Прикладные алгоритмы обучения с подкреплением и их применение

Тардова Александра

Декабрь 2023

Содержание

1. Введение

Очевидно, что с каждым годом математическая база работ усложняется вместе с увеличением вычислительной мощности, машины за единицу времени могут обрабатывать больше метрик от года к году и выдавать более точные решения. Применение глубокого обучения уже хорошо зарекомендовало себя, так как применение такого подхода дало качественный скачек для многих сфер деятельности. Обучение с подкреплением интересно тем,что в отличие от традиционного машинного обучения (сучителемибез него), позволяет изучить долгосрочные стратегии и применять их к сложным промышленным и бизнес-задачам. Данный подход в направлениях создания искусственных систем работает лучше всего, когда решения принима- ются последовательно, а действия связаны с исследованием окружающей среды. В статье рассмотрены особенности ряда программных систем имитации, применяющих методы обучения с под- креплением. Основное внимание уделено идейной стороне применяемых в имитационных программах безмо- дельных алгоритмов, которые теоретически применимы к любой задаче. Они изучают стратегии через взаимодействие, усваивая при этом любые правила окружающей среды. В результате сравнения алгоритмов выявлены их достоинства и недостатки.

2. Анализ содержания статей и оценка полученных в них результатов

На сегодняшний день многие компатентные авторы уделят внимние применению метода обучения с подкреплением для решения прикладных задач.

В данной секции я хочу рассказать о методах, сравнить статьи на основе новизны предлагаемых их ывторами алгоритмов и оптимизаций.

Начнем со статьи "ВОЗНИКНОВЕНИЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ"(Шарибаев А.Н., Шарибаев Р.Н., Абдулазизов Б.Т. доцен Тохиржонова М.Р,) в ней освящена история обучения с подкреплением. Она важна для понимания современного состояния искусственного интеллекта и его возможностей. Психологические исследования и изучение поведения животных сыграли ключевую роль в развитии этой области, а ранние работы ученых, таких как Скиннер, Торндайк и Халл, заложили основу для разработки алгоритмов обучения с подкреплением. Из интересного, в статье рассказывается, что концепция обучения с подкреплением основана на идее, что организм может научиться совершать определенные действия, основываясь на последствиях этих действий. Например, в 1930-х годах американский психолог Б.Ф. Скиннер ввел концепцию оперантного обусловливания, представляющий собой тип обучения, при котором поведение модифицируется его последствиями. Работа Скиннера была сосредоточена на том, как можно обучить организмы реагировать на раздражители, основываясь на последствиях их действий. Например, крыса могла бы научиться нажимать на рычаг, чтобы получить пищевое вознаграждение, или избегать нажатия на рычаг, если она получила удар электрическим током.

Продолжим знакомство со спецификой обучения с подкреплением статьей Ю. Г. Степина, В. М. Локтевича «Обучение с подкреплением для решения оптимизационных задачах». В ней объясняется разница различных видов обучения, применяемых в последние года. Это наиболее общее деление, тем не менее, данная статья хорошо подходит для ознакомления с тематикой, а так же для людей далеких от машинного обучения и программирования, но все же желающих назбираться в последних тенденциях хотя бы поверхностно. Итак, автор объясняет, что для обучения с учителем (Supervised Learning) агенту предоставляются заранее "помеченные"данные: пары , на которых он учится. (Например, так можно обучить спам-фильтр, предоставляя ему заранее отсортированный на спам и не-спам поток почты.) Обучение без учителя (Unsupervised Learning) – агенту предоставляются необработанные ("непомечен-

ные") данные, в которых он должен обнаружить скрытые взаимосвязи. Обучение с Подкреплением (Reinforcement Learning, RL) – здесь данных нет, агент учится, взаимодействуя со средой: наблюдает состояние, совершает действие, получает обратную связь – награду (которая может быть как поощрением, так и наказанием). Автор объясняет, что результат такого метода обучения это некая политика (стратегия) управления, то есть способ выбора агентом действия, исходя из наблюдения среды, а вовсе ен сами действия или другие измеримые метрики. После обучения политика сохраняется с агентом и может быть применена / усовершенствована в будущем. Здесь уклон не столько на алгоритмы МО в целом, сколько на углубление в конкретную сферу Deep Learning с агентом моделирования. Далее разбирается подробный пример глубокого обучения с подкреплением (Deep Reinforcement Learning,DRL), в котором нейронная сеть учится управлять светофором на перекрёстке.

