

Применение алгоритмов оптимизации для повышения надёжности работы энергосистем

В России, которая обладает большой территорией с высокой неравномерностью её заселения, важным аспектом развития страны является обеспечение надёжного энергоснабжения потребителей, проживающих как в зоне централизованного энергоснабжения, так и в зоне децентрализованных энергосистем. Согласно исследованиям [1, 2], в Российской Федерации имеется более 500 изолированных поселений и объектов с автономными системами энергоснабжения общей установленной мощностью более 1 ГВт. В настоящий момент основным видом топлива для изолированных населённых пунктов и объектов является дизельное топливо, однако его использование отличается как сложной логистикой (что отражается на стоимости топлива для конечного потребителя), так и высоким углеродным следом. Одним из вариантов альтернативного развития автономных энергосистем является внедрение в них возобновляемых источников энергии (ВИЭ), которые более волатильны по сравнению с дизельной генерацией. Перспективным способом повышения надёжности подобных автономных энергосистем является снижение неопределённости выработки энергии ВИЭ за счёт применения механизмов оптимизации и прогнозирования.

Задача прогнозирования

Проблема прогнозирования выработки (в том числе с помощью ВИЭ) и потребления электроэнергии является ключевой для всех задач оптимизации микросети (уменьшенной версии централизованной системы электроснабжения, созданной для увеличения надёжности поставок электроэнергии, повышения энергонезависимости за счёт диверсификации источников энергии, а также иных энергосетевых структур способных работать в том числе автономно). Благодаря прогнозированию появляются возможность принимать управляющие решения на основе как текущих системных условий,

так и условий, которые сложатся в микросетях в будущем — на горизонте планирования. Для прогнозирования энергопотребления на основании имеющейся истории строятся и постоянно уточняются прогностические модели, структура которых зависит от объёма имеющихся данных.



Полученные с помощью прогностических моделей прогнозы независимых параметров оптимизации (потребления и генерации ВИЭ) используются в краткосрочной и оперативной оптимизации. Для решения задачи оперативной оптимизации используется оперативный прогноз потребления (порядка нескольких минут), полученный с помощью методов анализа и формирования временных рядов, основанных на авторегрессионных подходах (Autoregressive Integrated Moving Average, ARIMA) либо на интерполяции наиболее «свежих» точек траектории из фактического режима и точек из будущего, формируемых в задачах более дальнего прогноза. Прогноз ВИЭ-генерации на ветровых (ВЭС) и солнечных (СЭС) электростанциях составляется с использованием прогноза погоды и математических моделей, связывающих мощность этих электростанций с прогнозируемыми на расчётный срок метеопараметрами.

Рецензия эксперта на статью получена 11.07.2023 [The expert review of the article was received on July 11, 2023]

УДК 620.9. Научная специальность: 2.4.5 (05.14.01).

Применение алгоритмов оптимизации для повышения надёжности работы энергосистем

А. И. Величко, генеральный директор ООО «Смартрен»; **М. С. Повернов**, технический директор ООО «Дельта П»; **Н. Шувалов**, магистр, Санкт-Петербургский политехнический университет Петра Великого (СПбПУ)

Статья посвящена использованию алгоритмов оптимизации в энергосистемах с высокой долей ВИЭ. На изолированных территориях России расположены сотни поселений и объектов с общей установленной электротехнической мощностью более 1 ГВт. Предлагаемые в статье решения по оптимизации энергобаланса способствуют повышению надёжности и предсказуемости работы локальных энергосистем.

Ключевые слова: автономные энергосистемы, оптимизация энергобаланса, возобновляемые источники энергии, надёжность энергосистем.

UDC 620.9. The number of scientific specialty: 2.4.5 (05.14.01).

Application of algorithms to ensure the reliability of power systems

A. I. Velichko, general director of "Smartren", LLC; **M. S. Povernov**, technical director of "Delta P", LLC; **N. Shuvalov**, Master's Degree, [Peter the Great St. Petersburg Polytechnic University \(SPbPU\)](#)

The article is devoted to the use of optimization algorithms in power systems with a high share of renewable power sources. Hundreds of settlements and facilities with a total installed electric power capacity of more than 1 GW are located in the isolated territories of Russia. The solutions proposed in the article to optimize the energy balance contribute to improving the reliability and predictability of the operation of local energy systems.

Key words: autonomous power systems, energy balance optimization, renewable energy sources, energy system's reliability.



