

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Смирнов Сергей Николаевич
Должность: врио ректора
Дата подписания: 11.06.2025 10:02:46
Уникальный программный ключ:
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации

ФГБОУ ВО «Тверской государственный университет»

Утверждаю:

Руководитель ООП:

И.С. Солдатенко /

« 30 » 03 2023 года



Рабочая программа дисциплины (с аннотацией)

ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ

Направление подготовки

15.03.06 МЕХАТРОНИКА И РОБОТОТЕХНИКА

Профиль подготовки

Интеллектуальное управление в мехатронных и робототехнических системах

Для студентов 4-го курса

Форма обучения – очная

Составитель:

к.ф.-м.н., доц. И.С. Солдатенко

Тверь, 2023

I. Аннотация

1. Цели и задачи дисциплины:

Целью освоения дисциплины является:

Формирование знаний и навыков в области методов машинного обучения с подкреплением.

Задачами освоения дисциплины являются:

Изучение теоретических основ различных алгоритмов машинного обучения (МО) с подкреплением и получение опыта их практической реализации.

2. Место дисциплины в структуре ООП

Данная дисциплина относится к разделу 5 «Дисциплины профиля подготовки» части, формируемой участниками образовательных отношений Блока 1.

Предварительные знания и навыки:

Основой для освоения дисциплины являются знания, получаемые в рамках дисциплины «Практикум на ЭВМ», «Методы программирования», «Дискретная математика», «Алгоритмы и анализ сложности», «Теория вероятностей и математическая статистика», «Методы оптимизации и исследование операций».

Дальнейшее использование:

Полученные в ходе изучения дисциплины знания используются в научно-исследовательской работе, учебной и производственной практике, при подготовке выпускной квалификационной работы.

3. Объем дисциплины: 4 зачетных единицы, 144 академических часа, в том числе:

контактная аудиторная работа: лекции 30 часа, в т.ч. практическая подготовка 0 часа, лабораторные работы 30 часов, в т.ч. практическая подготовка 0 часа;

контактная внеаудиторная работа: контроль самостоятельной работы ____ 0 ____, в том числе курсовая работа ____ 0 ____;

самостоятельная работа: 84 часа, в том числе контроль 36 часов

4. Планируемые результаты обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
ПК-1 Способен участвовать в качестве исполнителя в научно-исследовательских	ПК-1.1 Разрабатывает математические модели мехатронных и робототехнических

разработках робототехнических мехатронных систем	новых и	систем, их подсистем и отдельных элементов и модулей ПК-1.2 Разрабатывает экспериментальные макеты управляющих, информационных и исполнительных модулей мехатронных и робототехнических систем и проводит их экспериментальное исследование с применением современных информационных технологий ПК-1.3 Анализирует научно-техническую информацию, обобщает отечественный и зарубежный опыт в области средств автоматизации и управления, проводит патентный поиск ПК-1.4 Проводит эксперименты на действующих макетах, образцах мехатронных и робототехнических систем по заданным методикам и обрабатывает результаты с применением современных информационных технологий и технических средств ПК-1.5 Проводит вычислительные эксперименты с использованием стандартных программных пакетов с целью исследования математических моделей мехатронных и робототехнических систем ПК-1.6 Участвует в составлении аналитических обзоров и научно- технических отчетов по результатам выполненной работы, в подготовке публикаций по результатам исследований и разработок
--	------------	--

5. Форма промежуточной аттестации: экзамен, 7 семестр

6. Язык преподавания русский.

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

1. Для студентов очной формы обучения

Учебная программа – наименование разделов и тем	Всего (час.)	Контактная работа (час.)				Контроль самостоят ельной работы (в том числе курсовая работа)	Самост. работа (час.)
		Лекции		Лаборатор ные работы			
		всего	в т.ч. практи ческая подгот овка	всего	в т.ч. практи ческая подгот овка		
Введение	12	2	0	0	0	0	10
Задачи обучения с подкреплением	12	2	0	0	0	0	10
Жадные и ϵ -жадные стратегии	14	2	0	4	0	0	8
Динамическое программирование	18	4	0	4	0	0	10
Метод Монте-Карло	16	4	0	4	0	0	8
Метод временных разностей	16	4	0	4	0	0	8
Следы приемлемости	18	4	0	4	0	0	10
Обобщение и аппроксимация функций	19	4	0	5	0	0	10
Обучение и планирование	19	4	0	5	0	0	10
ИТОГО	144	30	0	30	0	0	84

