ПРОБЛЕМАТИКА ПЕРЕНОСА АЛГОРИТМОВ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ С ИМИТАЦИОННЫХ МОДЕЛЕЙ НА РЕАЛЬНЫЕ ОБЪЕКТЫ

К.Ю. Усенко, студент гр. 8ЕМ02
А.Ю. Зарницын, ст. преподаватель ОЭИ
Томский политехнический университет
Email: kyu2@tpu.ru

Введение

На сегодняшний день алгоритмы обучения с подкреплением широко исследуются в применении к различным робототехническим задачам и задачам управления. Исследуются данные алгоритмы в большинстве случаев на симуляциях. Применение и обучение алгоритмов на реальных объектах имеет ряд ограничений таких как: несоответствие симуляции реальным объектам и окружающей среде, зависимость от полученных выборок данных, длительность обучения, проблема формирования функции награды. А также применение алгоритмов обучения с подкреплением вызывает беспокойство с точки зрения целостности реальных объектов.

Целью данной работы является описание проблематики переноса, предварительно обученных на симуляции алгоритмов обучения с подкреплением для решения задач управления на реальные объекты.

Обзор тематики

Алгоритмы обучения с подкреплением получили широкое распространение для выработки функции политики решения комплексных задач управления, начиная от прямого управления исполнительными механизмами, заканчивая высокоуровневым планированием и приятием решений [1-3]. Однако применение данных алгоритмов на реальных объектах вызывает множество сложностей [4]. Большая часть которых связаны с получением и обработкой данных, обучением алгоритма, несоответствие симуляций реальным объектам, перенос обученных алгоритмов на реальные объекты. Также существует беспокойство, связанное с безопасностью применения данных алгоритмов на реальных объектах. Применение может вызвать ускоренный износ деталей, а также выход из строя оборудования [5]. В данном тезисе внимание будет сосредоточено на рассмотрении проблематики переноса обученных алгоритмов на реальные объекты.

Постановка задачи

Пусть алгоритм обучения с подкреплением предварительно обучен и имеет функцию политики (управления) $\boldsymbol{\pi}(\boldsymbol{s})$, зависимой от состояния среды $\boldsymbol{s}$ для решения задачи автоматического управления на симуляции или математической модели реального объекта. Результат достиг заданных значений выбранного функционала качества $\boldsymbol{I}$ для управляемой, динамической, нестационарной системы. Где $\boldsymbol{I}$:

$$I = \int f(s,t)dt \to min$$

Тогда алгоритм, после переноса политики $\boldsymbol{\pi}(\boldsymbol{s})$, на реальный объект должен удовлетворять заданному функционалу качества системы.

$$I \approx I_m,$$

где I -достигнутый функционал качества системы;

I_m -достигнутый функционал качества модели.

Проблематика

Проблематика переноса может заключаться в трех основных аспектах: ресурсные ограничения, модельные ограничения, алгоритмические ограничения.

Ресурсные ограничения в первую очередь связаны с техническими сложностями применения нейросетевых алгоритмов на производстве с обучением в реальном времени (требование к вычислительным мощностям и достаточности памяти для хранения весов алгоритма и истории обучения и т.д.). Однако в большей степени результаты работы алгоритмов зависят от модели и ограничений связанный с моделированием и валидацией реальных объектов [6].

В первую очередь результат работы алгоритма зависит от степени приближённости модели к реальному объекту. При зависимости алгоритма от данных, полученных в процессе работы, такие случаи часто сопровождаются расхождением алгоритма при переносе вплоть до полной неспособности выполнять поставленную задачу. Наличие параметрических и структурных неопределенностей может, с одной стороны ухудшить прямые показатели качества переходного процесса, но с другой стороны в некоторых работах неопределенности вводятся искусственно для повышения робастности алгоритма управления.

При наличии таких неопределенностей в реальной системе, они обязательно должны быть отражены в модели, что накладывает дополнительные требования к моделированию системы.
Однако, невозможно учесть все неопределенности и параметры реального объекта. Соответственно, необходимо выработать критерий приближенности модели к реальному объекту и определить условия при котором модель можно описать как адекватную и допустимую для предварительного обучения алгоритма.

Алгоритмы также имеют ряд ограничений, связанных с формированием политик, сбором и анализом выборки для обучения алгоритмов. Функция политики алгоритма управления должна быть робастна и/или адаптивна для быстрой подстройки под реальный объект и учета всех параметров, которые не были внесены в модель. С обеспечением данных свойств есть трудности в виду длительности и сложности обучения и адаптации нейронных сетей к задачам в процессе управления, а также их зависимость от размерности и репрезентативности выборки полученных состояний объекта.

Функция политики, как при предварительном обучении, так и при работе на реальном объекте может быть подвержена техническим ограничениям на выработку управляющих воздействий.

Функция награды, используемая при обучении алгоритма, должна быть напрямую связана с поставленным функционалам качества. Функция награды должна быть монотонна, непрерывна, не должна противоречить функционалу качества, а также не должна содержать большее количество критериев, чем заявленный функционал качества.

Заключение

В заключении можно сказать, что в данной работе была обозначена проблематика переноса предварительно обученных на моделях для задач управления алгоритмов обучения с подкреплением на реальные объекты.

Были рассмотрены основные проблемы переноса на различных этапах построения системы управления: построение модели, формирование функции награды и функционала качества, требования к формированию функций политики и их свойствам. Исходя из обозначенной проблематики предлагается следующие направления работ по ее разрешению:

1. выработка критерия приближённости модели к реальному объекту, а также условий достаточности для применения и обучения алгоритма на модели;
2. рассмотрение влияния смоделированных и реальных неопределенностей на процесс обучения алгоритма и на результат его работы;
3. выработка единой системы требований к формированию функций наград и связи их с функционалом качества.

Список использованных источников

1. Mnih, V., K. Kavukcuoglu, D. Silver, A.A. Rusu and J. Veness, 2015. Human-level control through deep reinforcement learning. Nature, 518. [Электронный ресурс] – URL: web.stanford.edu/class/psych209/Readings/MnihEtAlHassibis15NatureControlDeepRL.pdf (дата обращения: 27.02.2022).
2. Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. A brief survey of deep reinforcement learning. // arXiv.org. [Электронный ресурс] - URL: https://arxiv.org/abs/1708.05866 (дата обращения: 27.02.2022).
3. Challenges of Real-World Reinforcement Learning // Arxiv.org [Электронный ресурс] - URL: https://arxiv.org/pdf/1904.12901.pdf (дата обращения: 22.12.2020).
4. Nguyen, T.T., N.D. Nguyen and S. Nahavandi, 2020. Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications. Transactions on Cybernetics (issue PP(99)), IEEE [Электронный ресурс] - URL: https://www.researchgate.net/publication/340068468_Deep_Reinforcement_Learning_for_Multiagent_Systems_A_Review_of_Challenges_Solutions_and_Applications (дата обращения: 27.02.2022).

5. Cheng, R., G.M. Orosz, R. Murray and J.W. Burdick, 2019. End-to-End Safe Reinforcement Learning through Barrier Functions for Safety-Critical Continuous Control Tasks. Arxiv, 1903 [Электронный ресурс] - URL: https://arxiv.org/pdf/1903.08792.pdf (дата обращения: 27.02.2022).
6. Zhao, W., J.P. Queralta and T. Westerlund, 2021. Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: a Survey. Arxiv, 2009. [Электронный ресурс] - URL: arxiv.org/pdf/2009.13303.pdf (дата обращения: 27.02.2022).