Приложение к рабочей программе дисциплины

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Наименование дисциплины	Автономные роботизированные системы в условиях неопределённости внешней среды
Направление подготовки / Специальность	16.04.01 Техническая физика
Направленность (профиль) / Специализация	Робототехника и автономные системы
Форма обучения	очная
Разработчик(и)	Аксёнов С.В., доцент, к.н.

1. Темы дисциплины для самостоятельного освоения обучающимися: отсутствуют.

2. План самостоятельной работы

№ п/п	Учебные встречи	Виды самостоятельной работы	Форма отчетности/ контроля	Количество баллов	Рекомендуемый бюджет времени на выполнение (ак.ч.)*
1	2	3	4	5	6
1.	Автономные роботизированные системы в условиях неопределённости внешней среды	1. Написание реферата 2. Проработка вопросов	1. Реферата 2. Ответы на Вопросы	0-10	26
2.	Ключевые аспекты в проектировании АРС. Взаимосвязь характеристик среды с общим кругом задач, типами и технологиями управления АРС				
3.	Автономное управление наземными мобильными роботами Outdoor-типа в условиях априори неизвестной физически неоднородной среды				
4.	Групповое управление автономными робототехническими системами с учётом различных подмножеств характеристик свойств среды				

3. Требования и рекомендации по выполнению самостоятельных работ обучающихся, критерии оценивания

Самостоятельная работа охватывает темы, изучаемые в течение дисциплины (модуля).

Вид: Написание реферата
Краткая характеристика: написания реферата - подразумевает самостоятельная работа над рефератом на выбранную тему
Критерии оценивания:
- полное раскрытие выбранной темы по дисциплине (модулю), оценивается максимальным количеством баллов;
- отсутствие / неполный раскрытие темы по дисциплине (модулю) оценивается в зависимости

от их количества и рассчитывается в процентах от максимального балла.

Вид: Проработка вопросов.
Краткая характеристика: письменные ответы на заданные вопросы
Критерии оценивания:
- наличие полных законспектированных ответов на вопросы по дисциплине (модулю), оценивается максимальным количеством баллов;
- отсутствие / неполный наличие законспектированных ответов по дисциплине (модулю) оценивается в зависимости от их количества и рассчитывается в процентах от максимального балла.

Темы рефератов

1. Безмодельное непрямое обучение с подкреплением с помощью метода Монте-Карло
2. Безмодельное непрямое обучение с подкреплением с помощью метода временной разницы
3. Безмодельное непрямое обучение с подкреплением с помощью динамического программирования
4. Непрямое обучение с подкреплением с аппроксимацией функций
5. Прямое обучение с подкреплением с помощью градиента политик
6. Аппроксимированное динамическое программирование
7. Ограничения среды и вопросы безопасности
8. Надёжное обучение с ограниченной неопределенностью
9. Частично обозреваемые марковские процессы принятия решений
10. Мета обучение с подкреплением
11. Стохастические мультиагентные игры
12. Полностью кооперативное обучение с подкреплением
13. Полностью соревновательное обучение с подкреплением
14. Мультиагентное обучение с гибридными наградами
15. Обратное обучение с подкреплением
16. Офлайн обучение с подкреплением
17. Сравнение платформ симуляции обучения с подкреплением

4. Рекомендации по самоподготовке к промежуточной аттестации по дисциплине

Оценка результатов самостоятельной работы организуется как самоконтроль.
При выполнении самостоятельной работы рекомендуется использовать:
- комплект учебно-методической документации по дисциплине, основную и дополнительную литературу,
- интернет-ресурсы:
https://grebennikon.ru/ Электронная библиотека Grebennikon
https://eduvideo.online/ Видеотека «Решение»
https://icdlib.nspu.ru/ Межвузовская электронная библиотека (МЭБ)
https://rusneb.ru/ Национальная электронная библиотека

Материалы

1. Саттон Р. С., Барто Э. Дж. Обучение с подкреплением: Введение. 2-е изд. / пер. с англ. А. А. Слинки-
2. на. – М.: ДМК Пресс, 2020. – 552 с.: ил. ISBN 978-5-97060-097-9