В следующей статье - Кабыш А.С., профессор Головко В.А. МОДЕЛЬ КООРДИНАЦИИ ПОВЕДЕНИЯ АГЕНТОВ НА ОСНОВЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ описывается модель для нахождения оптимального поведения многоагентной структуры через организацию в ней оптимальных взаимодействий между агентами. Модель включает в себя две основные техники. Модель графов координации позволяет явно выразить зависимость между агентами, что позволяет разбить целевую функцию поведения в линейную сумму индивидуальных целевых функций. Модель оценки влияний позволяет оценить влияния других агентов на действия друг друга и в результате позволяет им координировать свои действия. В работе приведена реализация данной модели на основе обучения с подкреплением и экспериментальные результаты применения данной модели.

Также в ней рассматривается Метод обучения Q-learning, применяемый при агентном подходе в управлении и имитационном моделировании. Если среда поощряет агента за «правильные» действия (целевая функция Q представля- ется в виде награды), агент продолжает их со- вершать, т. е. стимулируем агента совершать хорошие для среды действия. При этом на вход алгоритма не поступает обучающая выборка. Агент должен методом проб и ошибок сам собрать данные при взаимодействии со средой. Если алгоритмах машинного обучения с учителем при настройке ИНС коррекция проис- ходила за счет минимизации ошибки между ее выходом и эталоном, то здесь источником коррекции является награда. На основе получаемого от среды вознаграждения агент формирует функцию ценности Q, что дает ему возмож- ность уже не случайно выбирать стратегию по- ведения, а учитывать опыт предыдущего

взаи- модействия со средой. Чем выше награда, тем лучше для агента.

Далее рассмотрим статью "ОСНОВНЫЕ ПОДХОДЫ К НЕЙРОЭВОЛЮЦИОННОМУ ИМИТАЦИОННОМУ МОДЕЛИРОВАНИЮ МАКРОЭКОНОМИЧЕСКИХ СИСТЕМ"Коротеев М.В. (Финансовый университет при правительстве РФ). Автор переходит к более прикладным вещам, и рассматривает обучение с подкреплением и моделирование действий агента в конкретной сфере. Он рассматривает насущную проблему ожиданий и нерациональность поведения экономических агентов, и последствия, к которы это может привести, а именно, к нестабильности и кризисам в экономической системе. Таким образом, возрастающий интерес к агентному и имитационному моделированию экономических систем связан с необходимостью учета структурной гетерогенности и асимметрии информации, а также с признанием того, что экономические агенты далеки от рационального и равномерного поведения. Моделирование на основе агентных подходов позволяет учитывать весь спектр поведения экономических агентов и их взаимодействие в условиях неопределенности и ограниченной рациональности, что делает их более реалистичными и адекватными для анализа экономических кризисов и рецессий.

В следующей статье "ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ"А.А.Емельянов, Е.А.Власова обсуждаются агентные и имитационные модели экономических систем, которые являются методами машинного обучения и моделирования. В агентном моделировании экономическая система представляется как динамическая система взаимодействующих агентов с ограниченной рациональностью. Эти агенты могут быть как индивидуальными участниками рынка, так и фирмами, банками или другими экономическими субъектами. В процессе агентного моделирования используются алгоритмы машинного обучения, которые позволяют моделировать поведение и взаимодействие агентов на основе полученных данных. Это может включать в себя использование различных методов обучения с подкреплением, глубокого обучения, методов кластеризации и т.д.

Имитационное моделирование также используется для анализа экономических систем и включает создание моделей на основе симуляции взаимодействия множества агентов в условиях неопределенности. Для этого могут применяться методы статистического моделирования, теории игр, а также алгоритмы машинного обучения для анализа больших объемов данных.

Следующая статья "Агентное моделирование как современный метод исследования инновационных экономических систем"посвящена применению обучению с

подкреплдением в бизнес-задачах, для адаптации и инновациям для выживания на рынке. В ней рассматриваются методологические подходы в имитационном моделировании на основе агентного моделирования, ориентированного на решение практических проблем. Авторы рассматривают процесс принятия решений в крупных компаниях и проблему недоверия сотрудников к решениям руководства, так как первые не в состоянии просчитать все возможные исходы, а так же нечувствительны к желаниям сотрудников. В противопоставлении обыкновенному процессу принятия решений предлагает агентный подход в имитационном моделировании основан на индивидуальном поведении агентов и компьютерном моделировании для анализа и оптимизации бизнес-процессов.

Логичным прродолженнием данной тематики может оказаться статья ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ КАК ТЕХНОЛОГИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ РЕШЕНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЗАДАЧ: ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ АЛГОРИТМОВ Е.В. Орлова. Здесь описывается использование методов RL для управления человеческими ресурсами на уровне предприятия. В тексте затрагиваются алгоритмы, основанные на полезности, на стратегии и комбинированные алгоритмы, но по большей части текст предоставляет детальный обзор областей, в которых используется обучение с подкреплением, а также приводит примеры применения RL для различных задач в экономике, финансах и бизнесе. Например, рассказано, что в финансовой отрасли RL используется для управления рисками, оптимизации инвестиций, прогнозирования цен на финансовых рынках. Аналогичные методы активно применяются в маркетинге для оптимизации ценовой политики, управления рекламными кампаниями и улучшения пользовательского опыта.