III. Образовательные технологии

Учебная программа – наименование разделов и тем (в строгом соответствии с разделом II РПД)	Вид занятия	Образовательные технологии
Введение	Лекции, практические занятия	1. Изложение теоретического материала 2. Решение задач
Задачи обучения с подкреплением	Лекции, практические занятия	1. Изложение теоретического материала 2. Решение задач

Жадные и ϵ -жадные стратегии	Лекции, практические занятия	1. Изложение теоретического материала 2. Решение задач
Динамическое программирование	Лекции, практические занятия	1. Изложение теоретического материала 2. Решение задач
Метод Монте-Карло	Лекции, практические занятия	1. Изложение теоретического материала 2. Решение задач
Метод временных разностей	Лекции, практические занятия	1. Изложение теоретического материала 2. Решение задач
Следы приемлемости	Лекции, практические занятия	1. Изложение теоретического материала 2. Решение задач
Обобщение и аппроксимация функций	Лекции	1. Изложение теоретического материала 2. Решение задач

Преподавание учебной дисциплины строится на сочетании лекций, лабораторных работ и различных форм самостоятельной работы студентов. В процессе освоения дисциплины используются следующие образовательные технологии, способы и методы формирования компетенций: традиционные лекции, практические занятия в диалоговом режиме, выполнение индивидуальных заданий в рамках самостоятельной работы.

Дисциплина предусматривает выполнение контрольных работ, письменных домашних заданий.

Учебная программа дисциплины

1. Введение

Задача об n -руком бандите. Метод вычисления значений ценности действий. Выбор действия с помощью операции softmax. Пошаговая реализация обучения. Настационарные задачи. Оптимистичные начальные оценки.

2. Задачи обучения с подкреплением

Взаимосвязь агент-окружающая среда. Цели и вознаграждения. Выгода. Марковские процессы принятия решений. Функции ценности. Оптимальные функции ценности. Оптимальность и аппроксимация.

3. Жадные и ϵ -жадные стратегии

Жадная стратегия. ϵ -жадная стратегия. Сравнение стратегий. Адаптивные стратегии.

4. Динамическое программирование

Оценка стратегии. Улучшение стратегии. Итерация по стратегиям. Итерация по ценностям. Асинхронное динамическое программирование. Эффективность динамического программирования.

5. Метод Монте-Карло

Основные понятия и история. Общая схема метода. Точность вычислений. Оценка стратегии методами Монте-Карло. Оценка ценности действия методом Монте-Карло. Формирование управления методом Монте-Карло. Управление по методу Монте-Карло с интегрированной оценкой ценности стратегий. Оценивание одной стратегии при использовании другой. Управление по методу Монте-Карло с разделенной оценкой ценности стратегий. Пошаговая реализация.

6. Метод временных разностей

Предсказание на основе временных различий. Методы временных разностей TD(0) (Temporal Difference), SARSA (state-action-reward-state-action), Q-обучения. Преимущества методов TD-прогнозирования. Оптимальность метода TD(0).

7. Следы приемлемости

Многошаговое TD-прогнозирование. Прямой подход к методам TD(λ). Обратный подход к методам TD(λ). Эквивалентность прямого и обратного представлений. SARSA(λ), Q(λ). Замещающие следы. Проблемы реализации.

8. Обобщение и аппроксимация функций

Прогнозирование ценности при помощи аппроксимации функции. Методы наискорейшего спуска. Линейные методы. Управление с аппроксимацией функции. Самонастройка с разделенной оценкой ценности стратегий.

9. Обучение и планирование

Модели и планирование. Объединение планирования, исполнения и обучения. Когда модель неверна. Приоритетная прогонка. Сравнение полного и выборочного вариантов дублирования. Траекторная выборка. Эвристический поиск.

Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

1. Типовые задания для лабораторных работ, домашней работы и рейтингового контроля
2. Правила прохождения промежуточной аттестации
3. Примерный список вопросов на экзамен
4. Методические рекомендации по организации самостоятельной работы

IV. Оценочные материалы для проведения текущей и промежуточной аттестации

1. Типовые контрольные задания для проверки уровня сформированности компетенции ПК-1 Способен участвовать в качестве исполнителя в научно-исследовательских разработках новых робототехнических и мехатронных систем

Уровень формирования компетенции, в котором участвует дисциплина	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания компетенции, шкала оценивания
Базовый, владеть	Домашнее задание: Реализуйте программу, для решения задачи алгоритмом $Sarsa(\lambda)$ с линейной интерполяцией: Агенту, управляющему автомобилем, необходимо выбраться из ямы. Движение автомобиля в яме описывается уравнениями: $x_{t+1} = x_t + x'_t$ $x'_{t+1} = x'_t + 0.001a - 0.0025\cos(3x_t)$ $-1.2 \leq x \leq 0.5$ $-0.7 \leq x' \leq 0.07$	Корректно реализованная программа – 20 баллов

	Действие агента a может принимать значения $\{-1, 0, 1\}$ Целью агента является попасть в точку $x = 0.5$.	
Базовый, уметь	Допустим, наблюдались следующие эпизоды: $\{A, 1, B, 1\}, \{A, 1, B, 0\}, \{A, 1, B, 1\}, \{A, 1, B, 0\}$ (буквы – состояния, цифры – подкрепления). Какие значения функций ценности состояний A и B будут определены методом Монте-Карло?	Правильно решённая задача – 5 баллов. Ход решения верный, но имеются ошибки в расчётах – 3 балла.
Базовый, знать	Тест закрытого типа, примеры вопросов: 1) Какие из методов обучения с подкреплением могут быть применены без модели окружающей среды? ☐ Динамическое программирование ☐ Метод Монте-Карло ☐ Метод временных разностей 2) Какие из наборов знаний достаточны для составления оптимальной стратегии? ☐ Переходные вероятности и ожидаемые подкрепления ☐ Переходные вероятности и оптимальная функция ценности состояний ☐ Ожидаемые подкрепления и оптимальная функция ценности состояний	Правильный ответ – 2 балла. Для вопросов с множественным выбором правильность выбора каждого пункта оценивается независимо.

	[] Ожидаемые подкрепления и оптимальная функция ценности действий	
--	--	--

V. Учебно-методическое и информационное обеспечение дисциплины

а) Основная литература:

1. Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. Second edition. MIT Press, 2018. URL: <http://www.incompleteideas.net/book/RLbook2020.pdf>
2. Саттон, Р. С. Обучение с подкреплением: введение : практическое руководство / Р. С. Саттон, Э. Барто ; пер. с англ. А. А. Слинкина. - Москва : ДМК Пресс, 2020. - 552 с. - ISBN 978-5-97060-097-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1210617>

б) Дополнительная литература

1. Неделько В.М. Основы статистических методов машинного обучения [Электронный ресурс]: учебное пособие/ Неделько В.М.– Электрон. текстовые данные.– Новосибирск: Новосибирский государственный технический университет, 2010.– 72 с.– Режим доступа: <http://www.iprbookshop.ru/45418>.– ЭБС «IPRbooks»

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 ИС000000027 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

- <http://lms.tversu.ru>

VI. Методические материалы для обучающихся по освоению дисциплины

1. Типовые задания для практических занятий, домашней работы и рейтингового контроля

Типовые задания для контрольной работы (промежуточный контроль):

- 1) Допустим, наблюдались следующие эпизоды: {A, 1, B, 1}, {A, 1, B, 0} {A, 1, B, 1} {A, 1, B, 0} (буквы – состояния, цифры – подкрепления) . Какие значения функций ценности состояний A и B будут определены методом Монте-Карло?
- 2) Пусть известно, что для некоторой задачи и случайной стратегии с равновероятным выбором действия значения функции ценности для состояний s_1, s_2, s_3 равны, соответственно, -2.4, 3.3, 1.5. Чему равно

значение функции ценности состояния s_4 , если при выполнении доступных в этом состоянии действий a_1, a_2, a_3 агент попадает в состояния s_1, s_2, s_3 соответственно и получает подкрепление $r = 1$ для любого перехода.

