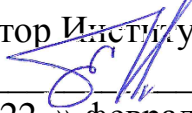


Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Макаренко Елена Николаевна
Должность: Ректор
Дата подписания: 29.07.2022 17:56:13
Уникальный программный ключ:
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ
Директор Института магистратуры

Иванова Е.А.
«_22_» февраля_2022__г.

**Рабочая программа дисциплины
Обучение с подкрепление и его приложения**

Направление 01.04.02 Прикладная математика и информатика
магистерская программа 01.04.02.04 "Искусственный интеллект: математические модели и прикладные решения"

Для набора 2022 года

Квалификация
Магистр

Составитель(и) программы:

Рутга Н.А., к.э.н., доц, и.о заведующий кафедрой фундаментальной и прикладной математики

Программа одобрена на заседании кафедры высшей фундаментальной и прикладной математики
«22» февраля 2021 г., протокол №6

I. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цели освоения дисциплины: выработка у студентов компетенций, связанных с применением алгоритмов обучения с подкреплением при решении практических задач.

Задачи:

- изучение основных классов задач, при решении которых используются алгоритмы обучения с подкреплением;
- освоение навыка оценки эффективности алгоритмов и их реализаций при решении задач, связанных с обучением с подкреплением;
- получение практических навыков программирования при решении задач;
- освоение навыка переформулирования задачи с языка предметной области на математический язык.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

2.1. Учебная дисциплина «Обучение с подкреплением и его приложения» (2 курс магистратуры, 3 семестр) относится к части блока дисциплин (модулей), формируемой участниками образовательных отношений, и является дисциплиной по выбору.

2.2. Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые, изучаемыми в курсе бакалавриата «Математический анализ», «Теория вероятностей», «Основы программирования».

2.3. Знания и навыки, полученные в ходе изучения данной дисциплины (модуля), могут использоваться для решения профессиональных задач в научно-исследовательской, научно-производственной и проектной деятельности, в частности, при выполнении выпускной квалификационной работы (магистерской диссертации).

III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс изучения дисциплины направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОП ВО по данному направлению подготовки (специальности):

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

Шифр и формулировка компетенций (результаты освоения ОП)	Индикаторы компетенций	Элементы компетенций, формируемые дисциплиной
<i>Общекультурные и профессиональные компетенции (ОК и ПК)</i>		
ПК-5. Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов.	ПК-5.1. Руководит работами по оценке и выбору моделей искусственных нейронных сетей и инструментальных средств для решения поставленной задачи.	ПК-5.1. 3-1. Знает функциональность современных инструментальных средств и систем программирования в области создания моделей искусственных нейронных сетей. ПК-5.1. У-1. Умеет проводить оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения задач машинного обучения. ПК-5.1. У-2. Умеет применять современные инструментальные средства и системы программирования для разработки и обучения моделей искусственных нейронных сетей.

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 5 зачетных единиц, 180 часов, из них 34 часа лекционных занятий, 34 часа лабораторных занятий, 76 часов на самостоятельную работу в течение семестра, 36 часов на подготовку к экзамену.

Форма отчетности: экзамен

4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов

№ п/п	Раздел дисциплины/темы	Семестр	Виды учебной работы, включая самостоятельную работу обучающихся и трудоемкость (в часах)				Формы текущего контроля успеваемости Форма промежуточной аттестации (по семестрам)
			Контактная работа преподавателя с обучающимися			Самостоятельная работа	
			Лекции	Семинарские (практические занятия)	Лабораторные занятия		
1	Раздел 1. Мета эвристики.	3	8		8	18	Выполнение лабораторных работ. Ответ на экзамене
1.1	Модель взаимодействия агента со средой.	3	2		2	4	
1.2	Алгоритмы обучения с подкреплением.	3	2		2	4	
1.3	Бэйзлайны.	3	2		2	5	
1.4	Эволюционные стратегии.	3	2		2	5	
2	Раздел 2. Классическая теория.	3	10		10	22	Выполнение лабораторных работ. Ответ на экзамене
2.1	Оценочные функции.	3	2		2	4	
2.2	Relative Performance Identity.	3	2		2	4	

