ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«07» марта 2024 г.
Протокол №1

Рабочая программа дисциплины (модуля)
«Обучение с подкреплением»

Направление подготовки: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика»*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2024*

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0215B0C000D1B3FEB44A37EE631E0C27D4
Владелец: Ивашкевич Евгений Васильевич
Действителен: с 13.01.2026 до 13.04.2027

Электронный документ

Москва
2024

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Перечень планируемых результатов обучения** 5
3. **Тематический план** 7
4. **Содержание дисциплины (модуля)** 7
5. **Учебно-методическое обеспечение** 9
6. **Материально-техническое обеспечение** 9
7. **Методические и оценочные материалы** 11

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Обучение с подкреплением» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению 02.03.01 «Математика и компьютерные науки», профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Обучение с подкреплением» играет ключевую роль в разработке автономных систем, способных обучаться в сложных и неопределённых средах, таких как робототехника, игры и адаптивные интерфейсы. Полученные знания и навыки позволяют эффективно применять методы обучения с подкреплением в современных задачах искусственного интеллекта, включая управление, оптимизацию и выравнивание генеративных моделей.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 «Математика и компьютерные науки», профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика» и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений, как дисциплина по выбору.

Дисциплина (модуль) изучается на 3 или 4 курсе в 5, 6, 7 семестре на выбор. Доступна к изучению после успешного освоения дисциплины (модуля) «Глубокое обучение».

Цель изучения дисциплины (модуля): сформировать системное понимание теоретических основ и практических методов обучения с подкреплением, а также развить навыки проектирования, реализации и анализа агентов, способных принимать оптимальные решения в динамических средах.

Задачи изучения дисциплины (модуля):

— изучить основные понятия обучения с подкреплением: агент, среда, состояние, наблюдение, действие, награда, эпизод, стратегия и функция ценности, а также формализовать задачи в виде марковского процесса принятия решений;

— освоить уравнения беллмана и принцип оптимальности, лежащие в основе методов динамического программирования и временных разностей;

— научиться различать и применять методы на основе текущей стратегии и другой стратегии, а также реализовывать алгоритмы, основанные на функции ценности и на стратегии, включая метод подкрепления и архитектуру актор-критик;

— изучить архитектуру и методы стабилизации глубоких алгоритмов, такие как сеть с воспроизведением опыта, целевая сеть, двойное q-обучение и сеть с разделённой оценкой, и применять их в средах с высокой размерностью состояний;

— освоить современные подходы к обучению с подкреплением, включая методы с доверительной областью, мягкий актор-критик и их применение в задачах непрерывного управления, а также выравнивание больших языковых моделей по предпочтениям человека.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

— основные понятия обучения с подкреплением (reinforcement learning, rl): агент, среда, состояние, наблюдение, действие, награда, эпизод, стратегия (policy), функция ценности;

— формальную постановку задачи обучения с подкреплением через марковский процесс принятия решений (markov decision process, mdp);

— уравнения беллмана и принцип оптимальности беллмана;

— различия между методами на основе текущей политики (on-policy) и другой политики (off-policy), на основе функции ценности (value-based) и политики (policy-based);

— основные классические алгоритмы обучения с подкреплением: динамическое программирование, монте-карло методы, td-обучение (td-learning), sarsa, q-обучение (q-learning);

— причины нестабильности алгоритмов глубокого обучения с подкреплением (deep rl) и методы стабилизации обучения;

— архитектуру и основные идеи deep q-network (dqn): буфер воспроизведения (replay buffer), целевая сеть (target network), double dqn и dueling dqn;

— подходы на основе градиента политики: reinforce, функция преимущества (advantage function), актор-критик (actor-critic), gae;

— принципы методов с доверительной областью (trust-region): trpo и ppo;

— подходы к непрерывному управлению (continuous control) и алгоритмы типа soft actor-critic (sac);

— основы мультиагентного обучения с подкреплением;

— подходы rlhf и dpo для больших языковых моделей человеческим предпочтениям;

— связь обучения с подкреплением и задач оптимального управления: линейно-квадратичный регулятор, линейно-гауссовский регулятор, фильтр калмана;

— области применения обучения с подкреплением: игры, робототехника, рекомендательные системы, управление ресурсами, финансы, большие языковые модели.

уметь:

— формализовывать задачи в виде марковского процесса принятия решений;

— объяснять связь обучения с подкреплением с оптимальным управлением и классическими методами теории управления;

— реализовывать и запускать базовые алгоритмы;

— обучать и анализировать агентов в симуляционных средах;

— сравнивать различные подходы и выбирать подходящий алгоритм под задачу;

— настраивать базовые гиперпараметры моделей;

— анализировать сходимость и стабильность обучения агентов;

— использовать нейросетевые модели для приближения функций ценности и политик;

— применять актор-критик и ppo-подобные подходы для решения задач управления;

— интерпретировать результаты обучения агента и диагностировать типичные проблемы обучения;

— работать с современными библиотеками и инструментами для глубокого обучения с подкреплением.

