

УДК 519.711.3

ПРИМЕНЕНИЕ МУЛЬТИАГЕНТНОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ ДЛЯ УПРАВЛЕНИЯ ЭНЕРГОПОТРЕБЛЕНИЕМ ЗДАНИЯ

О.Ю. Марьясин

Ярославский государственный технический университет

Россия, 150023, Ярославль, Московский пр., 88

E-mail: maryasin2003@list.ru

А.Н. Плохотнюк

Ярославский государственный технический университет

Россия, 150023, Ярославль, Московский пр., 88

E-mail: admin@nixson.ru

Ключевые слова: мультиагентная система, обучение с подкреплением, мультиагентное обучение с подкреплением, оптимизация энергопотребления.

Аннотация: В работе описывается мультиагентная система для управления микроклиматом и энергопотреблением здания в которой агенты действуют на основе мультиагентного обучения с подкреплением. Из различных вариантов мультиагентного обучения с подкреплением в работе рассмотрен вариант марковской игры с общим множеством состояний для всех агентов и вариант с отдельным множеством состояний. В работе приводится сравнение результатов, полученных с помощью различных алгоритмов обучения с подкреплением и различных вариантов мультиагентного обучения, по значению суммарного энергопотребления для отопления и охлаждения помещений здания в течение суток. Установлено, что вариант марковской игры позволяет получить чуть лучшие результаты по сравнению с одноагентным обучением, а наименее привлекательным с точки зрения энергосбережения оказался вариант с отдельным множеством состояний.

1. Введение

Концепция многоагентных (мультиагентных) систем (Multi-Agent Systems – MAS) была предложена в еще в середине 1980-х годов и к настоящему времени хорошо проработана как в теоретическом и методологическом плане, так и подкреплена инструментальными средствами моделирования и поддержки. Однако, несмотря на все свои достоинства, технология MAS до сих пор не нашла широкого применения в индустрии [1].

Обучение с подкреплением (Reinforcement Learning – RL), как часть машинного обучения, также использует понятие агента [2]. В RL агент взаимодействует со средой, наблюдая текущее состояние и используя данную информацию при выборе

своего действия. Среда в результате действия агента переходит в следующее состояние и возвращает агенту новое состояние и вознаграждение. В настоящее время алгоритмы RL широко применяются для решения сложных задач в таких областях как игры, робототехника, автономные мобильные транспортные средства и другие. Растет интерес, связанный с использованием алгоритмов RL для управления сложными техническими системами, например, энергетическими [3].

Успехи, достигнутые в последние годы в области одноагентного обучения, мотивируют исследователей использовать потенциал этих методов для мультиагентного обучения. Это создает возможности для объединения подходов MAS и RL на основе мультиагентного обучения с подкреплением (Multi-Agent Reinforcement Learning – MARL). В данной работе описывается MAS для управления микроклиматом и энергопотреблением здания в которой агенты действуют на основе MARL.

MARL аналогично одноагентной системе, моделируется как обобщение марковского процесса принятия решений (Markov Decision Process – MDP) в виде кортежа $\langle S, A, T, R \rangle$, где S – множество состояний среды, A – множество действий агентов, T – функция перехода, возвращающая вероятность переходов из состояния s_t в состояние s_{t+1} после выполнения агентами действия a , R – вознаграждение, получаемое агентами после выполнения действия a .

В научной литературе приводится множество примеров классификации задач MARL [4] в зависимости от представления множества состояний, действий, вознаграждения и других факторов. Из всех вариантов отметим случаи: с общим множеством состояний S для всех агентов, с отдельными множествами состояний $S_i \in S$ для каждого из агентов $i = 1, \dots, N$, с совместными действиями агентов A , с независимыми действиями агентов A_i , с общей наградой R для всех агентов, с независимой наградой $R_i \in R$ для каждого из агентов. Например, случай $\langle S, A_i, T, R_i \rangle$ относится к марковским играм (Markov games), называемым также стохастическими играми [5].

Применение мультиагентного подхода сталкивается с рядом проблем, отсутствующих при одноагентном обучении. Одной из проблем является то, что, когда одновременно обучаются несколько агентов, они изменяют свое окружение, а это изменение в окружении влияет на них самих и других агентов. Цикл обратной связи приводит к нескольким фазам обучения, каждая из которых зависит от предыдущей.

2. Решение задачи оптимизации энергопотребления здания с использованием MARL

Одной из перспективных областей применения MAS является управление микроклиматом и энергопотреблением зданий [6]. Ранее одним из авторов была предложена архитектура MAS управления зданием [7]. Полный перечень агентов MAS включает: локальных агентов, зональных агентов, агентов-людей, центральных агентов, агентов поставщиков, агентов внутренней и внешней среды. Конкретные реализации MAS управления зданием могут содержать только отдельные типы агентов. Например, описанная в данной работе MAS управления микроклиматом и энергосбережением здания включает только зональных агентов. Зональные агенты

относятся к наиболее интеллектуальным агентам MAS управления зданием [8]. Одной из задач, решаемых зональными агентами является задача оптимизации энергопотребления здания при соблюдении условий комфортности проживания людей.