№ п/п	Раздел дисциплины/темы	Семестр	Виды учебной работы, включая самостоятельную работу обучающихся и трудоемкость (в часах)				Формы текущего контроля успеваемости Форма промежуточной аттестации (по семестрам)
			Контактная работа преподавателя с обучающимися		Самостоятельная работа		
2.3	Динамическое программирование.	3	2		2	4	
2.4	Табличные алгоритмы.	3	2		2	5	
2.5	Bias-Variance Trade-Off.	3	2		2	5	
3	Раздел 3. Value-based подход.	3	8		8	18	Выполнение лабораторных работ. Ответ на экзамене
3.1	Deep Q-learning.	3	2		2	6	
3.2	Модификации DQN.	3	3		3	6	
3.3	Distributional RL.	3	3		3	6	
4	Раздел 4. Policy Gradient подход.	3	8		8	18	Выполнение лабораторных работ. Ответ на экзамене
4.1	Policy Gradient Theorem.	3	2		2	6	
4.2	Схемы «Актёр-критик».	3	3		3	6	
4.3	Policy Gradient с переиспользованием сэмплов.	3	3		3	6	
Итого часов			34	0	34	76	
Подготовка к экзамену		3				36	
Всего часов - 180			34	0	34	76	

4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной
		Вид самостоятельной работы	Сроки выполнения	Затраты времени (час.)		

3	Раздел 1. Мета эвристики.	Изучение лекций и написание программ.	4 недели	19	Проверка выполненных заданий. Ответ на экзамене	работы
3	Раздел 2. Классическая теория.	Изучение лекций и написание программ.	4 недель	19	Проверка выполненных заданий. Ответ на экзамене	
3	Раздел 3. Value-based подход.	Изучение лекций и написание программ.	4 недели	19	Проверка выполненных заданий. Ответ на экзамене	
3	Раздел 4. Policy Gradient подход.	Изучение лекций и написание программ.	5 недели	19	Проверка выполненных заданий. Ответ на экзамене	
Общая трудоемкость самостоятельной работы по дисциплине (час)				76		
Бюджет времени самостоятельной работы, предусмотренный учебным планом для данной дисциплины (час)				76		

4.3 Содержание учебного материала

Раздел 1. Мета эвристики

Тема 1.1. Модель взаимодействия агента со средой

Связь с оптимальным уравнением. Марковская цепь. Среда. Действия. Траектории. Марковский процесс принятия решений (MDP). Эпизодичность. Дисконтирование.

Тема 1.2. Алгоритмы обучения с подкреплением

Условия задачи RL. On-policy. Концепция model-free алгоритмов. Классификация RL-алгоритмов. Критерии оценки RL-алгоритмов. Сложности задачи RL. Бенчмарки.

Тема 1.3. Бэйзлайны

Задача безградиентной оптимизации. Случайный поиск. Hill Climbing. Имитация отжига. Эволюционные алгоритмы. Weight Agnostic Neural Networks (WANN). Видовая спецификация. Генетические алгоритмы.

Тема 1.4. Эволюционные стратегии

Идея эволюционных стратегий. Оценка вероятности редкого события. Метод Кросс-Энтропии для стохастической оптимизации. Метод Кросс-Энтропии для обучения с подкреплением (SEM). Натуральные эволюционные стратегии (NES). OpenAI-ES. Адаптация матрицы ковариации (CMA-ES).

Раздел 2. Классическая теория

Тема 2.1. Оценочные функции

Свойства траекторий. V-функция. Уравнение Беллмана. Оптимальная стратегия. Q-функция. Принцип оптимальности Беллмана. Отказ от однородности. Вид оптимальной стратегии (доказательство через отказ от однородности). Уравнение оптимальности Беллмана. Критерий оптимальности Беллмана.

Тема 2.2. Relative Performance Identity

Advantage-функция. Relative Performance Identity (RPI). Policy Improvement.

