos R. Distributional Reinforcement Learning with Quantile Regression // Proceedings of the AAAI Conference on Artificial Intelligence. – 2018. – Vol. 32, № 1. – P. 2892–

1. 2901. [ojs.aaai.org+1](https://ojs.aaai.org)

Краткое описание: В работе вводится новый подход к распределённому обучению с подкреплением (Distributional RL), в котором вместо оценки ожидаемой функции ценности $V(s)$ моделируется полное распределение возможных вознаграждений. Авторы представляют алгоритм QR-DQN, использующий квантильную регрессию для аппроксимации распределения ценности с помощью набора квантилей. Метод позволяет точнее учитывать неопределённость и вариабельность отдачи, улучшает устойчивость обучения и приводит к значимым улучшениям качества по сравнению с классическим DQN и его модификациями. Работа стала фундаментальной для современного развития Distributional RL, включая алгоритмы IQN и FQF.

5) Эффективным средством осуществления самостоятельной работы является электронная информационно-образовательная среда университета, которая обеспечивает доступ к образовательной программе, рабочей программе дисциплины, к электронным библиотечным системам, профессиональным базам данных и информационным справочным системам. А также рекомендованные ОТКРЫТЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ (на платформе LMS Moodle (elearning.volgatech.net), интегрированного с федеральной платформой ГИС ЦОС online.edu.ru открытый образовательный портал (mooped.net) и ресурсы открытых образовательных платформ Яндекс-Практикум, Сберуниверситет.

Рекомендуемые образовательные ресурсы

Курсы по машинному обучению и искусственному интеллекту

1. "Введение в Data Science и машинное обучение" (<https://stepik.org/course/4852/>)
 - Основы анализа данных
 - Классические алгоритмы ML
 - Библиотеки NumPy, pandas, scikit-learn
1. "Нейронные сети и обработка текста" (<https://stepik.org/course/54098/>)
 - Архитектуры нейронных сетей
 - Natural Language Processing
 - Практика на реальных задачах
1. "Введение в глубокое обучение" от HSE (<https://stepik.org/course/135003/>)
 - Теория глубокого обучения
 - PyTorch и TensorFlow
 - Компьютерное зрение и NLP

Специализированные курсы

1. "MLOps: практики и инструменты" (<https://stepik.org/course/181476/>)
 - Версионирование моделей и данных
 - CI/CD для ML-проектов
 - Мониторинг моделей в продакшене

Интерактивные платформы для практики

1. **LeetCode** (<https://leetcode.com/>) — алгоритмические задачи на Python
2. **HackerRank** (<https://www.hackerrank.com/domains/python>) — задачи по программированию и ML
- 3.

1. **Kaggle Competitions** (<https://www.kaggle.com/competitions>) — соревнования по анализу данных

Раздел 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-методическое обеспечение

№№ п/п	Список используемой литературы	Количество экземпляров печатных изданий, имеющихся в библиотеке, или электронный адрес издания (ресурса) в сети Интернет
УЧЕБНЫЕ, УЧЕБНО-МЕТОДИЧЕСКИЕ И НАУЧНЫЕ ИЗДАНИЯ		
1.	Хабаров, С. П. Построение распределенных систем на базе WebSocket [Текст] : учебное пособие для вузов / Хабаров С. П., Шилкина М. Л. 2-е изд., стер. Санкт-Петербург: Лань, 2022. - 216 с. ISBN 978-5-8114-9572-6.	https://e.lanbook.com/book/200510
2.	Борзунов, С. В. Алгебра и геометрия с примерами на Python [Электронный ресурс] / Борзунов С. В., Кургалин С. Д. 3-е изд., стер. Санкт-Петербург: Лань, 2022. - 444 с. ISBN 978-5-8114-9980-9.	https://e.lanbook.com/book/202154
3.	Сузи, Р. А. Язык программирования Python [Электронный ресурс] / Сузи Р. А. 2-е изд. Москва: ИНТУИТ, 2016. - 350 с. ISBN 5-9556-0058-2.	https://e.lanbook.com/book/100546
4.	Романов, П. С. Математические основы теории систем. Практикум [Электронный ресурс] : учебное пособие / Романов П. С., Романова И. П. Санкт-Петербург: Лань, 2022. - 172 с. ISBN 978-5-8114-3645-3.	https://e.lanbook.com/book/206675

