СИНТЕЗ СИСТЕМЫ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ НА ОСНОВЕ ПОДХОДА ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

А.Ю. Зарницын, ассистент ОАР,
К. Ю. Усенко, студент гр.8ЕМ02.
Томский политехнический университет
E-mail: kyu2@tpu.ru

Введение

Актуальность задач автоматического управления с заданными критериями в системах с параметрическими возмущениями с течением времени не спадает. Это обусловлено тем, что практически во всех реальных системах автоматического управления есть изменчивость параметров. Однако при изменении характера системы или изменения характера внешних воздействий картина меняется. Существует множество работ по применению классических адаптивных алгоритмов управления нелинейных систем с плавающими параметрами. Большинство таких алгоритмов не способны нивелировать те параметрические возмущения, которые не были учтены, а тем более не обладают свойством дообучения с целью достижения желаемой динамики. Для того чтобы нивелировать эти недостатки применяют интеллектуальные методы правления.

Целью данной работы является создание интеллектуального алгоритма автоматического управления на основе подхода обучения с подкреплением, и апробация данного алгоритма на модели.

Описание алгоритма

Поставленная в работе цель может быть декомпозирована на следующие задачи:

- синтезировать алгоритм интеллектуального автоматического управления на основе подхода обучения с подкреплением;
- предварительно обучить и апробировать алгоритм на математической модели стенда по изучению алгоритмов автоматического управления.

Для решения поставленных задач необходимо разработать архитектуру агента и подготовить математическую модель (среду) для его обучения. Предварительное обучение на реальных системах нежелательно из-за длительности обучения, сложности вычислений и излишней стохастичности, предварительное обучение будет проводится с помощью построенной математической модели.

Обучение с подкреплением строится на моделях агента и среды. Сам метод заключается во взаимодействии агента и среды, где агент за совершенные действия получает от среды ее состояние и награду (R). Задача агента максимизировать награду.

При разработке агента сначала необходимо выделить параметры среды на основе которых будет обучаться агент, множество действий агента, а также спроектировать структуру поощрений и наказаний агента.

Для реализации метода были выделены следующие параметры состояния среды:

- ток на двигателе турбины;
- позиция шара;
- скорость вращения двигателя;
- целевое значение позиции шара (для задания уставки).

Агент совершает действие каждую 0.1 с. и передает значение напряжения на двигатель от 0 до 10 вольт.

Награда агента зависит от близости реального положения шара к заданной уставке и определяется гауссовой зависимостью, с эвристически выведенными коэффициентами. Агент получает максимальную награду, когда реальное положение шара совпадает с уставкой. При этом награда убывает пропорционально возможному удалению шара от заданного установившегося значения.

Диаграмма сущностей системы управления роботом представлена на рисунке 1.

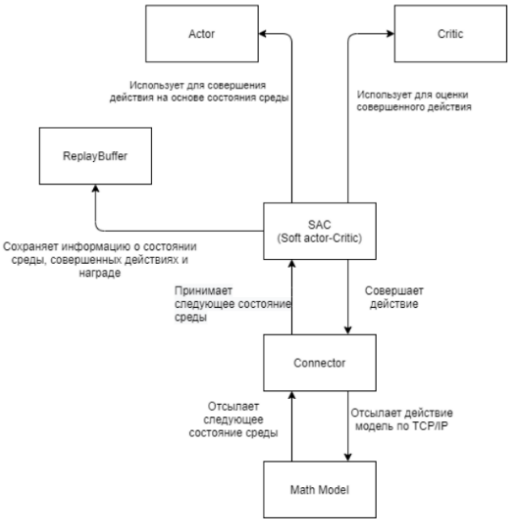

Рис. 1. Диаграмма сущностей системы управления.

Подробное описание принципа работы алгоритма SAC приведено в работе [1]

Апробация алгоритма

Алгоритм был предобучен на математической модели, составленной в MatLab Simulink. Сообщение между агентом и мат. Моделью MatLab произведено с помощью протокола TCP.

Модель агента апробирована и предварительно обучена на математической модели стенда. В будущем планируется перенос предобученной модели на реальный стенд. На рисунке 2 можно наблюдать переходный процесс системы.