Тема 2.3. Динамическое программирование

Метод простой итерации. Policy Evaluation. Value Iteration. Policy Iteration. Generalized Policy Iteration.

Тема 2.4. Табличные алгоритмы

Монте-Карло алгоритм. Экспоненциальное сглаживание. Стохастическая аппроксимация. Temporal Difference. Exploration-exploitation дилемма. Q-learning. SARSA.

Тема 2.5. Bias-Variance Trade-Off

Дилемма смещения-разброса. N-step Temporal Difference. Backward View. Eligibility Trace. TD(lambda).

Раздел 3. Value-based подход

Тема 3.1. Deep Q-learning

Q-сетка. Переход к параметрической Q-функции. Декорреляция сэмплов. Таргет-сеть. DQN.

Тема 3.2. Модификации DQN

Twin DQN. Clipped Twin DQN. Double DQN. Dueling DQN. Шумные сети (Noisy Nets). Приоритизированный реплей (Prioritized DQN). Multi-step DQN.

Тема 3.3. Distributional RL

Идея Distributional подхода. Z-функция. Distributional-форма. Distributional Policy Evaluation. Distributional Value Iteration. Категориальная аппроксимация Z-функций. Categorical DQN. Квантильная аппроксимация Z-функций. Quantile Regression DQN. Implicit Quantile Networks. Rainbow DQN.

Раздел 4. Policy Gradient подход

Тема 4.1. Policy Gradient Theorem

Вывод первым способом. Вывод вторым способом. Физический смысл. REINFORCE. State visitation frequency. Расцепление внешней и внутренней стохастики. Связь с policy improvement.

Тема 4.2. Схемы «Актёр-критик»

Бейзлайн. Введение критика. Bias-variance trade-off. Generalized Advantage Estimation (GAE). Обучение критика. Advantage Actor-Critic (A2C).

Тема 4.3. Policy Gradient с переиспользованием сэмплов

Проблема переиспользования сэмплов. Суррогатная функция. Нижняя оценка. Оптимизация нижней оценки. Trust Region Policy Optimization (TRPO). Proximal Policy Loss. Clipped Value Loss. Proximal Policy Optimization (PRO).

4.4 Вопросы, выносимые на самостоятельное изучение

Самостоятельная реализация изучаемых алгоритмов на языке Python.

V. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

При проведении лекций и лабораторных занятий используются следующие образовательные технологии:

- классические лекции
- мультимедийные лекции
- семинары-презентации
- семинары-дискуссии
- онлайн-поддержка самостоятельной работы студентов

Учебный процесс базируется на концепции компетентностного обучения, ориентированного на формирование конкретного перечня профессиональных компетенций, актуализацию получаемых теоретических знаний. Развертывание компетентностной модели обучения предполагает широкое применение инновационных способов организации учебного процесса, в т.ч. применение метода проектного обучения, технологий управляемого самостоятельного обучения в том числе балльно-рейтинговой системы, а также внедрение системы онлайн-поддержки внеаудиторной работы студентов.

VI. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Полный комплект контрольно-оценочных материалов (Фонд оценочных средств) оформляется в виде приложения к рабочей программе дисциплины.

VII. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

7.1. Основная литература.

{Печатный ресурс} Кадуринов А., Николенко С. И., Архангельская Е. В. Глубокое обучение. Погружение в мир нейронных сетей //СПб.: Питер.-2018.-480 с. – 2018.

{Печатный ресурс} Гудфеллоу Я., Иошуа Б., Курвилль А. Глубокое обучение. – Litres, 2018.

{Электронный ресурс: https://github.com/FortsAndMills/RL-Theory-book/blob/main/RL_Theory_Book.pdf] Конспект по обучению с подкреплением, 2021. - 219 с.

7.2. Дополнительная литература.

{Электронный ресурс: https://www.youtube.com/playlist?list=PL_iWQOsE6TfURIIhCrlt-wj9ByIVpbfGc] Курс Сергея Левина.

{Электронный ресурс: https://github.com/yandexdataschool/Practical_RL] Курс Practical RL.