6.2. Материально-техническая база и программное обеспечение

№№ п/п	Аудитории для проведения учебных занятий, самостоятельной работы и проведения государственной итоговой аттестации	Перечень основного оборудования	Программное обеспечение
1.	333б (II)	Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач, Mathcad University Classroom Perpetual - 40,

			LABVIEW
2.	333г (III)	Измерительный прибор "BerCut-E" (1), Комплекс ПАИК/77100/КПВ (1), Комплект дополнит.оборудования к ПАИК/7710/КПВ(автогенератор AnCom и автоответчи (1), Компьютер P4-3.0/2*256Mb/HDD 200Gb/128 6600GT/DVD-RW/KM/FDD/MBi945P/UPS (1), Ксерокс Canon FC-860 (1), Междисциплинарная лабораторная платформа в комплекте с аппаратно-программным кон (10), Монитор 19"Samsung 940N (LKSB) TFT (1), Принтер HP Laser Jet 1100 (1), Систем.блок Core2 DUOE6300/1024Mb*2/320Gb/DVD-RW/клав.мышь.ковр. (1), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач, Mathcad University Classroom Perpetual - 40, LABVIEW
3.	431 (III)	Доска маркерная 120x240 см. ауд.431 (2), Кабель VGA 30.5 М KRAMER (П-П) (1), Проектор мультимедийный Hitachi CP-X8150 (1), Экран настенный с электроприводом 400x300см (1), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач, Mathcad University Classroom Perpetual - 40, LABVIEW
4.	433 (III)	Акустический комплект с микшером CRATE PA 8FX (1), Микшерный пульт компактный малошумящий Behringer XENYX 1202FX (1), Подавитель шума ALTAIR NG-440 4-х канальный (1), Экран настенный с электроприводом 400*300 Braun (1), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных

			пользовательских задач, Mathcad University Classroom Perpetual - 40, LABVIEW
5.	538 (III)	Системный блок RAY P360 3 ,клав,мышь оптич, коврик+ монитор 19" ViewSonic VA916 (1), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ- Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач, Mathcad University Classroom Perpetual - 40, LABVIEW

Раздел 7. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ/ ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Критерии оценивания индикаторов достижения компетенций направлены на:

- усвоение теоретического материала (объем знаний, глубина усвоения), предусмотренного рабочей программой;
- умение излагать материал (четкость, грамотность изложения материала, точность и полнота воспроизведения учебного материала);
- умение применять теоретические знания при решении практических заданий.

Шкала оценивания представлена ниже.

Уровень сформированности элементов компетенции	Критерии оценивания	Шкала оценивания
Пороговый уровень	Обучающийся имеет знания основного материала, проявляет умение логично его излагать, но может допускать неточности в изложении материала, недостаточно правильные формулировки, испытывает затруднения в выполнении практических заданий.	удовлетворительно
Продвинутый уровень	Обучающийся твердо знает программный материал, излагает его грамотно и по существу, не допускает существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения	хорошо
Высокий уровень	Обучающийся глубоко и прочно усвоил программный материал, грамотно и логически стройно его излагает, дает исчерпывающие ответы на поставленные вопросы. В ответе тесно увязывается теория с практикой, при этом	отлично

	обучающийся не затрудняется с ответом при видоизменении задания, свободно справляется с задачами, вопросами и другими видами применения знаний, показывает знакомство с монографической литературой, периодическими изданиями, правильно обосновывает принятые решения, свободно владеет разносторонними навыками, приемами выполнения практических работ	
--	---	--