Рис. 1. Принципиальная схема микросети (microgrid)

Задача оптимизации

Управление производством и/или потреблением электроэнергии в микросети является сложной задачей, основной целью которой является достижения оптимального использования энергоресурсов. Оптимизация работы микросетей может включать следующие задачи:

- максимизация выработки в определённый промежуток времени;
- минимизация операционных затрат микросети в определённый промежуток времени;
- минимизация использования дизельного топлива;
- максимизация заработка на рынке (в случае параллельной работы с основной, централизованной сетью) [3].

Выбор методологии для решения задачи оптимизации в микросетях зависит от многих факторов: целевых функций, установленной мощности микросети, её конфигурации, объёма данных и других.

Нелинейное программирование

Некоторые исследователи используют модели на основе линейного (Linear Programming, LP) и нелинейного (Non-Linear Programming, NLP) программирования. Задача нелинейного программирования, включающая в себя нелинейность целевой функции и (или) ограничений:

$$f(x) \rightarrow \min,$$

где $g_i(x)$:

$$\text{Minimize } f(x)$$

$$\text{Subject to } g_i(x) \leq 0 \text{ for each } i \in \{1, \dots, m\}$$

$$H_j(x) = 0 \text{ for each } j \in \{1, \dots, p\}$$

$$x \in X.$$

Так, в статье [4] использовано смешанное целочисленное программирование для определения оптимальной площадки и конфигурации микросети на этапе проектирования. В работе [5] была предложена модель оптимизации микросети, состоящей из дизельных генераторов, на основе линейного программирования.

Мультиагентная система

«Агентом» в мультиагентной платформе считается система, расположенная в среде, имеющая возможность выполнять действия автономно, чтобы соответствовать цели проектирования системы (рис. 2). Агенты, в отличие от других систем, способны выполнять действия иначе, чем программно-аппаратные системы.

Благодаря прогнозированию появляется возможность принимать управляющие решения на основе как текущих системных условий, так и условий, которые сложатся в микросетях в будущем. Для этого используют модели на основе линейного и нелинейного программирования

Так, агентам присущи следующие свойства [6, 7]:

1. **Реактивность** — способность агента реагировать на изменения внутри среды без прямого за счёт использования интеллектуальной системы, не влияя напрямую на окружающую среду.
2. **Автономность** — способность агента самостоятельно выполнять свои задачи в сети без внешнего вмешательства со стороны других агентов или людей.
3. **Ответная реакция** — способность агента полностью наблюдать за текущим состоянием среды и реагировать на него за минимально возможное время для внесения изменений в среду.
4. **Проактивность** — способность агента реагировать на различные условия, не обращая внимания на общее поведение среды.
5. **Социальность** — способность агента взаимодействовать с внешними источниками (с окружающей средой), такими как другой агент или человек.

Алгоритмы на основе теории игр

Под алгоритмами на основе теории игр подразумевают алгоритмы, имитирующие действие игроков в различных конфликтах. Различают корпоративные и некорпоративные теории игр.

В работе [8] предложен алгоритм для оптимизации агрегированного спроса для снижения издержек с использованием алгоритмов на основе теории игр, снизив издержки как пула* потребителей, так и каждого потребителя по отдельности.

* Пул (pool) — разновидность монополии, объединение нескольких участников рынка, причём с общим фондом, состоящим из всех расходов и доходов участников.