владеть:

— ориентация в основных направлениях современного обучения с подкреплением и примерах их практического применения;

— реализация и тестирование классических и глубоких алгоритмов обучения с подкреплением;

— анализ качества обучения агентов и интерпретация поведения стратегий;

— использование нейросетевых моделей для решения задач обучения с подкреплением.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области искусственного интеллекта, основные принципы критической оценки источников информации и их релевантности.
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1.	Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач
		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики	ОПК-1.1.	Знает основные методы и подходы к решению задач прикладной и компьютерной математики, включая алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области

		ОПК-1.2.	Умеет анализировать и формулировать математические задачи, применять соответствующие методы и алгоритмы для их решения, а также интерпретировать и представлять результаты в понятной и доступной форме
		ОПК-1.3.	Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая участие в конкурсах, олимпиадах или научных публикациях, где были решены актуальные и значимые задачи
ПК-1.	Способен формулировать задачи с математической точностью, обосновывать утверждения строго и анализировать полученные результаты в области математики и компьютерных наук	ПК-1.1.	Знает методы и подходы к формулированию задач, а также основные принципы математического доказательства и анализа результатов.
		ПК-1.2.	Умеет корректно ставить и формулировать математические задачи, применять строгие методы доказательства и анализировать полученные результаты.
		ПК-1.3.	Имеет опыт работы с задачами в области математики и компьютерных наук, включая применение математических методов для решения практических задач
ПК-2.	Способен решать типовые задачи профессиональной деятельности в области искусственного интеллекта, опираясь на информационную и библиографическую культуру, используя информационно-коммуникационные технологии и учитывая основные требования информационной безопасности	ПК-2.1.	Знает основы информационной и библиографической культуры, а также принципы информационной безопасности и применения информационно-коммуникационных технологий в профессиональной деятельности
		ПК-2.2.	Умеет эффективно использовать информационно-коммуникационные технологии для решения стандартных задач профессиональной деятельности, учитывая требования информационной безопасности
		ПК-2.3.	Имеет опыт работы с информационными ресурсами и технологиями в области искусственного интеллекта, включая соблюдение норм информационной безопасности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятельная работа	
		Лекции	Семинары			
1	Введение в обучение с подкреплением	4	4		16	Домашнее задание
2	Классическая теория обучения с подкреплением	4	4		16	Домашнее задание
3	Глубокое обучение с подкреплением	4	4		16	Домашнее задание
4	Алгоритмы градиента политики (Policy Gradient) и Актор-Критик (Actor-Critic)	8	8	2	34	Домашнее задание Контрольная работа
5	Алгоритмы непрерывного управления (Continuous Control) и современные применения обучения с подкреплением	10	10	2	40	Домашнее задание Контрольная работа Кейс
	Экзамен			4		
	Итого:	***30***	***30***	***8***	***122***	
Объем дисциплины (модуля) (в ак. ч.)		***190***				
Объем дисциплины (модуля) (в зач. ед.)		***5***				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Введение в обучение с подкреплением	Введение в обучение с подкреплением Формальная постановка обучения с подкреплением
2	Классическая теория обучения с подкреплением	Динамическое программирование Табличное обучение с подкреплением
3	Глубокое обучение с подкреплением	Введение в глубокое обучение с подкреплением. Алгоритм DQN Модификации алгоритма DQN
4	Алгоритмы градиента политики (Policy Gradient) и Актор-Критик (Actor-Critic)	Алгоритмы на основе градиента политики (Policy Gradient) Актор-Критик (Actor-Critic) алгоритмы Алгоритмы на основе доверительной области (trust-region) Алгоритм Proximal Policy Optimization (PPO)
5	Алгоритмы непрерывного управления (Continuous Control) и современные	Алгоритмы непрерывного управления (Continuous Control) Мультиагентное обучение с подкреплением Алгоритмы для дообучения больших языковых моделей Обучение с подкреплением и оптимальное управление

	применения обучения с подкреплением	

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Карау, Х. Изучаем Spark. Молниеносный анализ данных : практическое руководство / Х. Карау, Э. Конвински, П. Венделл, З. Матей ; пер. с англ. - 2-е изд. - Москва : ДМК Пресс, 2023. - 305 с. - ISBN 978-5-89818-320-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2102607.