Типовые задания для домашней работы:

- 1) Реализуйте программу, для решения задачи алгоритмом $Sarsa(\lambda)$ с линейной интерполяцией:
Агенту, управляющему автомобилем, необходимо выбраться из ямы.
Движение автомобиля в яме описывается уравнениями: $x_{t+1} = x_t + x'_t$,
 $x'_{t+1} = x'_t + 0.001a - 0.0025\cos(3x_t)$, $-1.2 \leq x \leq 0.5$, $-0.7 \leq x' \leq 0.07$.
Действие агента a может принимать значения $\{-1, 0, 1\}$.
Целью агента является попасть в точку $x = 0.5$.
- 2) Разработайте программу для решения следующей задачи методом динамического программирования. Игрок делает ставки (целое число) на результат бросания монетки:
 - a. Если выпал орёл игрок получает столько денег, сколько он поставил
 - b. Если решка – теряет деньги
 - c. Вероятность выпадения орла известна и равна p .Игра заканчивается выигрышем игрока, если он набирает 100 монет, проигрышем – если у него заканчиваются монеты.
- 3) Сравните эффективность ε -жадных алгоритмов, с алгоритмом, использующим распределение Гиббса (мягкий максимум).
Сравнение проводите на наборе тестов следующего вида:
 - каждому алгоритму предлагается 2000 случайно сгенерированных задач n -руких бандитов, где $n=10$.
 - в каждой задаче 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 на 1 шаг и процента выбора оптимального решения на протяжении 1000 раундов игры разными алгоритмами (усреднено по 2000 задачам).

- 4) Реализуйте алгоритм SARSA для задачи поиска выхода из лабиринта. Лабиринт представляется прямоугольной решёткой, каждая клетка которой может быть доступной для агента, препятствием или выходом. Агент имеет возможность перемещаться в 4 стороны, в случае попытки выхода за пределы поля или перемещения на препятствие агент остаётся на месте.

2. Правила прохождения промежуточной аттестации

Для успешной сдачи экзамена студент должен:

- 1) успешно сдать промежуточный контроль, представляющий собой контрольную работу и тест по тематике упражнений, перечисленных выше;
- 2) выполнить домашние задания на программирование систем машинного обучения;
- 3) ответить в ходе экзамена на устные вопросы по тематике учебной программы.

3. Примерный список вопросов на экзамен

- 1) Уравнения оптимальности Беллмана.
- 2) Оценка функции ценности состояний методом динамического программирования
- 3) Оценка функции ценности состояний методом Монте-Карло
- 4) Управление методом Монте-Карло: оценка стратегии, используемой агентом.
- 5) Управление методом Монте-Карло: оценка стратегии, отличающейся от стратегии, управляющей агентом.
- 6) Оценка функции ценности состояний методом временных разностей.
- 7) Управление методом временных разностей: алгоритм SARSA.
- 8) Управление методом временных разностей: оценка стратегии, отличающейся от текущей.
- 9) Оценка функции ценности состояний методом $TD(\lambda)$.
- 10) Управление с использованием алгоритма Sarsa(λ).
- 11) Использование линейной аппроксимации для представления функций ценности.

4. Методические рекомендации по организации самостоятельной работы

Организуя свою учебную работу, студенты должны, во-первых, выявить рекомендуемый режим и характер учебной работы по изучению теоретического курса, практическому применению изученного материала, по выполнению заданий для самостоятельной работы, по использованию информационных технологий и т.д. Во-вторых, ознакомиться с указанным в методическом материале по дисциплине перечнем учебно-методических

изданий, рекомендуемых студентам для подготовки к занятиям и выполнения самостоятельной работы, а также с методическими материалами на бумажных и/или электронных носителях, выпущенных кафедрой своими силами и предоставляемые студентам во время занятий.

Самостоятельная работа студентов, предусмотренная учебным планом, должна соответствовать более глубокому усвоению изучаемого курса, формировать навыки исследовательской работы и ориентировать студентов на умение применять теоретические знания на практике.

1. Работа с учебными пособиями.