7.1. Текущий контроль успеваемости

Текущий контроль успеваемости обеспечивает оценивание хода освоения дисциплины (модуля) и производится с применением технологии рейтингового контроля в соответствии с технологической картой дисциплины. Порядок составления технологической карты и алгоритм проведения процедуры оценивания видов деятельности обучающихся, направленных на освоение знаний, умений, навыков и/или опыта деятельности, по накопительной системе в баллах устанавливается положением о системе РИТМ в ФГБОУ ВО «ПГТУ»

7.2. Промежуточная аттестация обучающихся

Промежуточная аттестация обучающихся направлена на оценивание результатов обучения по дисциплине (модулю) и проводится с использованием фондов оценочных средств.

Примеры типовых контрольных заданий из базы фонда оценочных средств по образовательной программе.

Компетенция MF-1: Статистическое машинное обучение

Тип 3: Множественный выбор

1 Укажите характеристики, которые отличают параметрические методы машинного обучения от непараметрических.

1. Фиксированное число параметров, не зависящее от размера выборки
2. Явные предположения о функциональной форме распределения данных
3. Сложность модели растёт с увеличением объёма обучающих данных
4. Отсутствие предположений о виде функциональной зависимости

2 Определите методы, относящиеся к дискриминативным моделям в статистическом машинном обучении.

1. Логистическая регрессия
2. Метод опорных векторов (SVM)
3. Наивный байесовский классификатор
4. Скрытые марковские модели (HMM)

3 Выберите факторы, которые необходимо учитывать при выборе модели для задачи с высокой размерностью признакового пространства и малым объёмом выборки.

1. Необходимость регуляризации для предотвращения переобучения
2. Выбор методов с встроенной селекцией признаков
3. Увеличение числа параметров модели для лучшей подгонки
4. Использование максимально сложных нелинейных зависимостей

4 Укажите метрики, которые подходят для оценки качества моделей бинарной классификации в условиях несбалансированных классов.

1. F1-мера
2. ROC-AUC
3. Accuracy
4. Среднеквадратичная ошибка (MSE)

5 Определите подходы к валидации моделей, которые обеспечивают несмещённую оценку обобщающей способности при ограниченном объёме данных.

1. K-fold кросс-валидация
2. Stratified кросс-валидация
3. Валидация на обучающей выборке
4. Однократное случайное разбиение train/test

6 Выберите свойства, характерные для линейных методов классификации и регрессии.

1. Интерпретируемость модели через веса признаков
2. Вычислительная эффективность при обучении
3. Способность автоматически выделять нелинейные паттерны
4. Построение иерархических представлений данных

7 Укажите техники, используемые для борьбы с переобучением в статистических моделях машинного обучения.

1. L1-регуляризация (Lasso)
2. L2-регуляризация (Ridge)
3. Увеличение числа параметров модели
4. Использование всех доступных признаков без селекции

8 Определите характеристики генеративных моделей в контексте статистического машинного обучения.

1. Моделирование совместного распределения $P(X,Y)$
2. Способность генерировать синтетические образцы данных
3. Прямая оптимизация границы принятия решения
4. Фокус только на разделяющей гиперплоскости

9 Выберите факторы, влияющие на выбор между генеративным и дискриминативным подходами к классификации.

1. Объём доступных обучающих данных
2. Необходимость генерации новых образцов или обработки пропусков
3. Требование максимальной вычислительной эффективности
4. Исключительно размерность признакового пространства

10 Укажите свойства нелинейных методов машинного обучения по сравнению с линейными.

