

УДК 004.89

Оптимизация демонстрации и переноса задач для роботизированных манипуляторов

Гао Тяньцы^(*) gaotianci0088@gmail.com

Ян Бо yangbo.123@hotmail.com

Жао Шэнжэнь raoshengren@gmail.com

МГТУ им. Н.Э. Баумана, Москва, Россия

Аннотация. Предложен метод оптимизации обучения промышленного роботизированного манипулятора, объединяющий методы обучения на основе демонстрации с алгоритмом глубокого обучения с подкреплением (Proximal Policy Optimization, PPO). Сначала используется обучение на основе демонстрации для генерации начальной траектории, затем PPO применяется для улучшения точности при минимальных затратах на обучающие выборки. Эксперименты показывают, что метод снижает объем данных на 15 % и улучшает адаптивность в переносе задач, открывая новые возможности для гибкого умного производства.

Ключевые слова: глубокое обучение с подкреплением, оптимизация стратегии, роботизированный манипулятор, обучение на основе демонстрации, перенос задач, эффективность выборки

Введение. Промышленные роботизированные манипуляторы применяются для задач, таких как сборка и сварка. Традиционные методы обучения требуют большого количества демонстрационных данных, что ограничивает адаптивность к динамическим условиям [1]. Обучение на основе демонстрации и глубокое обучение с подкреплением показали преимущества в оптимизации траекторий [2]. Однако их использование по отдельности часто недостаточно для многозадачности и ограниченных данных.

Материалы и методы; результаты. Для создания начальной траектории используется метод клонирования поведения, минимизирующий ошибку между предсказанным действием и действиями эксперта:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|\pi_{\theta}(s_i) - a_i\|^2. \quad (1)$$

После этого применяется PPO для точной настройки стратегии:

$$L^{PPO}(\theta) = E_t [\min(r_t(\theta) A^{\pi_{old}}(s_t, a_t), \text{clip}(r_t(\theta), 1 - \varepsilon, 1 + \varepsilon) A^{\pi_{old}}(s_t, a_t))]. \quad (2)$$

Стратегия переноса задач оценивается по сходству целей:

$$S(T_i, T_j) = \exp(-\|x_{\text{target}}^i - x_{\text{target}}^j\|^2 / \sigma^2). \quad (3)$$

Эксперименты проводились на платформе Mujoco для проверки метода в задаче сборки. Оценивались точность траектории, успешность выполнения задачи и снижение объема демонстрационных данных. Сравнивались методы

«традиционного обучения» и «обучения демонстрации + PPO». Рис. 1 иллюстрирует задачу сборки, где манипулятор собирает разбросанные детали в заданные позиции. Эта задача требует точного выполнения операций и служит для проверки адаптивности предложенного метода.

