

ГОУ ВПО Российско-Армянский (Славянский) университет

Утверждено
Директор Института математики и информатики
Арамян Р.Г.
«21» мая 2025, протокол №9.1



УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ДИСЦИПЛИНЫ

Наименование дисциплины: Глубокое обучение с подкреплением

Автор (ы) Саакян Вардан Рафикович

Ф.И.О, ученое звание (при наличии), ученая степень (при наличии)

01.04.02 Прикладная математика и информатика

Наименование образовательной программы: «Искусственный интеллект и робототехника»

1. АННОТАЦИЯ

1.1. Краткое описание содержания данной дисциплины;

Дисциплина посвящена изучению методов обучения с подкреплением и их развитию в рамках глубокого обучения. Вначале рассматриваются базовые алгоритмы (многоармие бандиты, TD-обучение, Q-learning, методы Монте-Карло), затем изучаются современные методы с использованием нейронных сетей: DQN, PPO, SAC, Actor-Critic и др. Особое внимание уделяется практической реализации алгоритмов, работе с симуляторами и анализу поведения обученных агентов. Курс завершается обзором актуальных направлений и приложений в области DRL.

1.2. Трудоемкость в академических кредитах и часах, формы итогового контроля (экзамен/зачет);

Виды учебной работы	Всего, в акад. часах
Общая трудоемкость изучения курса	144
Аудиторные занятия, в т. ч.:	64
Лекции	32
Практические занятия	32
Самостоятельная работа	53
Итоговый контроль	Зачет

1.3. Взаимосвязь дисциплины с другими дисциплинами учебного плана специальности (направления)

Дисциплина логически и содержательно связана с курсами «Машинное обучение», «Теория вероятностей и математическая статистика», «Выпуклая оптимизация», и «Теория управления». Знания, полученные в этих дисциплинах, необходимы для понимания методов обучения с подкреплением, построения моделей, реализации алгоритмов и анализа результатов. Курс также служит основой для проектной и исследовательской деятельности в области интеллектуальных агентов и автономных систем.

1.4. Результаты освоения программы дисциплины:

Код компетенции	Наименование компетенции	Код индикатора достижения компетенций	Наименование индикатора достижений компетенций
УК-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	1	Разрабатывает командную стратегию для достижения поставленной цели
		2	Умеет организовывать и руководить работой команды
		3	Демонстрирует понимание результатов работы команды и личных действий в ней
ПК-2	способностью разрабатывать и анализировать концептуальные и теоретические модели решаемых научных проблем и задач	1	Применяет современные методы и инструменты для представления результатов научно-исследовательской деятельности
		2	Умеет определять реалистические цели научный проектов в области программирования
		3	Определяет основные направления научный исследований в области программирования и применяемых в них подходов
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе	1	Критически анализирует проблемную ситуацию с

	системного подхода, вырабатывать стратегию действий		целью выработки стратегии действий, аргументировано формулирует собственные суждения и оценки
		2	Использует критический анализ, систематизацию и обобщение информации для решения проблемной ситуации
		3	Владеет навыками критического анализа проблемных ситуаций на основе системного подхода, вырабатывания стратегии действий
УК-2	Способен управлять проектом на всех этапах его жизненного цикла	1	Определяет этапы жизненного цикла проекта и выстраивает последовательность их реализации
		2	Формулирует проблему, на решение которой направлен проект, грамотно определяет цель проекта.
		3	Проектирует решение конкретных задач проекта, выбирая оптимальный способ их решения.
УК-4	Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах),	1	Обладает знаниями особенностей и правил личной и профессиональной устной и письменной

	для академического и профессионального взаимодействия		коммуникации, в том числе на иностранном(ых) языке(ах)
		2	Демонстрирует умение применять современные коммуникативные технологии для академического и профессионального взаимодействия в ситуации устной и письменной коммуникации, в том числе на иностранном(ых) языке(ах)
		3	Имеет навыки академического и профессионального взаимодействия, в том числе на иностранном(ых) языке(ах)

2. УЧЕБНАЯ ПРОГРАММА

2.1. Цели и задачи дисциплины

Цель:

Познакомить студентов с основами и современными методами обучения с подкреплением, включая использование нейронных сетей.