2. Харенслак, Б. Apache Airflow и конвейеры обработки данных : практическое руководство / Б. Харенслак, Д. дэ Руйтер ; пер. с англ. Д. А. Беликова. - Москва : ДМК Пресс, 2022. - 502 с. - ISBN 978-5-97060-970-5. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2155905.

Дополнительная литература:

1. Лесковец, Ю. Анализ больших наборов данных : практическое руководство / Д. Дж. Ульман, Ю. Лесковец, А. Раджараман ; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2023. - 500 с. - ISBN 978-5-89818-304-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2102592.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование"	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое

Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Обучение с подкреплением» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, тесты, контрольная работа, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удается, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар – это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Аудиторная работа – активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники

информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Кейс – практическая работа студентов над реальными или смоделированными задачами, что позволяет студенту применять теоретические знания на практике.

Студент самостоятельно разрабатывает стратегию решения поставленной задачи, что способствует развитию навыков критического мышления и самостоятельного принятия решений. Такой подход помогает подготовить будущих специалистов к реальным вызовам в их профессиональной деятельности.

Контрольная работа – письменная работа с набором задач, которые нужно решить за ограниченное время.

Цель контрольной работы – получить специальные знания по одной или нескольким темам дисциплины и продемонстрировать навыки их практического применения.

Бонусные баллы — это оценки, которые студенты могут получить за выполнение дополнительных заданий.

Формат бонусных баллов позволяет студентам улучшить общую оценку по дисциплине (модулю) и стимулирует углубленное изучение материала.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Обучение с подкреплением»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме ***экзамена***, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные
9	Отлично	
8	Отлично	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
7	Хорошо	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	
5	Удовлетворительно	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	
3	Не сдан	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	
1	Не сдан	

Дисциплина (модуль) «Обучение с подкреплением» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	50%	14	Набор задач по темам недели
Кейс	20%	1	Практическая работа студентов над реальными или смоделированными задачами
Контрольная работа	15%	2	Письменная работа с набором задач, которые нужно решить за ограниченное время
Экзамен	15%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

В рамках изучения дисциплины (модуля) возможно получение бонусных баллов.

Формула расчёта итоговой оценки по дисциплине (модулю) «Обучение с подкреплением»: «0,5 × среднее за домашние задания + 0,2 × кейс + 0,15 × контрольная работа + 0,15 × экзамен».

Электронный документ

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1.

1. Объясните, в чём заключается суть обучения с подкреплением и как взаимодействуют агент и среда в процессе обучения;
2. Опишите основные компоненты задачи обучения с подкреплением: состояние, действие, награда, стратегия и функция ценности;
3. Сформулируйте задачу обучения с подкреплением в виде марковского процесса принятия решений и объясните, зачем требуется марковское свойство;
4. Объясните, чем отличаются эпизодическое и непрерывное обучение, и приведите примеры задач для каждого типа;
5. Приведите примеры реальных приложений обучения с подкреплением в робототехнике, играх и управлении ресурсами.

Домашнее задание 2.

1. Объясните, как методы динамического программирования используют уравнения Беллмана для поиска оптимальной стратегии;
2. Опишите, как работает итерация по стратегиям и итерация по ценности в табличных методах обучения с подкреплением;
3. Реализуйте простой алгоритм Q-обучения для среды с конечным числом состояний и действий (например, сетка 4x4);
4. Объясните, в чём разница между методами Монте-Карло и временным разностным обучением (TD-learning) при оценке функции ценности;
5. Сравните алгоритмы SARSA и Q-обучение, указав различия в подходах текущей политики (on-policy) и другой политики (off-policy).

Домашнее задание 3.

1. Объясните, почему для сред с высокой размерностью состояний недостаточно табличных методов и требуется аппроксимация функции ценности;
2. Опишите архитектуру алгоритма DQN и роль буфера воспроизведения в стабилизации обучения;
3. Объясните, как целевая сеть (target network) помогает уменьшить корреляцию между обучающими выборками и стабилизирует сходимость;
4. Реализуйте или опишите схему модификации DQN с двойным обучением (Double DQN) и объясните её преимущества;
5. Опишите принцип работы Dueling DQN и как разделение оценки состояния и преимущества действия улучшает обучение.

Примерное описание задания и критерии оценивания к кейсу

Цель задания: применить теоретические знания и практические навыки по обучению с подкреплением для анализа, проектирования и обоснования решения реальной или смоделированной задачи в одной из предметных областей: робототехника, игры, управление ресурсами, рекомендательные системы или выравнивание языковых моделей.

