

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Смирнов Сергей Николаевич
Должность: врио ректора
Дата подписания: 27.06.2025 16:12:57
Уникальный программный ключ:
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Тверской государственный университет»

Утверждаю:
Руководитель ООП:

А.В.Язенин
«25» июня 2023 года

Рабочая программа дисциплины (с аннотацией)
МАШИННОЕ ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ

Направление подготовки

02.04.02 – «Фундаментальная информатика и информационные технологии»

Направленность (профиль)

«Информационные технологии в управлении и принятии решений»

Для студентов 1-го курса

Форма обучения - очная

Составитель:

к.ф.-м.н., Солдатенко И.С

Тверь, 2023

I. Аннотация

1. Цели и задачи дисциплины:

Целью освоения дисциплины является:

овладение концепциями и методологиями машинного обучения с подкреплением и без учителя.

Задачами освоения дисциплины являются:

- получение теоретических знаний в области методов обучения с подкреплением;
- получение теоретических знаний в области современных методов обучения без учителя;
- получение практических навыков решения задач анализа данных с использованием методологии машинного обучения.

2. Место дисциплины в структуре ООП

Дисциплина относится к элективным дисциплинам части, формируемой участниками образовательных отношений Блока 1.

Для освоения дисциплины, обучающиеся должны иметь навыки программирования, работы с основными структурами данных, анализа сложности алгоритмов, а также иметь базовые знания теории вероятности.

Полученные знания в последующем используются при выполнении выпускной квалификационной работы и при прохождении научно-исследовательской практики, а также в дальнейшей трудовой деятельности.

3. Объем дисциплины: 3 зачетных единицы, 108 академических часов, в том числе:

контактная аудиторная работа: лекции 16 часов, в т.ч. практическая подготовка 0 часов; практические занятия 16 часов, в т.ч. практическая подготовка 0 часов;

самостоятельная работа: 76 часов, в том числе контроль 36.

4. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
ПК-2 Способен применять в научно-исследовательской и прикладной деятельности современные языки программирования и методы представления и обработки знаний и данных, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии	ПК-2.1 Знает и применяет современные языки программирования и методы обработки знаний и данных ПК-2.2 Реализовывает численные методы решения задач профессиональной деятельности, пакеты программного обеспечения, операционные системы, сетевые технологии

	ПК-2.3 Применяет существующие методы и алгоритмы цифровой обработки сигналов и изображений, разработки информационных систем
ПК-3 Способен проектировать информационные системы, их компоненты и протоколы их взаимодействия, собирать, обрабатывать и интерпретировать экспериментальные данные и знания, необходимые для проектной и производственно-технологической деятельности; разрабатывать новые алгоритмические, методические и технологические решения в конкретной сфере профессиональной деятельности	ПК-3.1 Знает основы проектирования и элементы архитектурных решений информационных систем ПК-3.2 Применяет в практической деятельности профессиональные стандарты в области информационных технологий ПК-3.3 Составляет техническое задание на разработку информационной системы

5. Форма промежуточной аттестации и семестр прохождения: экзамен, 2 семестр

6. Язык преподавания русский.

II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Учебная программа – наименование разделов и тем	Всего (час.)	Контактная работа (час.)					Самост оатель ная работа, в том числе Контро ль (час.)
		Лекции		Практические занятия		Контроль самостоят ельной работы (в том числе курсовая работа)	
		всего	в т.ч. практическа я подготовка	всего	в т.ч. практическа я подготовка		
Введение	3	1					2
Обратная связь в задачах обучения с подкреплением	10	1		1			8
Задача обучения с подкреплением	12	2		2			8

Динамическое программирование	13	2		3			8
Метод Монте-Карло	12	2		2			8
Метод временных разностей	12	2		2			8
Следы преемственности	10	1		1			8
Обобщение и аппроксимация	12	2		2			8
Планирование в обучении с подкреплением	12	2		2			8
Обучение с подкреплением.	12	1		1			10
ИТОГО	108	16		16			76

III. Образовательные технологии

Учебная программа – наименование разделов и тем (в строгом соответствии с разделом II РПД)	Вид занятия	Образовательные технологии
Введение	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ, проведение вычислительных экспериментов
Обратная связь в задачах обучения с подкреплением	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ, проведение вычислительных экспериментов
Задача обучения с подкреплением	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ, проведение вычислительных экспериментов
Динамическое программирование	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ, проведение вычислительных экспериментов
Метод Монте-Карло	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ, проведение вычислительных экспериментов

Метод временных разностей	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ, проведение вычислительных экспериментов
Следы преемственности	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ, проведение вычислительных экспериментов
Обобщение и аппроксимация	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ, проведение вычислительных экспериментов
Планирование в обучении с подкреплением	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ, проведение вычислительных экспериментов
Обучение с подкреплением. Итоги	Лекции, практические занятия	Презентации, выполнение компьютерных лабораторных работ.