Задачи:

Изучить ключевые алгоритмы RL и DRL (DQN, PPO, SAC и др.);

Развить навыки реализации и настройки RL-агентов

Научить применять методы RL к практическим задачам;

Ознакомить с текущими направлениями в области глубокого RL.

2.2. Трудоемкость дисциплины и виды учебной работы (в академических часах и зачетных единицах) (удалить строки, которые не будут применены в рамках дисциплины)

Виды учебной работы	Всего, в акад. часах	Распределение по семестрам
		<u>1</u> сем
1	2	3
1. Общая трудоемкость изучения дисциплины по семестрам, в т. ч.:	64	
1.1. Аудиторные занятия, в т. ч.:	64	
1.1.1. Лекции	32	
1.1.2. Практические занятия, в т. ч.	32	
1.1.3. Лабораторные работы		
1.1.4. Другие виды (указать)		
1.2. Самостоятельная работа, в т. ч.:		
1.2.1. Подготовка к экзаменам		
1.2.2. Другие виды самостоятельной работы, в т.ч. (подготовка к практикам, работа с учебной литературой, базовая работа с SDR)		
1.3. Другие методы и формы занятий		
Итоговый контроль (Экзамен, Зачет, диф. зачет - указать)	Зачет	

2.3. Содержание дисциплины

2.3.1. Тематический план и трудоемкость аудиторных занятий (модули, разделы дисциплины и виды занятий) по рабочему учебному плану

Разделы и темы дисциплины	Всего (ак. часов)	Лекции (ак. часов)	Практ. Занятия (ак. часов)	Семинары (ак. часов)	Лабор. (ак. часов)
1	2=3+4+5+6 +7	3	4	5	6
Тема 1.1. Введение в обучение с подкреплением	1	1	0		
Тема 1.2. Многоруккие бандиты и исследование-эксплуатация	4	2	2		
Тема 1.3. Методы динамического программирования	4	2	2		
Тема 1.4. Обучение на основе Монте-Карло	5	2	3		
Тема 1.5. Обучение с временными разностями (TD)	4	2	2		
Тема 1.6. Аппроксимация функций ценности	5	2	3		

Тема 2.1. Deep Q-Network (DQN)	5	3	2		
Тема 2.2. Политики на основе градиента	4	2	2		
Тема 2.3. Actor-Critic методы	4	2	2		
Тема 2.4. Proximal Policy Optimization (PPO)	4	2	2		
Тема 2.5. Off-policy методы: DDPG, TD3, SAC	4	2	2		
Тема 2.6. Современные направления и вызовы	4	2	2		
ИТОГО	64	32	32		

2.3.2. Краткое содержание разделов дисциплины в виде тематического плана

Тема 1.1. Введение в обучение с подкреплением

Рассматриваются базовые элементы обучения с подкреплением: агент, среда, политика, награда, функция ценности и взаимодействие во временной перспективе.

Тема 1.2. Многорукие бандиты

Изучаются задачи выбора действий с целью максимизации выигрыша в условиях неопределённости.

Тема 1.3. Динамическое программирование

Рассматриваются методы оптимизации политик в задачах Маркова при известной модели среды.

Тема 1.4. Методы Монте-Карло

Изучаются методы оценки ценности и улучшения политик без знания модели среды.

Тема 1.5. TD-методы, Q-learning и Sarsa

Изучение итеративных методов, объединяющих плюсы динамического программирования и Монте-Карло.

Тема 1.6. Аппроксимация функций ценности

Рассматриваются подходы к обучению с подкреплением с использованием приближений — линейных и нейросетевых.

Модуль 2. Глубокое обучение с подкреплением

Тема 2.1. DQN и улучшения

Анализируется алгоритм Deep Q-Network и его модификации для стабильного обучения.

Тема 2.2. Policy Gradient и REINFORCE

Изучение градиентных методов оптимизации стохастических политик.

Тема 2.3. Actor-Critic и обобщённая оценка преимущества

Рассматриваются методы, сочетающие обучение ценности и политик, включая A2C и GAE.

Тема 2.4. Proximal Policy Optimization (PPO)

Изучение одного из наиболее популярных современных методов, устойчивого к разрушению политик.

Тема 2.5. Off-policy методы: DDPG, TD3, SAC

Изучаются подходы для непрерывных действий с акцентом на стабильность и эффективность.

