

ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ И ЕГО ИНТЕГРАЦИЯ С КЛАССИЧЕСКИМИ МЕТОДАМИ УПРАВЛЕНИЯ В РОБОТОТЕХНИКЕ

Евсеенко В. С.

Научный руководитель – Сиротин Ф.Л., к.т.н., доцент

Промышленные роботы, используемые сегодня в различных отраслях промышленности, в основном выполняют монотонные циклические задачи. Общая производительность задачи зависит от точности их систем управления для отслеживания заданных траекторий. Однако способность роботов производить сложные действия и принимать решения в различных средах довольно ограничена.

Например – это выборка ранее не встречавшихся объектов или вставка новых деталей в задачи сборки. Существует множество примеров впечатляющих роботов, демонстрирующих ловкость и продвинутый контроль, например, робот Fanta Challenge, или роботы, играющие в пинг-понг. Однако эти приложения сложно программировать и поддерживать, обычно они являются результатом докторской диссертации, и они не сделали скачок для промышленности. Желательно наделять машины более высоким уровнем интеллекта для самостоятельного приобретения навыков. Основная задача состоит в том, чтобы разработать адаптируемые, но надежные алгоритмы управления с учетом трудностей, присущих моделированию всех возможных поведений системы и необходимости обобщения поведения. Методы обучения с подкреплением (далее – RL) позволяющие агентам изучать поведение через взаимодействие с окружающей средой и идеально обобщать для новых сценариев.

Обучение с подкреплением – это метод контроля и обучения, который позволяет агентам изучать поведение через взаимодействие с окружающей средой [1]. В отличие от традиционных методов управления роботами, основная идея обучения с подкреплением состоит в том, чтобы показать и научить роботов тому, что конкретно делать в каждой возможной ситуации, а не как это делать. Таким образом, агент взаимодействует с окружающей средой и собирает наблюдения и сведения о состоянии среды и его положении в ней. В качестве реакции среды на действия агента выступают так называемые награды. Иногда это может быть абстрактным понятием, а иногда имеет физическую подоплеку (например подзарядка мобильного робота от станции, если он найдет путь к ней). Алгоритм «усиливает, подкрепляет» успешную политику – наборы действие-последствие, которые дают высокие награды. Все алгоритмы можно грубо разделить на два класса: основанные на максимизации функции награды и на поиске и улучшении политики. При поиске политики роботы изучают прямое сопоставление

состояний с действиями. В подходах, основанных на функции награды, роботы изучают функцию награды, промежуточную структуру, которая оценивает значение награды в данном состоянии и производят действия из значения этой функции.

Как поиск политики, так и подходы на основе значений-функций могут быть либо на основе модели, либо без модели. Безмодельные методы не учитывают динамику мира [2].

Как мы видим, методы управления роботами могут быть сгруппированы вдоль некоей линии, где на одном конце мы находим «жесткие» законы управления с обратной связью, которые разработаны вручную, включают знание предметной области, а структура управления не адаптируется данными. На другом конце спектра у нас есть RL, который позволяет изучать стратегии управления исключительно на основе данных наблюдений. Оба метода имеют свои преимущества и недостатки. Традиционные методы управления с обратной связью могут очень эффективно решать различные типы задач управления роботом, такие как отслеживание траектории в свободном пространстве, путем построения явных моделей, таких как уравнения движения твердого тела. Однако многие проблемы управления в современном производстве связаны с контактами и трением, которые трудно уловить при физическом моделировании. И если требуются рассуждения более высокого уровня (например, какой конкретный предмет необходимо выбрать из всех, что есть на конвейере), нынешним системам управления роботом не хватает гибкости. Применение методологий управления с обратной связью для решения подобных проблем часто приводит к хрупким и неточным системам, которые необходимо настраивать вручную. RL, с другой стороны, может, в принципе, изучить любую структуру управления. Однако для реальных роботов пространство исследования велико и, следовательно, требуются большие объемы данных и, следовательно, длительное время обучения. Кроме того, в отличие от обычного управления с обратной связью, подтверждение о сходимости и устойчивости трудно получить для методов RL.

Оптимальная системы управления для гибкого производства должна сочетать как традиционную теорию управления, так и управляемую данными RL. Традиционный контроль может обеспечить гарантии безопасности и производительности, в то время как RL может обеспечить гибкость и адаптивность, если настроен правильно.

Предлагается разложить сложные многоуровневые системы управления, на подзадачи, которые могут быть явно решены с помощью традиционных методов и подзадач, которые решаются с помощью RL. Окончательные стратегии управления представляют собой композицию как управляемых данными компонентами, так и стратегии управления из моделей первого порядка. Данный подход сочетает в себе преимущества традиционной теории управления с гибкостью RL. Например, управление положением осуществляется с помощью ПИД-регулятора, а RL представляет ту часть управления, которая имеет дело с трением и контактами. Уже были

проведены исследования по различным промышленно значимым случаям использования, среди которых, среди прочего, роботы для выполнения реальных сборочных задач с использованием контактов и нестабильных компонентов [3].

Однако у данного подхода все еще есть недостатки. Каждая итерация в среде с высоким коэффициентом трения может повредить деталь, контактирующую с захватом. Точные датчики и адекватное управление ограничениями могут облегчить проблему. Вся система лучше работает при подходах, использующих традиционное управление, которое может фильтровать выход RL алгоритма. Кроме того, чтобы сократить количество итераций в реальном мире, было доказано, что новые подходы в моделировании, направленные на изучение и сокращение различий между симуляцией и реальностью, ускоряют процесс обучения [4].

В заключение мы полагаем, что возросший интерес к применению обучения с подкреплением в робототехнике и на производстве в целом имеет вескую причину; однако, это не основной ингредиент, гарантирующий успех. Сквозные методы обучения показали низкую производительность в задачах, требующих точности. Поэтому несмотря на все успехи в области машинного обучения и др, необходимо помнить и применять классические подходы, гарантирующие результат.

Литература

1. Reinforcement Learning: An Introduction Second edition, in progress Richard S. Sutton and Andrew G. Barto, 2017. -С.1-2
2. Part1: Key concepts in RL // Spinning RL [Электронный ресурс]. — 2018. Режим доступа: https://spinningup.openai.com/en/latest/spinningup/rl_intro.html — Дата доступа : 21.04.2019.
3. Reinforcement learning shows promise for industrial robotics// Therobotreport [Электронный ресурс]. — 2018. Режим доступа: <https://www.therobotreport.com/reinforcement-learning-industrial-robotics> — Дата доступа : 20.04.2019.
4. Sim2Real View Invariant Visual Servoing by Recurrent Control [Электронный ресурс]. — 2017. Режим доступа: <https://arxiv.org/pdf/1712.07642.pdf>. — Дата доступа : 27.04.2019.