Преподавание учебной дисциплины строится на сочетании практических занятий и различных форм самостоятельной работы студентов. В процессе освоения дисциплины используются следующие образовательные технологии, способы и методы формирования компетенций: семинары, сопровождаемые презентациями; компьютерное тестирование; выполнение индивидуальных заданий в рамках самостоятельной работы.

Дисциплина предусматривает выполнение контрольных работ, домашних заданий на программирование, проведение и интерпретацию результатов вычислительных экспериментов.

Электронные презентации по материалам курса размещаются на сайте поддержки учебного процесса по дисциплине: <http://lms.tversu.ru>.

IV. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

1. Типовые контрольные задания для проверки уровня сформированности компетенций:

ПК-2 Способен применять в научно-исследовательской и прикладной деятельности современные языки программирования и методы представления и обработки знаний и данных, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии

Результат (индикатор)	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания компетенции, шкала оценивания
ПК-2.1 ПК-2.2 ПК-2.3	Тест закрытого типа, примеры вопросов: 1) Вы планируете разработать систему обучения с подкреплением, задачей которой является вывод агента из лабиринта. Какие варианты подкрепления подойдут для этой задачи: ☐ Подкрепление -1 на каждом шаге внутри лабиринта, 0 при выходе. ☐ Нулевое подкрепление на каждом шаге внутри лабиринта, +1 при выходе. ☐ Подкрепление +1 на каждом шаге внутри лабиринта, +10 при выходе. 2) Какие из методов обучения с подкреплением могут вычислить значение функции ценности для одного состояния, не используя её значения для других состояний? ☐ Динамическое программирование ☐ Метод Монте-Карло ☐ Метод временных разностей	Правильный ответ – 2 балла. Для вопросов с множественным выбором правильность выбора каждого пункта оценивается независимо.

ПК-3 Способен проектировать информационные системы, их компоненты и протоколы их взаимодействия, собирать, обрабатывать и интерпретировать экспериментальные данные и знания, необходимые для проектной и производственно-технологической деятельности; разрабатывать новые алгоритмические, методические и технологические решения в конкретной сфере профессиональной деятельности

Результат (индикатор)	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания компетенции, шкала оценивания
ПК-3.1 ПК-3.2	Домашнее задание: Реализуйте программу, которая	Корректно определены подкрепления и

ПК-3.3	методом динамического программирования найдёт оптимальные ценности состояний и стратегию для следующей задачи. Состояние определяется доступной агенту суммой от 0 до 100. Действием игрока является ставка, не превышающая имеющуюся у него сумму. На каждом шаге с вероятностью p сумма ставки добавляется к его деньгам. С вероятностью $1-p$ сумма ставки вычитается. Целью игрока является набрать 100. Если игрок тратит всю сумму, он проигрывает. Изучите зависимость ценностей и стратегии от значения p. Проанализируйте, что происходит если $p < 0.5$ и $p > 0.5$	составлена модель – 4 балла. Корректная реализация метода динамического программирования – 4 балла. Корректно проведен анализ – 4 балла.
	Домашнее задание: Реализуйте программу для сравнения эффективности ϵ-жадных алгоритмов, с алгоритмом, использующим распределение Гиббса на задаче n-руких бандитов. Сравнение проводите на наборе тестов следующего вида: - каждому алгоритму предлагается 2000 случайно сгенерированных задач n-руких бандитов, где $n=10$. - в каждом раунде вознаграждение за выбор действия a ($a \in \overline{1:10}$) определяется случайным числом с нормальным распределением со средним значением $Q^*(a)$ и дисперсией 1. - для генерации каждой из 2000 задач значения $Q^*(a)$ выбираются из нормального распределения со средним значением 0 и дисперсией 1. Постройте графики изменения среднего вознаграждения и процента выбора оптимального решения на	Корректная реализация каждого метода – 8 баллов

	протяжении 1000 раундов игры разными алгоритмами (усреднено по 2000 задач).	
--	---	--

V. Учебно-методическое и информационное обеспечение дисциплины

1) Рекомендуемая литература

а) Основная литература:

1. Лубенцова, Е.В. Системы управления с динамическим выбором структуры, нечеткой логикой и нейросетевыми моделями: монография / Е.В. Лубенцова; Министерство образования и науки Российской Федерации, Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Северо-Кавказский федеральный университет». - Ставрополь: СКФУ, 2014. - 248 с.: ил. - Библиогр. в кн. - ISBN 978-5-88648-902-6; [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=457413>
2. Зыков, Р.И. Системы управления базами данных / Р.И. Зыков. - М.: Лаборатория книги, 2012. - 162 с.: табл., схем. - ISBN 978-5-504-00394-8; То же [Электронный ресурс]. - Режим доступа <http://biblioclub.ru/index.php?page=book&id=142314>