{Электронный ресурс: https://deeppavlov.ai/rl_course_2020] Цикл докладов Advanced RL.

{Электронный ресурс: <https://www.davidsilver.uk/teaching/>] Курс Дэвида Сильвера.

{Электронный ресурс: https://www.youtube.com/watch?v=Isk80iLhdfU&list=PLqYmG7hTraZBKeNJ-JE_eyJH7XgBoAyb&index=1] Курс DeepMind.

Электронный ресурс: <https://www.udacity.com/course/reinforcement-learning--ud600>] Курс GeorgiaTech.

7.3. Список авторских методических разработок.

7.4. Периодические издания (*при необходимости*)

7.5. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

7.6. Программное обеспечение информационно-коммуникационных технологий

VIII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

8.1. Учебно-лабораторное оборудование

При проведении дисциплины учащиеся должны быть обеспечены:

1. Лекционной аудиторией.
2. Аудиторией для лабораторных занятий с аппаратными и программными средствами в соответствии с реализуемой учебной тематикой.

8.2. Программные средства

Операционная система Microsoft Windows (любая версия).

Любая среда для разработки программ на Python (например, Anaconda)

Любой браузер, например, Google Chrome v47 или выше.

8.3. Технические и электронные средства

При проведении дисциплины используются презентации лекционных материалов, для которых требуется: проектор, экран, компьютер для управления.

IX. МЕТОДОЛОГИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Методические указания приведены в учебных пособиях, перечисленных в разделе VII.

X. УЧЕБНАЯ КАРТА ДИСЦИПЛИНЫ (МОДУЛЯ)

Учебная карта дисциплины оформлена в виде приложения к рабочей программе дисциплины.

УЧЕБНАЯ КАРТА ДИСЦИПЛИНЫ
Обучение с подкреплением и его приложения

Трудоемкость: 5 зач.ед.

Форма промежуточной аттестации: экзамен

Курс 2, семестр 3

Код и наименование направления подготовки (специальности): 01.04.02 «Прикладная математика и информатика» (академическая магистратура)

Магистерская программа: «Искусственный интеллект: математические модели и прикладные решения»

№	Виды контрольных мероприятий	Текущий контроль	Рубежный контроль (при наличии)
	Модуль 1. Мета эвристики.	15	
1.	Аудиторная активность	2	
2.	Выполнение лабораторного задания	13	
	Модуль 2 Классическая теория.	15	
2.	Аудиторная активность	2	
	Выполнение лабораторного задания	13	
	Модуль 3. Value-based подход.	15	
1.	Аудиторная активность	2	
2.	Выполнение лабораторного задания	13	
	Модуль 4. Policy Gradient подход.	15	
1.	Аудиторная активность	2	
2.	Выполнение лабораторного задания	13	
	Всего	60	
	Промежуточная аттестация <i>в форме экзамена</i>	40 баллов	Экзамен проводится письменно по билетам. Критерий оценивания сформулирован в ФОС
	ИТОГО	100 баллов	

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования «Ростовский государственный экономический университет (РИНХ)»
Факультет компьютерных технологий и защиты информации
Кафедра фундаментальной и прикладной математики

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
ПО ДИСЦИПЛИНЕ (МОДУЛЮ) / ПРАКТИКЕ
ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ И ЕГО ПРИЛОЖЕНИЯ

Направление подготовки / специальность
01.04.02 «Прикладная математика и информатика»

ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ, ФОРМИРУЕМЫХ ДИСЦИПЛИНОЙ

«Обучение с подкреплением и его приложения»

Код компетенции	Формулировка компетенции
1	2
ПК	ПРОФЕССИОНАЛЬНЫЕ КОМПЕТЕНЦИИ
ПК-5	Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов.

ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

«Обучение с подкреплением и его приложения»

<i>№ п/п</i>	<i>Контролируемые разделы дисциплины*</i>	<i>Код контролируемой компетенции</i>	<i>Наименование оценочного средства**</i>
1.	Мета эвристики	ПК-5.1	Лабораторные задания. Экзаменационные билеты.
2.	Классическая теория	ПК-5.1	Лабораторные задания. Экзаменационные билеты.
3.	Value-based подход	ПК-5.1	Лабораторные задания. Экзаменационные билеты.
4.	Policy Gradient подход	ПК-5.1	Лабораторные задания. Экзаменационные билеты.

* Наименование раздела указывается в соответствии с рабочей программой дисциплины.

**Наименование оценочного средства указывается в соответствии с учебной картой дисциплины.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования «Ростовский государственный экономический университет (РИНХ)»
Факультет компьютерных технологий и защиты информации
Кафедра фундаментальной и прикладной математики

Вопросы к экзамену

по дисциплине «Обучение с подкреплением и его приложения»

1. Связь с оптимальным уравнением. Марковская цепь. Среда. Действия. Траектории. Марковский процесс принятия решений (MDP). Эпизодичность. Дисконтирование.
2. Условия задачи RL. On-policy. Концепция model-free алгоритмов. Классификация RL-алгоритмов.
3. Критерии оценки RL-алгоритмов. Сложности задачи RL. Бенчмарки.
4. Задача безградиентной оптимизации. Случайный поиск. Hill Climbing. Имитация отжига.
5. Эволюционные алгоритмы. Weight Agnostic Neural Networks (WANN). Видовая спецификация. Генетические алгоритмы.
6. Идея эволюционных стратегий. Оценка вероятности редкого события. Метод Кросс-Энтропии для стохастической оптимизации.
7. Метод Кросс-Энтропии для обучения с подкреплением (CEM). Натуральные эволюционные стратегии (NES). OpenAI-ES. Адаптация матрицы ковариации (CMA-ES).
8. Свойства траекторий. V-функция. Уравнение Беллмана.
9. Оптимальная стратегия. Q-функция. Принцип оптимальности Беллмана.
10. Отказ от однородности. Вид оптимальной стратегии (доказательство через отказ от однородности).
11. Уравнение оптимальности Беллмана. Критерий оптимальности Беллмана.
12. Advantage-функция. Relative Performance Identity (RPI). Policy Improvement.
13. Метод простой итерации. Policy Evaluation. Value Iteration. Policy Iteration. Generalized Policy Iteration.
14. Монте-Карло алгоритм. Экспоненциальное сглаживание. Стохастическая аппроксимация. Temporal Difference. Exploration-exploitation дилемма. Q-learning. SARSA.
15. Дилемма смещения-разброса. N-step Temporal Difference.
16. Backward View. Eligibility Trace. TD(λ).
17. Q-сетка. Переход к параметрической Q-функции. Декорреляция сэмплов. Таргет-сеть. DQN.
18. Twin DQN. Clipped Twin DQN. Double DQN. Dueling DQN.
19. Шумные сети (Noisy Nets). Приоритизированный реплей (Prioritized DQN). Multi-step DQN.
20. Идея Distributional подхода. Z-функция. Distributional-форма. Distributional Policy Evaluation. Distributional Value Iteration.
21. Категориальная аппроксимация Z-функций. Categorical DQN.
22. Квантильная аппроксимация Z-функций. Quantile Regression DQN.
23. Implicit Quantile Networks. Rainbow DQN.
24. Вывод первым способом. Вывод вторым способом. Физический смысл.
25. REINFORCE. State visitation frequency.
26. Расцепление внешней и внутренней стохастики. Связь с policy improvement.
27. Бейзлайн. Введение критика. Bias-variance trade-off.
28. Generalized Advantage Estimation (GAE).
29. Обучение критика. Advantage Actor-Critic (A2C).
30. Проблема переиспользования сэмплов. Суррогатная функция.
31. Нижняя оценка. Оптимизация нижней оценки. Trust Region Policy Optimization (TRPO). Proximal Policy Loss.
32. Clipped Value Loss. Proximal Policy Optimization (PRO).