3. Уиндер Ф. Обучение с подкреплением для реальных задач: Пер. с англ. — СПб.: БХВ-Петербург, 2023. — 400 с.: ил. ISBN 978-5-9775-6885-2
4. Моралес Мигель Грокаем глубокое обучение с подкреплением. — СПб.: Питер, 2023. — 464 с.: ил. — (Серия «Библиотека программиста»). ISBN 978-5-4461-3944-6
5. Грессер Лаура, Кенг Ван Лун Глубокое обучение с подкреплением: теория и практика на языке Python. — СПб.: Питер, 2022. — 416 с.: ил. — (Серия «Библиотека программиста»). ISBN 978-5-4461-1699-7
6. Michael Hu The Art of Reinforcement Learning Fundamentals, Mathematics, and Implementations with Python Apress, New York, 2023. ISBN-13 (pbk): 978-1-4842-9605-9
7. Uwe Lorenz Reinforcement Learning From Scratch Springer, Cham, 2022. ISBN 978-3-031-09029-5
8. Christos Dimitrakakis, Ronald Ortner Decision Making Under Uncertainty and Reinforcement Learning Theory and Algorithms Springer, Cham, 2022. ISBN 978-3-031-07612-1
9. Shengbo Eben Li Reinforcement Learning for Sequential Decision and Optimal Control Springer, Singapore, 2023. ISBN 978-981-19-7783-1

Вопросы для самоподготовки

1. Чем обучение с подкреплением отличается от других парадигм машинного обучения?
2. Что называется средой?
3. В чем разница между детерминированной и стохастической политикой?
4. Что такое эпизод?
5. Зачем нам нужен коэффициент дисконтирования?
6. Чем функция ценности отличается от Q функции?
7. В чем разница между детерминированной и стохастической средой?
8. Дайте определение уравнению Беллмана.
9. В чем разница между уравнениями Беллмана и оптимальности Беллмана?
10. Как мы выводим функцию ценности из Q функции?
11. Как мы выводим Q функцию из функции ценности?
12. Какие шаги включены в итерацию ценности?
13. Какие шаги включены в итерацию политики?
14. Чем итерация политики отличается от итерации ценности?
15. Что такое метод Монте-Карло?
16. Почему метод Монте-Карло предпочтительнее динамического программирования?
17. Чем задачи прогнозирования отличаются от задач управления?
18. Как метод прогнозирования Монте-Карло предсказывает функцию значения?
19. В чем разница между Монте-Карло первого посещения и Монте-Карло каждого посещения?
20. Почему мы используем инкрементальные обновления среднего в методе прогнозирования Монте-Карло?
21. Чем контроль в соответствии с политикой отличается от контроля вне политики?
22. Что такое политика жадности эпсилон?
23. Чем отличается обучение методом временной разницы от метода Монте-Карло?
24. В чем преимущество использования методом временной разницы?

25. Что такое ошибка метода временной разницы TD?
26. Что такое правило обновления обучения методом временной разницы TD?
27. Как работает метод прогнозирования TD?
28. Что такое SARSA?
29. Чем Q-обучение отличается от SARSA?
30. Что такое проблема многоруких бандитов?
31. Как политика epsilon-greedy выбирает руку бандита?
32. Каково значение температуры в функции softmax?
33. Как мы вычисляем верхнюю границу доверия?
34. Что происходит, когда значение альфа выше значения бета в бета-распределении?
35. Какие этапы включает в себя выборка Томпсона?
36. Что такое контекстные бандиты?
37. Зачем нам нужен DQN?
38. Что такое буфер воспроизведения?
39. Зачем нам нужна целевая сеть?
40. Чем двойной DQN отличается от DQN?
41. Почему нам нужно расставлять приоритеты полученному опыту?
42. Что такое функция преимущества?
43. Зачем нам нужны слои LSTM в DRQN?
44. Что такое метод, основанный на ценности?
45. Зачем нам нужен метод, основанный на политике?
46. Как работает метод градиента политики?
47. Как мы вычисляем градиент в методе градиента политики?
48. Что такое выгода?
49. Что такое градиент политики с направляющей?
50. Определите направляющую функцию
51. Что такое метод актор-критик?
52. Какова роль сетей актора и критика?
53. Чем метод актор-критик отличается от градиента политики с направляющей?
54. Что такое уравнение обновления градиента сети актор?
55. Как работает A2C?
56. Что означает асинхронность в A3C?
57. Чем A2C отличается от A3C?
58. Какова роль сетей акторов и критиков в DDPG?
59. Как работает критик в DDPG?
60. Каковы основные особенности TD3?
61. Зачем нам нужно обучение с усеченным двойным Q?
62. Что такое сглаживание целевой политики?
63. Что такое обучение с подкреплением с максимальной энтропией?
64. Какова роль сети критиков в SAC?
65. Что такое доверительная область?
66. Почему TRPO полезен?
67. Чем метод сопряженных градиентов отличается от градиентного спуска?
68. Каково правило обновления TRPO?
69. Чем PPO отличается от TRPO?

70. Объясните метод PPO-clipped.
71. Что такое факторизация Кронекера?
72. Что такое распределенное обучение с подкреплением?
73. Что такое категориальный DQN?
74. Почему категориальный DQN называется алгоритмом C51?
75. Что такое квантильная функция?
76. Чем QR-DQN отличается от категориального DQN?
77. Чем D4PG отличается от DDPG?