б) Дополнительная литература:

1. Нужнов, Е.В. Компьютерные сети: учебное пособие / Е.В. Нужнов; Министерство образования и науки Российской Федерации, Южный федеральный университет, Инженерно-технологическая академия. - Таганрог: Издательство Южного федерального университета, 2015. - Ч. 2. Технологии локальных и глобальных сетей. - 176 с.: схем., табл., ил. - Библиогр. в кн. - ISBN 978-5-9275-1691-9 ; То же [Электронный ресурс]. - Режим доступа <http://biblioclub.ru/index.php?page=book&id=461991>
2. Аверченков В.И. Эволюционное моделирование и его применение [Электронный ресурс]: монография / В.И. Аверченков, П.В. Казаков. — Электрон. текстовые данные. — Брянск: Брянский государственный технический университет, 2012. — 200 с. — 5-89838-441-X. — Режим доступа: <http://www.iprbookshop.ru/7012.html>
3. Гладков, Л. А. Генетические алгоритмы [Электронный ресурс] / Под ред. В. М. Курейчика. - 2-е изд., исправл. и доп. - М.: ФИЗМАТЛИТ, 2010. - 368 с. - ISBN 978-5-9221-0510-1 Режим доступа <http://znanium.com/go.php?id=544626>

2) Программное обеспечение

Компьютерный класс факультета прикладной математики и кибернетики № 46 (170002, Тверская обл., г.Тверь, Садовый переулок, д.35)	
Adobe Acrobat Reader DC - Russian	бесплатно
Apache Tomcat 8.0.27	бесплатно
Cadence SPB/OrCAD 16.6	Государственный контракт на поставку лицензионных программных продуктов 103 - ГК/09 от 15.06.2009
GlassFish Server Open Source Edition 4.1.1	бесплатно
Google Chrome	бесплатно
Java SE Development Kit 8 Update 45 (64-bit)	бесплатно
JetBrains PyCharm Community Edition 4.5.3	бесплатно
JetBrains PyCharm Edu 3.0	бесплатно
Kaspersky Endpoint Security 10 для Windows	Акт на передачу прав ПК545 от 16.12.2022
Lazarus 1.4.0	бесплатно
Mathcad 15 M010	Акт предоставления прав ИС00000027 от 16.09.2011
MATLAB R2012b	Акт предоставления прав № Us000311 от 25.09.2012
Многофункциональный редактор ONLYOFFICE бесплатное ПО	бесплатно
ОС Linux Ubuntu бесплатное ПО	бесплатно
MiKTeX 2.9	бесплатно
MSXML 4.0 SP2 Parser and SDK	бесплатно
NetBeans IDE 8.0.2	бесплатно
NetBeans IDE 8.2	бесплатно
Notepad++	бесплатно
Oracle VM VirtualBox 5.0.2	бесплатно
Origin 8.1 Sr2	договор №13918/M41 от 24.09.2009 с ЗАО «СофтЛайн Трейд»
Python 3.1 pygame-1.9.1	бесплатно
Python 3.4 numpy-1.9.2	бесплатно
Python 3.4.3	бесплатно
Python 3.5.1 (Anaconda3 2.5.0 64-bit)	бесплатно
WCF RIA Services V1.0 SP2	бесплатно
WinDjView 2.1	бесплатно
R Studio	бесплатно
Anaconda3 2019.07 (Python 3.7.3 64-bit)	бесплатно

3) Современные профессиональные базы данных и информационные справочные системы

1. ЭБС «ZNANIUM.COM» www.znanium.com;

2. ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru/>;
3. ЭБС «Лань» <http://e.lanbook.com>.

- 4) Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины
 - Сайт поддержки учебного процесса по дисциплине: <http://lms.tversu.ru>
 - Научная библиотека ТвГУ (<http://library.tversu.ru>)
 - Сайт ТвГУ (<http://www.tversu.ru>)