Задание для студентов:

— выбрать или получить кейс, описывающий задачу принятия решений в динамической среде;

— формализовать задачу в терминах обучения с подкреплением: определить состояние, действие, награду, стратегию и тип среды (эпизодическая / непрерывная, дискретная / непрерывная);

— обосновать выбор алгоритма (например, q-обучение, dqn, ppo, sac) с учётом особенностей задачи и доступных данных;

— описать архитектуру решения: структуру нейросети (если используется), способ аппроксимации функции ценности или стратегии, метод стабилизации обучения;

— провести анализ возможных проблем: нестабильность, сходимость, исследование/эксплуатация, смещение политики;

— подготовить презентацию и защитить решение перед комиссией или группой, ответив на вопросы по обоснованности выбора и альтернативным подходам.

Примеры кейсов:

— обучение агента проходить лабиринт с изменяющейся структурой;

— оптимизация управления температурой в умном доме;

— дообучение языковой модели по предпочтениям пользователя (RLHF);

— управление автономным транспортным средством на перекрёстке;

— балансировка нагрузки в облачной инфраструктуре.

Критерии оценивания кейса:

— корректность формализации задачи в терминах обучения с подкреплением: чёткое определение состояния, действия, награды и среды;

— обоснованность выбора алгоритма с учётом типа задачи (дискретное/непрерывное управление, размерность пространства, доступ к среде);

— глубина анализа архитектуры решения: использование современных подходов (буфер воспроизведения, целевые сети, gae, энтропия и др.);

— способность выявить и прокомментировать потенциальные проблемы обучения: сходимость, нестабильность, компромисс между исследованием и эксплуатацией;

— качество презентации и защита: логичность изложения, ответы на вопросы, сравнение с альтернативными методами.

Примерные задания для контрольной работы

1. Объясните, в чём заключается отличие обучения с подкреплением от обучения с учителем и без учителя, и приведите пример задачи, где требуется именно обучение с подкреплением;
2. Опишите, как формализуется задача обучения с подкреплением в виде марковского процесса принятия решений: укажите состояния, действия, награду и стратегию;
3. Объясните, как уравнение Беллмана используется для оценки функции ценности и почему оно лежит в основе большинства алгоритмов;
4. Сравните методы Монте-Карло и временного разностного обучения (TD) по способу обновления Q-функции и сходимости;
5. Опишите, как работает алгоритм Q-обучения и в чём его основное отличие от SARSA;
6. Объясните, зачем в DQN используется буфер воспроизведения и целевая сеть, и как эти механизмы стабилизируют обучение;
7. Опишите, как архитектура актор-критик сочетает оценку ценности и непосредственное моделирование стратегии;
8. Объясните, как функция преимущества (advantage function) и метод GAE уменьшают дисперсию градиента в алгоритмах на основе политики;
9. Опишите, как алгоритм PPO использует обрезку вероятностей для ограничения шага обновления стратегии;
10. Объясните, почему алгоритмы, такие как DQN, не подходят для задач с непрерывным пространством действий, и как Soft Actor-Critic решает эту проблему.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1	Назовите ключевую концепцию MDP в обучении с подкреплением.	Марковское свойство	УК-1
2	Укажите метод поиска научных статей по обучению с подкреплением.	Поисковая система; база данных	УК-1
3	Как называется процесс оценки достоверности источника информации о RL?	Критический анализ	УК-1
4	Назовите подход к объединению данных из разных источников для решения задач в RL.	Синтез	УК-1
5	Назовите нормативный акт, регулирующий использование ИИ в России.	Федеральный закон	УК-2
6	Как называется процесс определения задач в проекте по RL?	Формулировка	УК-2
7	Какой способ решения задач учитывает ограниченность ресурсов в RL?	Оптимизация	УК-2
8	Как называется документ, подтверждающий соответствие решения правовым нормам в RL?	Сертификат	УК-2
9	Назовите метод математического моделирования, применяемый в обучении с подкреплением.	Алгоритм	ОПК-1
10	Как называется алгоритм, используемый в вариационных методах RL?	Вариационный байесовский	ОПК-1
11	Назовите теорию, применяемую для нахождения экстремумов в функциях RL.	Оптимизация	ОПК-1
12	Как называется участие в научных публикациях по RL?	Публикация	ОПК-1
13	Назовите основную теоретическую концепцию машинного обучения в RL.	Обучение с подкреплением	ПК-1
14	Как называется закономерность, описывающая зависимость данных в RL?	Корреляция	ПК-1
15	Назовите метод исследования, основанный на эксперименте в RL.	Эмпирический	ПК-1
16	Как называется практическая работа по RL?	Проект	ПК-1
17	Назовите элемент информационной культуры, связанный с этикой в RL.	Конфиденциальность	ПК-2
18	Как называется принцип защиты данных в RL?	Шифрование	ПК-2
19	Назовите технологию для обмена данными в ИИ и RL.	API	ПК-2
20	Как называется опыт работы с информационными ресурсами в RL?	Анализ данных	ПК-2