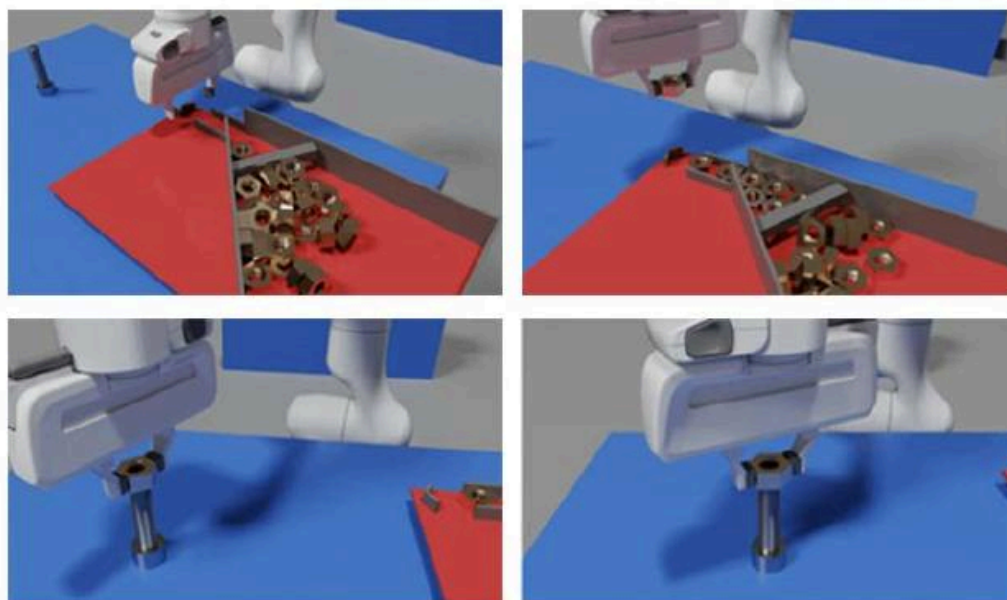


Рис. 1. Иллюстрация задачи размещения гаек промышленным манипулятором в процессе сборки

На рис. 2 показано сравнение трех методов: Имитационное обучение + PPO, Только PPO и Традиционное обучение. Имитационное обучение + PPO достигает наилучших результатов по горизонту принятия решений, накопленной отдаче и использованию памяти. Только PPO показывает умеренные результаты, тогда как традиционное обучение демонстрирует наименьшую производительность.

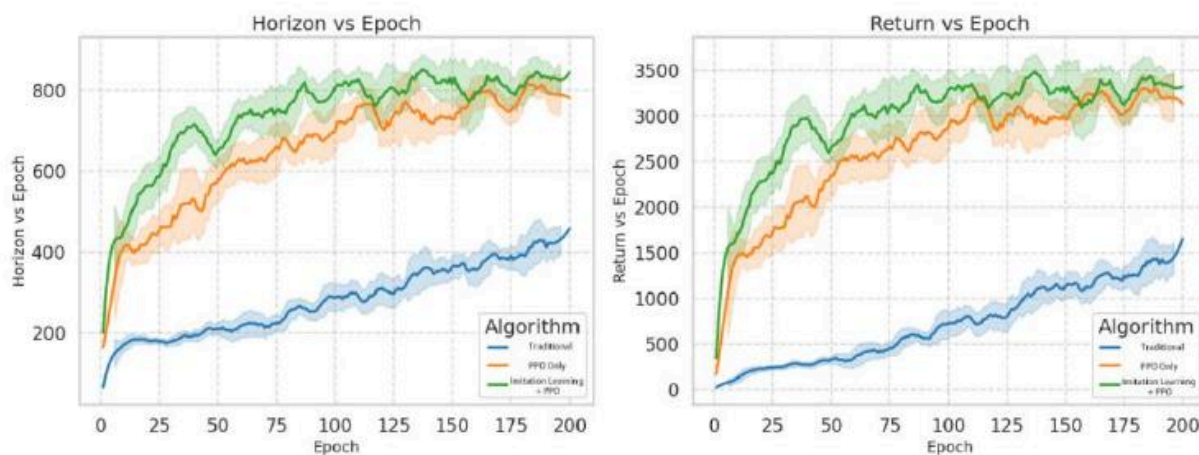


Рис. 2. Сравнение методов: показывает эффективность обучения демонстрации + PPO по сравнению с другими подходами

Заключение. Предложенный метод продемонстрировал высокую точность и адаптивность при минимальном объеме данных. В будущем планируется улучшение вычислительной эффективности и тестирование в реальных условиях.

Предложенный метод продемонстрировал высокую точность выполнения траектории и адаптивность к новым задачам при минимальном объеме демонстрационных данных. В будущем планируется оптим

Список источников

- [1] Levine S. End-to-End Training of Deep Visuomotor Policies. *Journal of Machine Learning Research*, 2016. <http://arxiv.org/abs/1504.00702>
- [2] Kramberger A. Deep Reinforcement Learning for Robotic Manipulation with Object Discovery. *IEEE Robotics and Automation Letters*, 2023.