Рассмотрим еще одну статью, посвященную оптимизации задач из бизнес-сферы "Манипуляторы с управлением на основе машинного обучения с подкреплением"Али Сажи Маннаа,Андрей О. Зарубин. Авторы демонстрируют пример проекта: обучение с подкреплением в управлении производством. Он рассматривает пример реального использования ИИ для решения бизнес задачи компании Лагор (Италия) производящей ферромагнитные сердечники для трансформаторов. Для повышения эффективности производства компанией Engineering Ingegneria Informatica был разработан цифровой двойник производства, который представляет собой детальную имитационную модель, разработанную в AnyLogic, интегрированную со SCADA-системой, предающей текущее состояние производственной линии и сердечников

в стадии производства. Применение эвристических алгоритмов для управления перемещением сердечников по цеху Пленарные доклады ИММОД - 2019 28 показало существенный прирост общей эффективности производства по сравнению с ручным планированием. Однако, автор отмечает, что в сложных случаях эвристики не справлялись с поставленной задачей. В качестве решения специалисты компании применили глубокое обучение с подкреплением агентов, управляющих перемещением сердечников на производственной линии. Для обучения агентов использовалась имитационная модель, разработанная в AnyLogic с использованием библиотеки для глубокого машинного обучения DL4J. Таким образом, основная задача управления процессом производства была разбита на более простые подзадачи (перемещение сердечника к конкретной целевой позиции), а для каждой подзадачи производилось обучение соответствующего агента. После дообучения в среде модель стала показывать хоршие результаты двже в трудных случаях, требовавших ранее консультации с экспертами.

Итак, с алгоритмами обучения с подкреплением, а также прикладными задачами, в которых он используется, мы уже немного познакомились, поэтому далее было бы интеерсно рассмотреть проблемы данной области. Она неплохо представлены в статье "ПРОБЛЕМЫ В ОБЛАСТИ ГЛУБОКОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ"Шарибаев А.Н, Шарибаев Р.Н., Абдулазизов Б.Т., Тохиржонова М.Р. Исследователи в статье обсуждают прогресс и проблемы обучения с подкреплением (RL). Они отмечают, что RL требует большого количества взаимодействий с окружающей средой и сталкивается с проблемами обобщения для новых задач, а также безопасности, стабильности и объяснимости. Для устранения этих проблем и ограничений исследователи разрабатывают новые подходы, такие как RL на основе моделей, многозадачность и мета-RL, обратный RL и безопасный RL. Эти подходы направлены на повышение эффективности, обобщенности, безопасности, стабильности и объяснимости систем RL, что может открыть новые возможности в робототехнике, автономных системах и взаимодействии человека и роботов.

Далее уделим внимание еще одной проблеме в области обучения с подкреплением - 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. Она освящвется в статье

В статье подчеркивается, что RL находится на пути к значительному прогрессу,

но, чтобы преодолеть текущие проблемы, необходимо разработать новые подходы, которые улучшат его возможности в различных сферах применения. Обнаружение методов, обеспечивающих безопасность, обобщение, и стабильность в RL, является приоритетной задачей, и возможно, что дальнейшее развитие RL окажет значительное влияние на науку и технику, а также произведет новую эру в робототехнике, автономных системах и взаимодействии человека и робота.

В продолжении рассмотрим статью ПРОБЛЕМАТИКА ПЕРЕНОСА АЛГОРИТМОВ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ С ИМИТАЦИОННЫХ МОДЕЛЕЙ НА РЕАЛЬНЫЕ ОБЪЕКТЫ К.Ю. Усенко, А.Ю. Зарницын. Изложенная тематика говорит о важности проблемы переноса обученных алгоритмов обучения с подкреплением на реальные объекты. Основные сложности связаны с ресурсными, модельными и алгоритмическими ограничениями. Ресурсные ограничения касаются технических сложностей применения нейросетевых алгоритмов на производстве, требования к вычислительным мощностям, памяти для хранения весов алгоритма, истории обучения и других аспектов, специфичных для реальных объектов. Модельные ограничения указывают на важность моделирования и валидации реальных объектов. Это включает ограничения, связанные с точностью математической модели, симуляцией и способностью модели предсказывать динамику реальных объектов. Алгоритмические ограничения связаны с результатами работы алгоритмов, их способностью удовлетворять заданным функционалам качества системы после переноса на реальные объекты. Эти проблемы подчеркивают важность развития методов, которые позволяют адаптировать обученные алгоритмы к реальным объектам, учитывая их специфические особенности и ограничения. Такой подход позволит более эффективно использовать алгоритмы обучения с подкреплением в реальных промышленных и производственных сферах.