1. Способность моделировать сложные нелинейные зависимости
2. Большая гибкость и выразительность модели
3. Простота интерпретации весов
4. Гарантированная глобальная оптимизация

Тип 4: Установление последовательности

11 Расположите этапы жизненного цикла разработки модели статистического машинного обучения:

1. Анализ задачи и исследовательский анализ данных
2. Предобработка данных и инженерия признаков
3. Выбор модели и обучение
4. Оценка качества и интерпретация результатов

12 Установите последовательность шагов при проведении K-fold кросс-валидации:

1. Разделение датасета на K фолдов
2. Обучение модели на K-1 фолдах

3. Оценка модели на k-м фолде
4. Усреднение метрик по итерациям

13 Расположите этапы применения регуляризации:

1. Выбор типа регуляризации
2. Определение диапазона значений λ
3. Подбор оптимального λ через кросс-валидацию
4. Обучение финальной модели и оценка

14 Расположите действия при сравнении двух моделей с использованием статистических критериев:

1. Формулирование гипотез
2. Выбор статистического критерия
3. Сбор результатов на одинаковых фолдах
4. Вычисление статистики и интерпретация

15 Расположите шаги feature engineering:

1. Анализ признаков
2. Создание новых признаков
3. Селекция признаков
4. Валидация улучшения качества

Тип 5: Установление соответствия

16 Установите соответствие между методами и характеристиками:

(Методы: Линейная регрессия, Дерево решений, kNN)

1. Параметрический метод с предположением о линейной зависимости
2. Нелинейный метод с автоматической селекцией признаков
3. Непараметрический метод на основе расстояний
4. Ансамблевый метод
5. Байесовский метод

17 Соответствие типов регуляризации и их эффектов:

(Типы: L1, L2, ElasticNet)

1. Приводит к разреженности модели
2. Равномерно уменьшает параметры
3. Комбинирует L1 и L2 штрафы
4. Увеличивает параметры
5. Применяет адаптивные веса

18 Соответствие метрик и типов задач:

(Группы: Регрессия, Классификация, Кластеризация)

1. MSE, RMSE
2. Accuracy, F1-score, ROC-AUC
3. Silhouette score, Davies-Bouldin index
4. Метрики скорости сходимости
5. Метрики вычислительной сложности

19 Соответствие источников ошибок и способов устранения:

(Проблемы: Переобучение, Недообучение, Смещение выборки)

1. Регуляризация или уменьшение сложности
2. Увеличение сложности или добавление признаков
3. Сбор репрезентативных данных
4. Нормализация признаков
5. Применение ансамблей

20 Соответствие методов валидации и применений:

(K-fold кросс-валидация, Leave-one-out кросс-валидация (LOOCV), Временная валидация (time series split))

1. Баланс эффективности и надёжности для независимых наблюдений среднего объёма
2. Максимальное использование данных при очень малых выборках
3. Данные с временной зависимостью, где важен хронологический порядок
4. Большие данные с возможностью параллельных вычислений

5. Данные с пространственной корреляцией и географической структурой

Компетенция MF-4: Динамические системы и обучение с подкреплением

Тип 3: Множественный выбор

21 Укажите необходимые компоненты для формального описания MDP.

1. Множество состояний S
2. Множество действий A
3. Гиперпараметры алгоритма
4. Архитектура нейросети

22 Определите свойства, характерные для марковского свойства.

1. Будущее зависит только от текущего состояния и действия
2. Состояние содержит всю необходимую информацию
3. Необходимо учитывать всю историю
4. Требуется запоминание всех предыдущих состояний

23 Методы динамического программирования:

1. Value Iteration
2. Policy Iteration
3. Q-learning
4. Actor-Critic

24 Характеристики методов Монте-Карло:

1. Использование полных возвратов
2. Model-free подход
3. Обновление после каждого шага
4. Применимость к бесконечным задачам

25 Компоненты функции $Q(s,a)$:

1. Ожидаемый суммарный дисконтированный доход
2. Немедленное вознаграждение
3. Градиент функции потерь
4. Гиперпараметры обучения

26 Методы временных разностей (TD):

1. SARSA
2. Q-learning
3. Value Iteration
4. PPO

27 Преимущества нейросетевой аппроксимации функций ценности:

