

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 25.05.2026 11:50:52
Уникальный программный ключ:
ca953a0120d891083f939673078ef1a989dae18a

**Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»**

Инженерная академия

(наименование основного учебного подразделения (ОУП) – разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ТЕОРИЯ И ПРАКТИКА ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

**07.04.01 АРХИТЕКТУРА /
27.04.04 УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ**

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

ТЕХНОЛОГИИ ИНТЕЛЛЕКТУАЛЬНОГО АРХИТЕКТУРНОГО ПРОЕКТИРОВАНИЯ

(наименование (профиль/специализация) ОП ВО)

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Теория и практика обучения с подкреплением» входит в программу магистратуры «Технологии интеллектуального архитектурного проектирования» по направлениям 07.04.01 Архитектура / 27.04.04 Управление в технических системах и изучается в 3 семестре 2 курса. Дисциплину реализует Кафедра механики и процессов управления. Дисциплина состоит из 7 разделов и 21 тема и направлена на изучение - базовых принципов обучения с подкреплением;

- основных алгоритмов обучения с подкреплением.
- терминологического аппарата обучения с подкреплением.
- библиотеки глубокого обучения Tensorflow.

Целью освоения дисциплины является познакомиться и овладеть базовыми принципами и основными алгоритмами обучения с подкреплением для применения их в своей профессиональной деятельности.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Теория и практика обучения с подкреплением» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
ОПК-7	Способен осуществлять обоснованный выбор, разрабатывать и реализовывать на практике схмотехнические, системотехнические и аппаратно-программные решения для систем автоматизации и управления	ОПК-7.1 Умеет разрабатывать и реализовывать на практике схмотехнические и системотехнические решения для систем автоматизации и управления;; ОПК-7.2 Умеет разрабатывать аппаратно-программные решения для систем автоматизации и управления;; ОПК-7.3 Владеет подходами для осуществления обоснованного выбора и реализации на практике схмотехнических, системотехнических и аппаратно-программных решений для систем автоматизации и управления.;
ОПК-9	Способен разрабатывать методики и выполнять эксперименты на действующих объектах с обработкой результатов на основе информационных технологий и технических средств	ОПК-9.1 Владеет современными информационными технологиями и техническими средствами для проведения экспериментов на действующих объектах;; ОПК-9.2 Имеет навыки разработки методик и волнения экспериментов на действующих объектах;; ОПК-9.3 Имеет навыки разработки методики и выполнения экспериментов на действующих объектах с обработкой результатов посредством информационных технологий.;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Теория и практика обучения с подкреплением» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Теория и практика обучения с подкреплением».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ОПК-7	Способен осуществлять	Научно-исследовательская	Технологическая практика;

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	обоснованный выбор, разрабатывать и реализовывать на практике схемотехнические, системотехнические и аппаратно-программные решения для систем автоматизации и управления	работа (научно-исследовательская деятельность в области искусственного интеллекта); Технологическая практика; Современные методы машинного обучения;	
ОПК-9	Способен разрабатывать методики и выполнять эксперименты на действующих объектах с обработкой результатов на основе информационных технологий и технических средств	Алгоритмы и структуры данных; Современные методы машинного обучения; Глубокое обучение и генеративные модели; Технологическая практика;	Технологическая практика;

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Теория и практика обучения с подкреплением» составляет «4» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			3
Контактная работа, ак.ч	36		36
Лекции (ЛК)	18		18
Лабораторные работы (ЛР)	18		18
Практические/семинарские занятия (СЗ)	0		0
Самостоятельная работа обучающихся, ак.ч.	81		81
Контроль (экзамен/зачет с оценкой), ак.ч.	27		27
Общая трудоемкость дисциплины ак.ч.	ак.ч.	144	144
	зач.ед.	4	4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы*