Для полноценного усвоения курса студент должен, прежде всего, овладеть основными понятиями этой дисциплины. Необходимо усвоить определения и понятия, уметь приводить их точные формулировки, приводить примеры объектов, удовлетворяющих этому определению. Кроме того, необходимо знать круг фактов, связанных с данным понятием. Требуется также знать связи между понятиями, уметь устанавливать соотношения между классами объектов, описываемых различными понятиями.

2. Самостоятельное изучение тем.

Самостоятельная работа студента является важным видом деятельности, позволяющим хорошо усвоить изучаемый предмет и одним из условий достижения необходимого качества подготовки и профессиональной переподготовки специалистов. Она предполагает самостоятельное изучение студентом рекомендованной учебно-методической литературы, различных справочных материалов, написание рефератов, выступление с докладом, подготовку к лекционным и практическим занятиям, подготовку к зачёту и экзамену.

3. Подготовка к практическим занятиям.

При подготовке к практическим занятиям студентам рекомендуется следовать методическим рекомендациям по работе с учебными пособиями, приведенным выше.

4. Составление конспектов.

В конспекте отражены основные понятия темы. Для наглядности и удобства запоминания используются схемы и таблицы.

5. Подготовка к экзамену.

При подготовке к экзамену студенты должны использовать как самостоятельно подготовленные конспекты, так и материалы, полученные в ходе лекций. Экзамен студенты могут сдавать в виде теста, письменной контрольной работы или устного ответа по вопросам, представленным в данной программе. Для получения положительной оценки на экзамене необходимо продемонстрировать знания, не ниже базового (минимального) уровня.

Процедура оценивания знаний, умений, владений (умений применять) и (или) опыта деятельности обучающихся по дисциплине производится в рамках балльно-рейтинговой системы, включая рубежную и текущую аттестации. Согласно подходам балльно-рейтинговой системы в рамках оценки знаний, умений, владений (умений применять) и (или) опыта деятельности дисциплины установлены следующие аспекты.

Содержание учебной дисциплины в рамках одного семестра делится на два модуля (периода обучения). По окончании модуля (периода обучения) осуществляется рейтинговый контроль успеваемости знаний студентов.

Сроки проведения рейтингового контроля:

осенний семестр – I рейтинговый контроль успеваемости проводится на 8-9 учебной неделе по графику учебного процесса, II рейтинговый контроль успеваемости – две последние недели фактического завершения семестра по графику учебного процесса;

весенний семестр – I рейтинговый контроль успеваемости проводится на 31-32 учебной неделе по графику учебного процесса, II рейтинговый контроль успеваемости – две последние недели фактического завершения семестра по графику учебного процесса.

Максимальное количество баллов, которое может быть получено в результате освоения дисциплины, составляет 100 баллов. Для дисциплин, заканчивающихся экзаменом, 60 баллов отводится на текущий контроль (например, по 30 баллов на каждый модуль) и 40 баллов на промежуточную аттестацию. Для дисциплин, заканчивающихся зачетом, общее количество баллов делится между первым и вторым модулями (например, по 50 баллов на каждый модуль).

Максимальная сумма рейтинговых баллов по учебной дисциплине, заканчивающейся экзаменом, по итогам промежуточной аттестации составляет 40 баллов.

Студент, набравший от 20 до 49 баллов включительно, сдает экзамен. Студенту, набравшему менее 20 баллов, в экзаменационной ведомости ставится оценка «неудовлетворительно». Применяется следующая шкала перевода баллов в оценки: от 50 до 69 – удовлетворительно, от 70 до 84 – хорошо, от 85 и выше – отлично.

VII. Материально-техническое обеспечение

Для аудиторной работы

Учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, курсового	Набор учебной мебели, компьютер, проектор.
---	--

проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, практических занятий, №200 (170002, Тверская область, г. Тверь, пер. Садовый, д. 35)	
Учебная аудитория для проведения лабораторных занятий, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, компьютерный класс №46 (170002, Тверская область, г. Тверь, пер. Садовый, д. 35)	Компьютер, экран, проектор, кондиционер.

Для самостоятельной работы

Помещение для самостоятельной работы, компьютерный класс общего доступа № 46 (170002, Тверская область, г. Тверь, пер. Садовый, 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 дисциплины 2) Программное обеспечение	Внесены изменения в программное обеспечение	От 24.08.2023 года, протокол № 1 ученого совета факультета