1. Обработка высокоразмерных пространств
2. Автоматическое выделение признаков
3. Гарантированная сходимость
4. Простота интерпретации

28 Стратегии баланса exploration–exploitation:

1. ϵ -greedy
2. Softmax (Boltzmann)
3. Жадный выбор max-Q
4. Детерминированная политика

29 Современные алгоритмы глубокого RL:

1. DQN
2. Actor-Critic
3. Policy Iteration
4. SARSA

30 Компоненты определения целей агента через доход:

1. Полное недисконтированное вознаграждение
2. Дисконтированный доход с фактором
3. Градиент функции потерь

Learning rate

Перечень вопросов для проведения промежуточной аттестации

Перечень вопросов для проведения промежуточной аттестации

1. Основы RL и MDP

4. Что такое обучение с подкреплением и чем оно отличается от обучения с учителем и без учителя?
5. Определите марковский процесс и объясните его свойства.
6. Что такое функция ценности $V(s)$ и $Q(s,a)$? В чем их различия?
7. Как определяется ожидаемое суммарное вознаграждение $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$ $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$?
8. Как выбор γ (дисконтирующего коэффициента) влияет на поведение агента?

2. Методы оценки и обновления функций ценности

9. Объясните разницу между TD и Monte-Carlo методами оценки $V(s)$.
10. Как $TD(\lambda)$ объединяет TD и Monte-Carlo подходы?
11. Что такое bootstrapping и зачем оно используется в RL?
12. Как рассчитывается TD-ошибка δ и для чего она нужна?
13. Опишите n-step TD и λ -return.

3. Политики и исследование–эксплуатация

14. Что такое политика (policy) и какие существуют способы её представления?
15. Как работают ϵ -жадная и Softmax-политики?
16. Объясните, что такое exploration–exploitation trade-off и как его контролировать.
17. Какие стратегии исследования применяются в байесовских подходах к RL?
18. Как параметр ϵ или τ влияет на поведение агента?

4. Алгоритмы Q-learning и SARSA

19. В чем ключевое отличие SARSA от Q-learning?
20. Что такое off-policy и on-policy методы? Приведите примеры.
21. Объясните алгоритм Q-learning и формулу обновления $Q(s,a)$.
22. Как влияет α (скорость обучения) на сходимость алгоритмов Q-learning и SARSA?
23. Как управлять стабильностью обучения в табличных методах?

5. Аппроксимация функции ценности

24. Когда требуется аппроксимация функций ценности?

25. Какие методы аппроксимации существуют для дискретных и непрерывных пространств состояний?
26. Как реализуется линейная аппроксимация Q-функции?
27. Как нейросетевые методы (DQN) аппроксимируют функции ценности?
28. Что такое experience replay и target network и зачем они нужны?
6. Actor-Critic и Policy Gradient
29. Что такое Actor-Critic и как он работает?
30. Как вычисляется advantage function $A(s,a)$?
31. Как обновляется Actor и Critic в алгоритме Actor-Critic?
32. Какие преимущества Policy Gradient методов по сравнению с Q-learning?
33. Как уменьшить дисперсию градиентов при Policy Gradient методах?
7. Современные методы глубокого RL
34. В чем отличие DQN, Double DQN и DDPG?
35. Как PPO ограничивает изменения политики при обновлении?
36. Какие методы RL подходят для непрерывных действий и состояний?
37. Что такое entropy regularization и зачем он нужен?
38. Объясните применение λ в $TD(\lambda)$ для глубоких RL-методов.
8. Статистическая оценка и валидация
39. Как проводится A/B тестирование стратегий RL?
40. Как применяются доверительные интервалы для оценки функции ценности?
41. Какие критерии сравнения стратегий RL существуют и как их использовать?
42. Как Bayes-тест применяется к оценке эффективности стратегий RL?
43. Как оценить различия в производительности двух RL-агентов статистически?