Тема 2.6. Современные направления

Краткий обзор продвинутых тем: многоагентные системы, мета-обучение, обучение с моделями, симуляции и трансфер.

2.3.3. Краткое содержание семинарских/практических занятий/лабораторного практикума

Практические занятия направлены на закрепление теоретических основ обучения с подкреплением и их реализацию в коде. Студенты реализуют базовые и продвинутые алгоритмы (TD, Q-learning, DQN, PPO, SAC), обучают агентов в средах OpenAI Gym и аналогичных симуляторах, проводят эксперименты с гиперпараметрами и анализируют поведение агентов. Также предусмотрены обсуждения современных подходов и работа над мини-проектами по выбранной теме.

2.3.4. Материально-техническое обеспечение дисциплины

Компьютеры с интернет-браузером.

2.4. Модульная структура дисциплины с распределением весов по формам контролей

Формы контролей	Вес формы (форм) текущего контроля в результирующей оценке текущего контроля (по модулям)		Вес формы промежуточного контроля в итоговой оценке промежуточного контроля		Вес итоговой оценки промежуточного контроля в результирующей оценке промежуточных контролей		Вес итоговой оценки промежуточного контроля в результирующей оценке промежуточных контролей (семестровой оценке)		Веса результирующей оценки промежуточных контролей и оценки итогового контроля в результирующей оценке итогового контроля	
	М1 ¹	М2	М1	М2	М1	М2				
Вид учебной работы/контроля										
Контрольная работа <i>(при наличии)</i>										
Устный опрос <i>(при наличии)</i>										
Тест <i>(при наличии)</i>										
Лабораторные работы <i>(при наличии)</i>										
Письменные домашние задания <i>(при наличии)</i>										
Реферат <i>(при наличии)</i>										
Эссе <i>(при наличии)</i>										
Проект <i>(при наличии)</i>										
Другие формы <i>(при наличии)</i>										
Веса результирующих оценок текущих контролей в итоговых оценках промежуточных контролей					0.4					
Веса оценок промежуточных контролей в итоговых оценках промежуточных контролей					0.6					
Вес итоговой оценки 1-го промежуточного контроля в результирующей оценке промежуточных контролей										

¹ Учебный Модуль

Вес итоговой оценки 2-го промежуточного контроля в результирующей оценке промежуточных контролей								
Вес результирующей оценки промежуточных контролей в результирующей оценке итогового контроля								0.4
Вес итогового контроля (Экзамен/зачет) в результирующей оценке итогового контроля								0.6
	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$

3. Теоретический блок

3.1. Материалы по теоретической части курса

3.1.1. Учебник(и);

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.).

3.1.2. Учебное(ые) пособие(я);

3.1.3. Курс лекций;

Pascal Poupart. Deep Reinforcement Learning.

3.1.4. Краткие конспекты лекций;

3.1.5. Электронные материалы (электронные учебники, учебные пособия, курсы и краткие конспекты лекций, презентации PPT и т.п.);

— Электронные версии учебников и учебных пособий;

3.1.6. Глоссарий/терминологический словарь;

3.1.7. др. варианты материалов, необходимых для освоения учебной программы дисциплины.

4. Фонды оценочных средств

Для оценки освоения дисциплины используются следующие формы:

- **Домашние задания**, включающие реализацию и тестирование алгоритмов обучения с подкреплением, основанных на практических темах курса.
- **Презентации**, в которых студенты представляют одну из современных тем или алгоритмов глубокого RL, объясняя принципы и демонстрируя практические примеры.
- **Итоговый экзамен**, состоящий из теоретических вопросов и заданий на понимание архитектуры и особенностей алгоритмов (в письменной форме).

Эти формы контроля позволяют оценить как теоретическую подготовку, так и способность применять методы обучения с подкреплением на практике.

5. Методический блок

5.1. Методика преподавания

Преподавание дисциплины строится на сочетании лекционного материала с активной практической работой. На лекциях излагаются теоретические основы и современные подходы в обучении с подкреплением, включая глубокие нейросетевые методы. Практические занятия направлены на реализацию алгоритмов, работу с симуляторами, анализ результатов и обсуждение их интерпретации. Студенты поощряются к самостоятельному экспериментированию, чтению оригинальных статей и участию в мини-проектах. Такой подход способствует глубокому усвоению материала и развитию практических навыков в области DRL.