На рисунке 3 можно увидеть выходной сигнал с SAC модели, подаваемый в математическую модель.

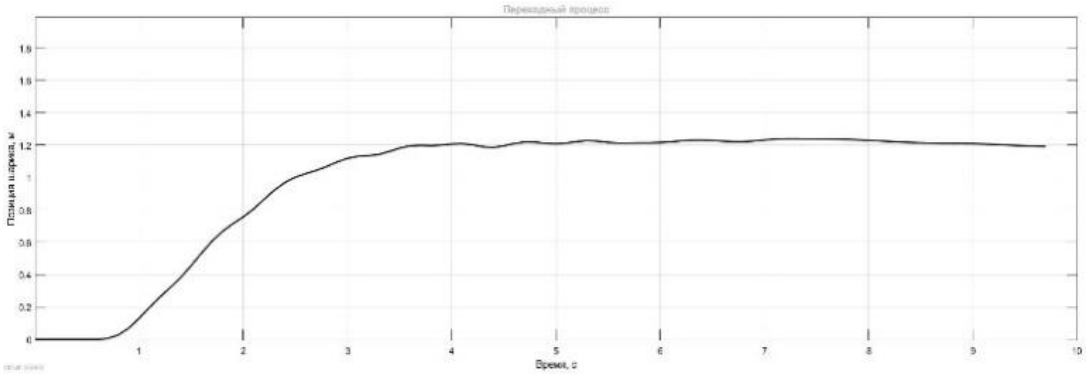

Рис. 2. Переходный процесс системы

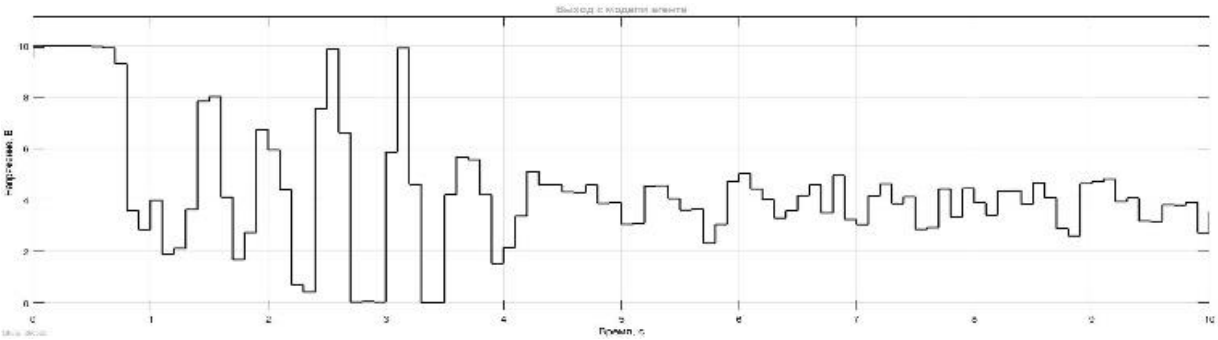

Рис. 3. Выход с модели, подаваемый в математическую модель.

Заключение

В заключении можно сказать, что в данной работе был успешно разработан и апробирован на математической модели алгоритм обучения с подкреплением Soft Actor-Critic. Метод, который

способен при правильной формализации задачи, а также входных и выходных данных, справиться с задачей автоматического управления на математической модели, без жесткой подстройки под конкретную задачу. Также несомненным преимуществом данного алгоритма является способность так называемого «активного обучения», то есть способность обучаться в процессе работы системы и дообучаться, что положительно скажется на работе системы при внешних возмущениях или перемены параметров самой системы.

В дальнейшем планируется апробировать алгоритм на реальном учебном стенде и сравнить с алгоритмами оптимального управления, а также апробировать алгоритм на стенде с изменяющимися параметрами объекта управления и сравнить работу алгоритма с алгоритмами оптимального управления в аналогичных условиях

Список использованных источников

1. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor // Arxiv.org [Электронный ресурс] - URL: https://arxiv.org/pdf/1801.01290.pdf (дата обращения: 22.12.2020).
2. Challenges of Real-World Reinforcement Learning // Arxiv.org [Электронный ресурс] - URL: https://arxiv.org/pdf/1904.12901.pdf (дата обращения: 22.12.2020).