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
Раздел 1	Введение в обучение с подкреплением	1.1	Постановка задачи обучения с подкреплением. Отличия обучения с подкреплением от обучения с учителем и обучения без учителя. Области и примеры использования задачи обучения с подкреплением	Обучение с подкреплением как парадигма машинного обучения, где агент обучается через взаимодействие со средой. Отличия от обучения с учителем: отсутствие правильных ответов, обучение через пробу и ошибку. Отличия от обучения без учителя: наличие целевой функции в виде награды. Области применения: робототехника, игры, управление, рекомендательные системы.	ЛК, ЛР
		1.2	Основные компоненты обучения с подкреплением: функция поведения агента, функция ценности действий агента и модель представления окружения агентом	Функция поведения агента как стратегия выбора действий в каждом состоянии. Функция ценности действий агента как оценка ожидаемой суммарной награды. Модель представления окружения агентом как описание динамики среды.	ЛК, ЛР
Раздел 2	Основные понятия обучения с подкреплением	2.1	Понятие рационального агента. Понятие окружения	Рациональный агент как субъект, стремящийся максимизировать суммарную награду. Окружение как внешняя среда, с которой взаимодействует агент.	ЛК, ЛР
		2.2	Понятие действия агента в обучении подкреплением. Понятие награды агента в обучении подкреплением. Понятие истории в обучении подкреплением. Понятие состояния агента в обучении подкреплением	Действие агента как выбор из доступных альтернатив. Награда как числовой сигнал от среды, оценивающий качество действия. История как последовательность наблюдений, действий и наград. Состояние агента как вся доступная информация, необходимая для принятия решения.	ЛК, ЛР
		2.3	Алгоритмы принятия решения и награждения агента в обучении с подкреплением	Алгоритмы принятия решений как правила выбора действий на основе текущего состояния. Награждение как формирование сигнала обратной связи для улучшения поведения.	ЛК, ЛР
Раздел 3	Задача "многоруких бандитов"	3.1	Постановка задачи "многоруких бандитов". Области и примеры использования задачи "многоруких бандитов"	Задача многоруких бандитов как упрощённая модель обучения с подкреплением без учёта перехода между состояниями. Выбор лучшего действия из нескольких при неизвестном распределении наград. Области применения: A/B-тестирование, медицинские испытания, рекомендательные системы.	ЛК, ЛР
		3.2	Обзор методов решения задачи "многоруких бандитов": жадный метод, эпсилон жадный (англ. epsilon-greedy) метод,	Жадный метод с постоянным выбором лучшего известного действия. Эпсилон-жадный метод с балансом между эксплуатацией и исследованием. Метод верхней доверительной границы с учётом неопределённости оценки. Метод gradient-bandit с предпочтениями действий. Методы ценности действия для оценки ожидаемой награды.	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
			метод верхней доверительной границы (англ. upper confidence bound), метод gradient-bandit и др. Методы ценности действия		
Раздел 4	Марковский процесс принятия решений	4.1	Постановка задачи марковского процесса принятия решений	Марковский процесс принятия решений как формализм для описания задач обучения с подкреплением. Марковское свойство: будущее не зависит от прошлого при известном настоящем.	ЛК, ЛР
		4.2	Основные компоненты марковского процесса принятия решений: множество состояний, множество действий, функции моментальных поощрений, функции перехода между состояниями	Множество состояний, в которых может находиться среда. Множество действий, доступных агенту. Функция моментальных поощрений как награда за переход между состояниями. Функция перехода между состояниями как вероятность изменения состояния под действием агента.	ЛК, ЛР
		4.3	Эпизодические и непрерывные задачи. Примеры марковских процессов принятия решений	Эпизодические задачи с естественным завершением взаимодействия. Непрерывные задачи без чёткого окончания. Примеры марковских процессов принятия решений из различных областей.	ЛК, ЛР
Раздел 5	Динамическое программирование	5.1	Постановка задачи динамического программирования. Применение динамического программирования в обучении подкреплением	Динамическое программирование как подход к решению марковских процессов принятия решений при известной модели среды. Применение динамического программирования в обучении с подкреплением для вычисления оптимальной политики.	ЛК, ЛР
		5.2	Применение уравнения Беллмана в обучении подкреплением	Уравнение Беллмана как рекуррентное соотношение, связывающее ценность текущего состояния с ценностью последующих состояний. Использование уравнения для вычисления функций ценности.	ЛК, ЛР
		5.3	Оценка политики рационального агента. Методы Policy Iteration, Policy Evaluation, Value Iteration и Value Evaluation. Проблемы динамического программирования	Метод итерации по политике с последовательным улучшением стратегии. Метод итерации по ценности с прямым вычислением оптимальной функции ценности. Проблемы динамического программирования: необходимость полного знания модели среды, вычислительная сложность.	ЛК, ЛР
Раздел 6	Методы Монте-Карло	6.1	Постановка задачи обучения с подкреплением на основе полного эпизода действий агента. Метод	Метод Монте-Карло как подход, обучающийся на завершённых эпизодах взаимодействия. Усреднение полученных наград для оценки ценности состояний.	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы	Содержание темы	Вид учебной работы*
		Монте-Карло в обучении с подкреплением		
		6.2 Оценка ценности действий с помощью метода Монте-Карло. Контроль Монте-Карло. Преимущества и недостатки метода метода Монте-Карло	Оценка ценности действий путём усреднения наград после выполнения каждого действия. Контроль Монте-Карло как улучшение политики на основе оценок ценности. Преимущества: отсутствие необходимости в модели среды. Недостатки: невозможность обучения в непрерывных задачах, высокая дисперсия оценок.	ЛК, ЛР
		6.3 Примеры задач обучения с подкреплением использующие метод Монте-Карло	Примеры: игры с чёткими эпизодами, управление роботами в эпизодических задачах, оптимизация рекламных кампаний.	ЛК, ЛР
Раздел 7	Нейронные сети и обучение с подкреплением	7.1 Понятие Q-обучения (англ. Q-learning)	Q-обучение как метод обучения без модели среды с использованием Q-функции ценности действий. Офф-политический характер: обучение оптимальной политике независимо от текущей стратегии.	ЛК, ЛР
		7.2 Архитектуры нейронных сетей на основе сверточных нейронных сетей и на основе рекуррентных нейронных сетей	Свёрточные нейронные сети для обработки визуальной информации в обучении с подкреплением. Рекуррентные нейронные сети для задач с частично наблюдаемой средой и необходимостью запоминания истории.	ЛК, ЛР
		7.3 Целевые функции	Целевая функция для обучения нейронных сетей в задачах обучения с подкреплением. Минимизация разницы между текущей и целевой оценками ценности.	ЛК, ЛР
		7.4 Особенности обучения: исследования и эксплуатации, политика принятия решений, параллельное обучение	Баланс между исследованием новых действий и эксплуатацией известных. Политика принятия решений для выбора действий в процессе обучения. Параллельное обучение для ускорения сходимости.	ЛК, ЛР
		7.5 Нейронная сеть Value Iteration Network	Архитектура Value Iteration Network для встраивания итераций ценностного планирования непосредственно в структуру нейронной сети. Применение в задачах, требующих планирования последовательности действий.	ЛК, ЛР