Продожим освещать проблемы использования RL статьей Обучение с подкреплением для управления: проблемы стабилизации динамических систем, Павел Осиненко. В данной статье помимо анализа основных часто используемых алгоритмов, каких как Q-обучение, глубокое обучение, PPS, и т.п., освещенных мной уже по предыдущим статьям, рассматривается такая оптимизация, как добавление constraint-ов, позволяющих приблизить действия, которые может выбирать ИИ к более реальным человеческим решения, продиктованным например страхом за жизнь, здоровье, свое или чужое, также с помощью constraint-ов обучающегося агента учат общечеловеческим принципам, таким как мораль, любовь к ближнему и т. п. чтобы его ре-

шения не основывались на чистой оптимизации полезности от итоговой функции.

В заключение, хочу рассмотреть статью "ТЕКУЩЕЕ СОСТОЯНИЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ И НАПРАВЛЕНИЯ НА БУДУЩЕЕ чтобы понять вектор в котором движется современная разработка алгоритмов машинного обучения. Авторы приходят к следующим выводам:

1. RL (обучение с подкреплением) переживает быстрый прогресс и широкое применение в различных областях, включая робототехнику, игры, системы рекомендаций и автономное вождение.

2. Одно из перспективных направлений в RL - разработка более эффективных алгоритмов для выборки, таких как мета-RL, многозадачное обучение и RL на основе моделей.

3. Интеграция глубокого обучения (Deep RL), сочетающая глубокие нейронные сети с RL, показывает значительные перспективы в различных приложениях, таких как игры и робототехника.

4. Растущий интерес в разработке алгоритмов RL, которые могут извлекать уроки из отзывов людей и руководств, включая RL "Человек в цикле".

5. Развитие безопасных и этичных систем RL, чтобы обеспечить соответствие человеческим ценностям и предпочтениям, а также предотвратить нежелательное или опасное поведение агентов RL.

3. Заключение

В заключение, можно отметить, что алгоритмы обучения с подкреплением играют важную роль в прикладных задачах экономики и моделирования бизнес-процессов. Они позволяют создавать эффективные стратегии принятия решений, оптимизировать процессы и улучшать результаты бизнеса. Благодаря развитию и применению этих алгоритмов, компании могут получать конкурентные преимущества и улучшать свои финансовые показатели. Таким образом, использование алгоритмов обучения с подкреплением в экономике и бизнесе имеет большой потенциал для оптимизации процессов и достижения успеха на рынке.

4. Источники

1) "ВОЗНИКНОВЕНИЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ"Шарибаев А.Н., Шарибаев Р.Н., Абдулазизов Б.Т. доцен Тохиржонова М.Р

2) Ю. Г. Степин, В. М. Локтевич «Обучение с подкреплением для решения оптимизационных задачах».

3) Кабыш А.С., профессор Головко В.А. МОДЕЛЬ КООРДИНАЦИИ ПОВЕДЕНИЯ АГЕНТОВ НА ОСНОВЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

4) ОСНОВНЫЕ ПОДХОДЫ К НЕЙРОЭВОЛЮЦИОННОМУ ИМИТАЦИОННОМУ МОДЕЛИРОВАНИЮ МАКРОЭКОНОМИЧЕСКИХ СИСТЕМ, Коротеев М.В. (Финансовый университет при правительстве РФ)

5) Агентное моделирование как современный метод исследования инновационных экономических систем

6) ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ, А.А.Емельянов, Е.А.Власова

7) ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ КАК ТЕХНОЛОГИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ РЕШЕНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЗАДАЧ: ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ АЛГОРИТМОВ, Е.В. Орлова

8) "Манипуляторы с управлением на основе машинного обучения с подкреплением"Али Сажи Маннаа,Андрей О. Зарубин.

9) Обучение с подкреплением для управления: проблемы стабилизации динамических систем, Павел Осиненко

10) ПРОБЛЕМАТИКА ПЕРЕНОСА АЛГОРИТМОВ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ С ИМИТАЦИОННЫХ МОДЕЛЕЙ НА РЕАЛЬНЫЕ ОБЪЕКТЫ К.Ю. Усенко, А.Ю. Зарницын

11) ПРОБЛЕМЫ В ОБЛАСТИ ГЛУБОКОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ Шарибаев А.Н, Шарибаев Р.Н., Абдулазизов Б.Т., Тохиржонова М.Р.

12) ТЕКУЩЕЕ СОСТОЯНИЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ И НАПРАВЛЕНИЯ НА БУДУЩЕЕ Шарибаев А.Н., Шарибаев Р.Н., Абдулазизов Б.Т. доцен Тохиржонова М.Р