В качестве модели системы в задаче оптимизации энергопотреблением здания выступает энергомодель одноэтажного здания, разделенного на 5 зон. Зона 1 находится в центре здания, а по ее сторонам располагаются зоны 2–5 соответственно. Для построения и выполнения расчетов по энергомодели применяется система энергомоделирования (Building Energy Modeling – BEM) зданий EnergyPlus [9]. С точки зрения моделирования энергомодель здания является сложной нелинейной моделью.

В качестве критерия оптимальности для задачи оптимизации энергопотребления будут выступать энергозатраты на отопление и охлаждения помещений зон здания в виде:

$$(1) \quad Q = \sum_{t=1}^H \sum_{i=1}^Z Q_{it}.$$

где Q_{it} – количество энергии, затраченное на отопления или охлаждения i -ой зоны здания в момент времени t , $t = 0, \dots, H$, H – горизонт управления, Z – число зон здания.

Введем ограничения, связанные с обеспечением комфортного микроклимата внутри помещений здания:

$$(2) \quad T_{itl} \leq T_{it} \leq T_{itu}, \quad i = 1, \dots, Z, \quad t = 0, \dots, H,$$

$$(3) \quad \varphi_{itl} \leq \varphi_{it} \leq \varphi_{itu}, \quad i = 1, \dots, Z, \quad t = 0, \dots, H,$$

где T_{itl} , T_{itu} – нижнее и верхнее значения температуры воздуха в i -ой зоне здания в момент времени t , φ_{itl} , φ_{itu} – нижнее и верхнее значения относительной влажности воздуха в i -ой зоне здания в момент времени t .

Таким образом, ставится задача определения значений уставок по температуре T_{ihst} и T_{icst} локальных регуляторов, минимизирующих критерий (1), при выполнении ограничений (2),(3).

Общая задача оптимизации энергопотребления здания (1)–(3) может быть разбита на подзадачи оптимизации энергопотребления отдельных зон здания в момент времени t . В этом случае в подзадачах, в качестве критерия оптимальности будет выступать Q_{zt} , а в качестве ограничений – подмножества ограничений (2),(3) для каждой конкретной зоны $i = 1, \dots, Z$ и момента времени $t = 0, \dots, H$. Для решения отдельных подзадач могут использоваться индивидуальные агенты, и, следовательно, общая задача оптимизации энергопотребления здания может быть решена с использованием подхода MARL. Применение MARL для управления микроклиматом и энергопотреблением зданий уже не является чем-то принципиально новым [10], однако работ, посвященных данной теме пока еще очень мало.

Из различных вариантов MARL в данной работе рассмотрен вариант $\langle S, A_i, T, R_i \rangle$ марковской игры с общим множеством состояний S для всех

агентов, с независимыми действиями агентов A_i и независимой наградой R_i (вариант 1), а также вариант $\langle S_i, A_i, T, R_i \rangle$ с отдельным множеством состояний S_i (вариант 2). Формирование стратегий агентов осложняется тем, что множества состояний S_i для отдельных зон здания не являются независимыми. Температура воздуха в каждой зоне здания зависит от температуры воздуха в соседних зонах из-за передачи тепла через ограждающие конструкции, разделяющие эти зоны. Поэтому действия агента по изменению температурного режима своей зоны будут влиять на температурный режим соседних зон и, следовательно, на действия агентов, управляющих этими зонами.

В данной работе рассматриваются только такие популярные алгоритмы RL которые могут работать с непрерывным пространством состояний и действий. Это алгоритмы: Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C) и Deep Deterministic Policy Gradient (DDPG). Полученные в результате работы алгоритма PPO суточные графики энергозатрат на отопление или охлаждение зон здания, связанные с функциями вознаграждения (reward) агентов для варианта 1 MARL показаны на рис. 1. Расчеты производились для весеннего периода года.