* - заполняется только по ОЧНОЙ форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве ____ шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Ростовцев, В. С. Искусственные нейронные сети : учебник / В. С. Ростовцев. - Санкт-Петербург : Лань, 2019. - 216 с. - ISBN 978-5-8114-3768-9. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/122180> (дата обращения: 06.03.2020). - Режим доступа: для авториз. пользователей

2. Осипов, Г. С. Методы искусственного интеллекта : монография / Г. С. Осипов. - Москва : Физматлит, 2011. - 296 с. - ISBN 978-5-9221-1323-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/544787> (дата обращения: 06.03.2020). - Режим доступа: по подписке

3. Жданов, А. А. Автономный искусственный интеллект : учебное пособие / А. А. Жданов. - 5-е изд. - Москва : Лаборатория знаний, 2020. - 362 с. - ISBN 978-5-00101-655-7. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/135544> (дата обращения: 06.03.2020). - Режим доступа: для авториз. пользователей

4. Вьюгин, В. В. Математические основы машинного обучения и прогнозирования : учебное пособие / В. В. Вьюгин. - Москва : МЦНМО, 2014. - 304 с. - ISBN 978-5-4439-2014-6. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/56397> (дата обращения: 06.03.2020). - Режим доступа: для авториз. пользователей

Дополнительная литература:

1. Алексеева, Т. В. Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - Москва : МФПУ Синергия, 2013. - 384 с. - (Университетская серия). - ISBN 978-5-4257-0092-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/451186> (дата обращения: 06.03.2020). - Режим доступа: по подписке

2. Смолин, Д. В. Введение в искусственный интеллект: конспект лекций : учебное пособие / Д. В. Смолин. - 2-е изд., перераб. - Москва : ФИЗМАТЛИТ, 2007. - 264 с. - ISBN 978-5-9221-0862-1. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/2325> (дата обращения: 06.03.2020). - Режим доступа: для авториз. пользователей

3. Гвоздева, В. А. Введение в специальность программиста : учебник / В.А. Гвоздева. - 2-е изд., испр. и доп. - М. : ИД 'ФОРУМ' : ИНФРА-М, 2019. - 208 с. : ил. - (Профессиональное образование). - ISBN 978-5-8199-0297-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/988422> (дата обращения: 06.03.2020). - Режим доступа: по подписке

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН <https://mega.rudn.ru/MegaPro/Web>
- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>
- ЭБС «Юрайт» <http://www.biblio-online.ru>
- ЭБС «Консультант студента» www.studentlibrary.ru
- ЭБС «Знаниум» <https://znanium.ru/>

2. Базы данных и поисковые системы

- Sage <https://journals.sagepub.com/>
- Springer Nature Link <https://link.springer.com/>
- Wiley Journal Database <https://onlinelibrary.wiley.com/>
- Наукометрическая база данных Lens.org <https://www.lens.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Теория и практика обучения с подкреплением».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИКИ

Доцент

Должность

Салтыкова О.А.

Фамилия И.О.

РАЗРАБОТЧИКИ

Доцент

Должность

Иванюхин А.В.

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП

Заведующий кафедрой

Должность

Разумный Ю.Н.

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО

Заведующий кафедрой

Должность

Разумный Ю.Н.

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО

Старший преподаватель

Должность

Халиль И.

Фамилия И.О.