Форма экзаменационного билета

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»
Факультет компьютерных технологий и защиты информации
Кафедра фундаментальной и прикладной математики

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №

По дисциплине "Обучение с подкреплением и его приложения"

Направление/специальность 01.04.02 «Прикладная математика и информатика»

1 Вопрос: Условия задачи RL. On-policy. Концепция model-free алгоритмов.
Классификация RL-алгоритмов.

2 Вопрос: Шумные сети (Noisy Nets). Приоритизированный реплей (Prioritized DQN). Multi-step DQN.

Составитель _____

(подпись)

Заведующий кафедрой _____

(подпись)

Критерии оценки

Баллы начисляются по результатам ответов на вопросы в соответствии с приведенной таблицей.

Оценочные таблицы экзамена по дисциплине «Обучение с подкреплением и его приложения»

К ответам на оба вопроса применяется следующая шкала баллов:

№	Точно "ДА"	Скорее "ДА"	Иногда "ДА", иногда "НЕТ"	Скорее "НЕТ"	Точно "НЕТ"
Вопрос 1	20	15	10	5	0
Вопрос 2	20	15	10	5	0

Таким образом, максимальное количество баллов, которое может получить студент за ответ на экзамене – 40.

Перечень лабораторных заданий

по дисциплине «Обучение с подкреплением и его приложения»

- Работа № 1.** Для игры «змейка», написанной на языке Python написать искусственный интеллект, управляющий агентом «змейка».
- Работа № 2.** Для игры «арканоид», написанной на языке Python написать искусственный интеллект, управляющий агентом «платформа».
- Работа № 3.** Для игры «пинг-понг», написанной на языке Python написать искусственный интеллект управляющий агентом «ракетка».
- Работа № 4.** Для игры «пятнашки», написанной на языке Python написать искусственный интеллект, собирающий головоломку.
- Работа № 5.** Для игры «сапер», написанной на языке Python написать искусственный интеллект, проходящий головоломку.
- Работа № 6.** Для игры «крестики нолики», написанной на языке Python написать искусственный интеллект, играющим против человека.
- Работа № 7.** Для среды, написанной на языке Python написать искусственный интеллект управляющий агентом, которому требуется пройти из одной части поля в другую, обходя препятствия.
- Работа № 8.** Для среды, написанной на языке Python написать искусственный интеллект управляющий агентом, которому требуется пройти из одной части поля в другую, обходя препятствия и уклоняясь от двигающихся объектов.

Критерии оценки

0-7 баллов оценивается программа с точки зрения выполнения поставленной задачи (программа делает то что требуется).

0-3 баллов оценивается программа с точки зрения соблюдения PEP-8.

0-3 баллов оценивается программа с точки зрения эффективности.

Итого, за правильную, хорошо написанную, эффективную программу студент получает максимум 13 баллов.

Общее количество баллов:

Раздел 1: 13 баллов;

Раздел 2: 13 баллов;

Раздел 3: 13 баллов;

Раздел 4: 13 баллов.

Таким образом, максимальное количество баллов, которое может получить студент за лабораторные работы – 52.

Перечень контрольных вопросов

- Вопрос 1. Что такое Марковский процесс принятия решений (MDP).
- Вопрос 2. Описать условия задачи RL.
- Вопрос 3. Что такое эволюционные алгоритмы.
- Вопрос 4. В чем заключается идея эволюционных стратегий.
- Вопрос 5. Описать алгоритм Монте-Карло.
- Вопрос 6. Написать уравнение Беллмана.
- Вопрос 7. Что такое Implicit Quantile Networks.
- Вопрос 8. Что такое Advantage Actor-Critic (A2C).

Критерии оценки

2 балла ставится в том случае, если студент ответил на вопрос полностью верно.

1 балл ставится тогда, когда студент ответил на вопрос скорее верно, чем не верно.

0 баллов ставится в том случае, когда студент не смог правильно ответить на поставленный вопрос.

Общее количество баллов:

Раздел 1: 2 балла;

Раздел 2: 2 балла;

Раздел 3: 2 балла;

Раздел 4: 2 балла.

Таким образом, максимальное количество баллов, которое может получить студент за аудиторную активность – 8.