VI. Методические указания для обучающихся по освоению дисциплины

1. Текущий контроль успеваемости

Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания компетенции, шкала оценивания
Домашнее задание: Реализуйте программу, которая методом динамического программирования найдёт оптимальные ценности состояний и стратегию для следующей задачи. Состояние определяется доступной агенту суммой от 0 до 100. Действием игрока является ставка, не превышающая имеющуюся у него сумму. На каждом шаге с вероятностью p сумма ставки добавляется к его деньгам. С вероятностью $1-p$ сумма ставки вычитается. Целью игрока является набрать 100. Если игрок тратит всю сумму, он проигрывает. Изучите зависимость ценностей и стратегии от значения p. Проанализируйте, что происходит если $p < 0.5$ и $p > 0.5$	Корректно определены подкрепления и составлена модель – 4 балла. Корректная реализация метода динамического программирования – 4 балла. Корректно проведен анализ – 4 балла.
Домашнее задание: Реализуйте программу для сравнения эффективности ϵ-жадных алгоритмов, с алгоритмом, использующим распределение Гиббса на задаче n-руких бандитов. Сравнение проводите на наборе тестов следующего вида: - каждому алгоритму предлагается 2000 случайно сгенерированных задач n-руких бандитов, где $n=10$. - в каждом раунде вознаграждение за выбор действия a ($a \in 1:10$) определяется случайным числом с нормальным распределением со средним значением $Q^*(a)$ и дисперсией 1.	Корректная реализация каждого метода – 8 баллов

- для генерации каждой из 2000 задач значения $Q^*(a)$ выбираются из нормального распределения со средним значением 0 и дисперсией 1. Постройте графики изменения среднего вознаграждения и процента выбора оптимального решения на протяжении 1000 раундов игры разными алгоритмами (усреднено по 2000 задач).	
Тест закрытого типа, примеры вопросов: 1) Вы планируете разработать систему обучения с подкреплением, задачей которой является вывод агента из лабиринта. Какие варианты подкрепления подойдут для этой задачи: ☐ Подкрепление -1 на каждом шаге внутри лабиринта, 0 при выходе. ☐ Нулевое подкрепление на каждом шаге внутри лабиринта, +1 при выходе. ☐ Подкрепление +1 на каждом шаге внутри лабиринта, +10 при выходе. 2) Какие из методов обучения с подкреплением могут вычислить значение функции ценности для одного состояния, не используя её значения для других состояний? ☐ Динамическое программирование ☐ Метод Монте-Карло ☐ Метод временных разностей	Правильный ответ – 2 балла. Для вопросов с множественным выбором правильность выбора каждого пункта оценивается независимо.

Вопросы для экзамена

1. Постановка задачи обучения с подкреплением в виде марковского процесса принятия решений. Примеры прикладных задач.
2. Уравнения Беллмана для функций ценности. Связь функций ценности между собой. Алгоритмы Policy/Value Iteration.
3. Дилемма exploration vs. exploitation. Подходы к её решению.
4. Понятие on-policy и off-policy алгоритмов обучения с подкреплением. Примеры.
5. Виды RL алгоритмов: model-based, model-free (value-based, policy gradient, эволюционный подходы).
6. Табличные методы: Монте-Карло, Q-learning, Sarsa.
7. Алгоритм DQN и его модификации: Double DQN, приоритизированный буфер, дуэльная архитектура, шумные сети, многошаговый DQN, память.
8. Подход Policy gradient. Алгоритмы Reinforce и A2C.
9. Понятия soft Q-функции и V-функции. Уравнения Беллмана для них. Оптимальная политика для них.

10. Задача многоруких бандитов, UCB-бандиты. Внутренняя мотивация: дистилляция случайной сети (RND) и внутренний модуль любопытства (ICM).

11. Monte Carlo Tree Search в общем виде. Методы AlphaZero и MuZero.

VII. Материально-техническое обеспечение

Для аудиторной работы.

Учебная аудитория № 212 (170002, Тверская обл., г.Тверь, Садовый переулок, д.35)	Набор учебной мебели, мультимедийный комплекс (доска, проектор, панель управления).
Учебная аудитория № 206 (170002, Тверская область, г.Тверь, пер. Садовый, д.35)	Набор учебной мебели, экран, проектор.

Для самостоятельной работы

Помещение для самостоятельной работы обучающихся: Компьютерный класс факультета ПМиК № 46 (170002, Тверская обл., г.Тверь, Садовый переулок, д.35)	Компьютер, экран, проектор, кондиционер.
--	--

VIII. Сведения об обновлении рабочей программы дисциплины

№ п.п.	Обновленный раздел рабочей программы дисциплины	Описание внесенных изменений	Дата и протокол заседания кафедры, утвердившего изменения
1.	V. Учебно-методическое и информационное обеспечение, необходимое для проведения практики 2) Программное обеспечение	Внесены изменения в программное обеспечение	От 24.08.2023 года, протокол № 1 ученого совета факультета
2.	I. Аннотация 3. Объем дисциплины II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий	Изменения в количестве зачетных единиц в часах на самостоятельную работу	От 29.02.2024 года, протокол № 8 ученого совета факультета
3.	I. Аннотация	Изменения в количестве зачетных	От 16.01.2025 года,

	3. Объем дисциплины 5. Форма промежуточной аттестации и семестр прохождения II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий	единиц, часов на контактную и самостоятельную работу, замена формы промежуточной аттестации	протокол № 7 ученого совета факультета
4.			
5.			