**УДК 004.85**

# ОЦЕНКА КАЧЕСТВА ДАННЫХ ДЛЯ ОФФЛАЙН ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

**Долговязов А. Р.** (ИТМО)
**Научный руководитель – аспирант Асадулаев А.А.**
(ИТМО)

**Введение.** Алгоритмы обучения с подкреплением без взаимодействия в реальном времени в последнее время привлекли значительное внимание исследователей. Алгоритмы обучения без доступа к окружающей среде полностью полагаются на наборы данных, предоставленные экспертом. Однако предоставленные данные не всегда являются экспертными - они могут быть неполноценными или даже случайными. Было показано, что различные алгоритмы предпочтительны в зависимости от уровня опыта данных. Однако, даже когда данные являются экспертными, некоторые алгоритмы не могут выучить наиболее эффективную стратегию. В то же время, высокие результаты могут быть достигнуты при использовании неполноценных данных. В нашей работе мы ставим вопрос: как можно измерить качество офлайн-данных, чтобы предсказать их производительность без обучения агента? Чтобы дать ответ, мы исследовали различные инструменты, которые могут быть использованы для измерения качества данных.

**Основная часть.** В данной работе мы представляем новую метрику, способную оценить степень случайности набора данных для оффлайн обучения с подкреплением. Мы назвали её расстояние Беллмана-Вассерштайна. Основная идея предложенного решения - это измерение расстояния между распределением действий на данных и действий, сгенерированных с помощью политики агента. В качестве меры расстояния была использована Wasserstein-1 distance [2]. Мотивация её использования заключается в следующем: мы предотвращаем переобучение на оффлайн данных, вычисляя дифференцируемое расстояние, которое может быть минимизировано во время обучения политики.

Было показано, что в случае, если функция стоимости не отражает структуры проблемы, можно использовать специально разработанную неевклидову функцию [1]. Поэтому в нашей работе мы представляем особую функцию, которая измеряет различия между нашей и случайной политикой.

**Выводы.** Была представлена новая метрика и были показана корреляция между её значениями и качеством данных

**Список использованных источников**:

1. Асадулаев, А., Коротин, А., Егиазарян, В., и Бурнаев, Е. Нейронный оптимальный транспорт с общими функционалами стоимости. Предварительная печать arXiv:2205.15403, 2022
2. Фудзимото, С., Хооф, Х., и Мегер, Д. Решение проблемы ошибки аппроксимации функции в методах актор-критик. В Международная конференция по машинному обучению, с. 1587–1596. PMLR, 2018.