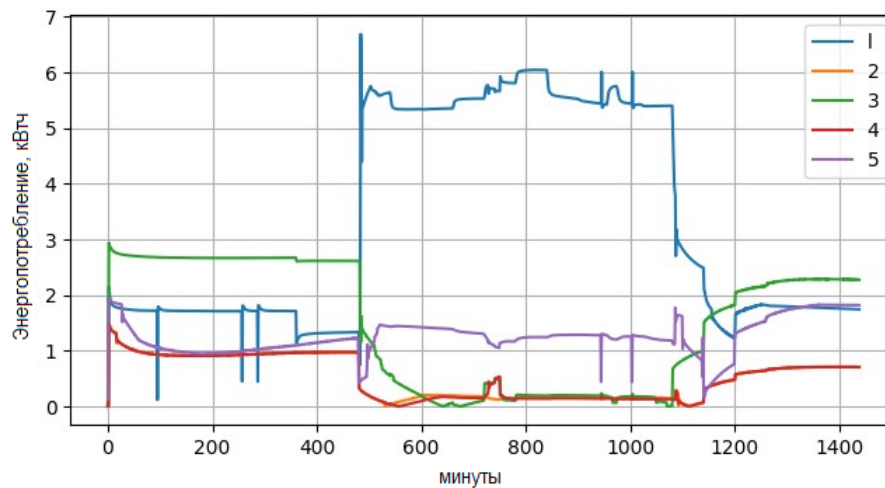


Рис. 1. Графики энергозатрат на отопление или охлаждение зон здания для алгоритма PPO

В таблице 1 приведены значения суммарного энергопотребления Q для отопления и охлаждения помещений здания в течение суток для всех рассмотренных в работе вариантов MARL и алгоритмов RL. Для сравнения в таблице 1 также приведены значения Q для алгоритмов RL при одноагентном обучении.

Таблица 1. Суммарное энергопотребление для различных вариантов

Алгоритм	Суммарное энергопотребление для варианта 1 MARL, кВт/час	Суммарное энергопотребление для варианта 2 MARL, кВт/час	Суммарное энергопотребление для одноагентного обучения, кВт/час
PPO	104.614	106.131	105.411
A2C	104.116	106.959	104.787
DDPG	101.148	105.817	104.753

Данные таблицы 1 позволяют сравнить результаты как в плане различных алгоритмов RL, так и в плане различных вариантов обучения с подкреплением. Анализ таблицы показывает, что из всех алгоритмов RL, наибольшее снижение энергопотребления дает использование алгоритма DDPG. Вариант 1 (марковская игра) позволяет получить чуть лучшие результаты (меньшее суммарное энергопотребление) по сравнению с одноагентным обучением (централизованное управление). Наименее привлекательным в точки зрения энергосбережения оказался вариант 2 (децентрализованное управление).

3. Заключение

В работе описывается пример MAS для управления микроклиматом и энергопотреблением здания в которой агенты действуют на основе MARL. Из различных вариантов MARL в работе рассмотрен вариант 1 марковской игры с общим множеством состояний для всех агентов и вариант 2 с отдельным множеством состояний. В работе приводится сравнение результатов, полученных с помощью различных алгоритмов RL и различных вариантов обучения с подкреплением, по значению суммарного энергопотребления для отопления и охлаждения помещений здания в течение суток. Установлено, что вариант 1 MARL позволяет получить чуть лучшие результаты по сравнению с одноагентным обучением, а наименее привлекательным в точки зрения энергосбережения оказался вариант 2.

Список литературы

1. Городецкий В.И., Скобелев П.О. Многоагентные технологии для промышленных приложений: реальность и перспектива // Труды СПИИРАН. 2017. Вып. 6(55). С. 11–45.
2. Гессер Л., Кенг В.Л. Глубокое обучение с подкреплением: теория и практика на языке Python. СПб.: Питер, 2022. 416 с.
3. Perera A.T.D., Kamalaruban P. Applications of reinforcement learning in energy systems // Renewable and Sustainable Energy Reviews. 2021. Vol. 137. P. 1–22.
4. Петренко В.И. Классификация задач мультиагентного обучения с подкреплением // Известия Кабардино-Балкарского научного центра РАН. 2021. № 3. С. 32–44.
5. Canese L., Cardarilli G.C., Nunzio L.D., Fazzolari R., Giardino D., Re M., Spano S. Multi-Agent Reinforcement Learning: A Review of Challenges and Applications // Applied Sciences. 2021. Vol. 11. P. 1–25.
6. Yang R, Wang L. Development of multi-agent system for building energy and comfort management based on occupant behaviors // Energy and Buildings. 2013. Vol. 56. P. 1–7.
7. Марьясин О.Ю. Проектирование мультиагентной системы управления зданием с использованием онтологий // Онтология проектирования. 2018. Т. 8, № 3. С. 387–399.
8. Марьясин О.Ю. Мультиагентная система управления зданием с интеллектуальными агентами // Труды XIII Всероссийского совещания по проблемам управления ВСПУ-2019. Москва, 17–20 июня 2019 г. М.: Институт проблем управления им. В.А. Трапезникова РАН, 2019. С. 2366–2370.
9. EnergyPlus. <https://energyplus.net/> (дата обращения: 14.11.2023).
10. Fu Q., Chen X., Ma S., Fang N., Xing B., Chen J. Optimal control method of HVAC based on multi-agent deep reinforcement learning // Energy and Buildings. 2022. Vol. 270. P. 1–19.