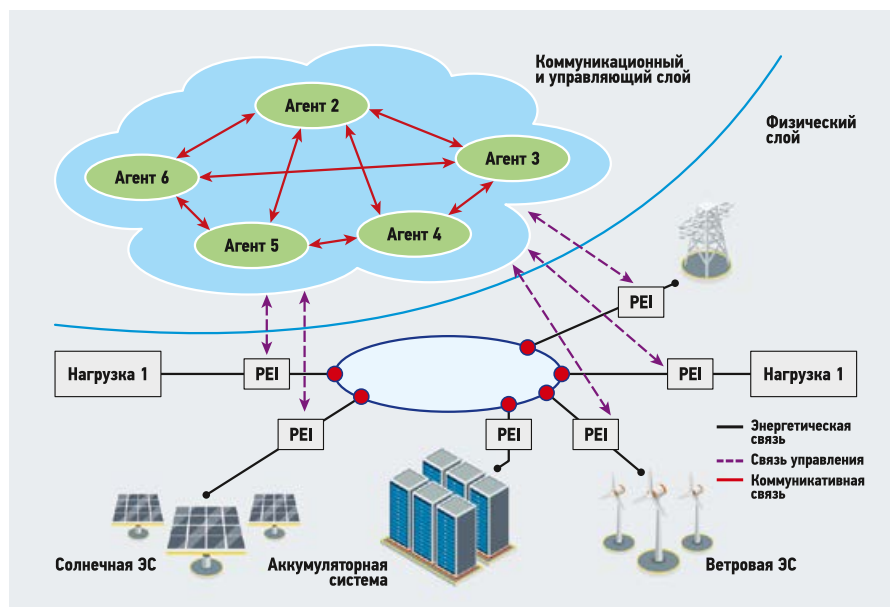


Рис. 2. Мультиагентная платформа

Стохастические методы оптимизации

Стохастическая оптимизация (Stochastic Optimization, SO) — это направление теории оптимизации, в котором модель или алгоритм используется для определения наилучшего распределения имеющихся ресурсов. Она особенно полезна для оптимизации микросетей благодаря своей способности учитывать неопределённые факторы, такие как прогноз погоды, изменения нагрузки пользователей и меняющееся производство энергии из возобновляемых источников. Цель стохастической оптимизации — найти наиболее эффективный и экономичный способ использования имеющейся энергии с учётом стохастической природы регрессоров модели.

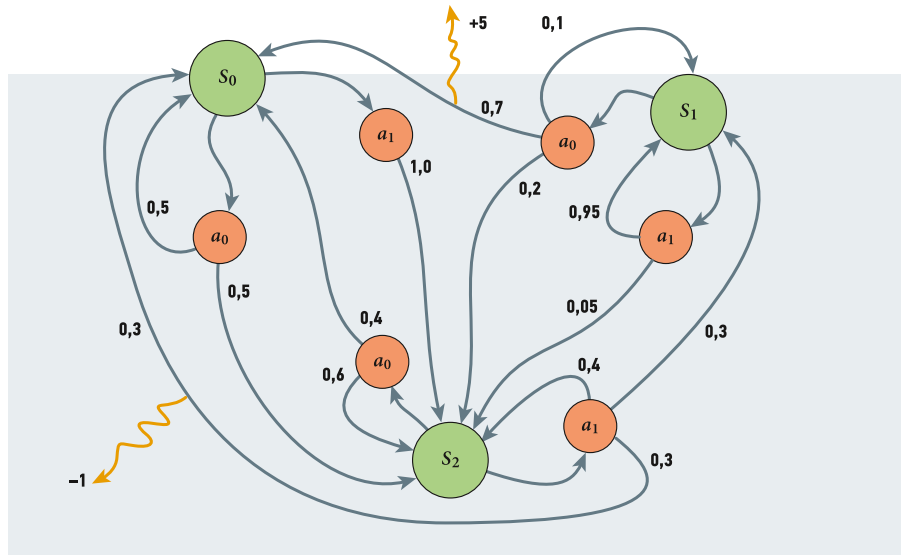


Рис. 3. Пример алгоритма MDP с тремя состояниями и двумя действиями



Стохастическая оптимизация также может использоваться и для управления хранением энергии в системе микросети. Принимая во внимание постоянные колебания в производстве возобновляемой энергии, алгоритм может решить, какие источники энергии использовать, когда хранить энергию и в каком количестве. Это помогает обеспечить микросеть достаточным количеством электроэнергии для удовлетворения спроса при минимизации затрат. Наконец, алгоритмы стохастической оптимизации могут быть использованы для оптимизации использования различных источников энергии в системе микросети. Алгоритм может оценить несколько вариантов, чтобы определить источник, который обеспечивает наиболее экономически эффективную электроэнергию. Он также может учитывать доступность энергии из различных источников и определять приоритетность этих источников, чтобы гарантировать, что энергия распределяется наиболее эффективным образом.

Рассматривая эти алгоритмы микросети, можно убедиться, что они оптимально используют свои источники энергии.

В результате вся система в целом может стать более устойчивой и менее уязвимой к колебаниям как производства электроэнергии, так и спроса на неё.

Выделяется два основных типа алгоритмов стохастической оптимизации: эволюционные алгоритмы (evolutionary algorithms) и алгоритмы роевого интеллекта (swarm algorithms). Выбор конкретного алгоритма стохастической оптимизации в значительной степени зависит от основных целей, а также от конкретных характеристик микросети.

Если целью создания микросети является минимизация эксплуатационных расходов, то будет выбран алгоритм управления с прогнозированием модели Model Predictive Control. Алгоритмы управления с прогнозированием модели Model Predictive Control оптимизированы для генерации наилучшего решения по управлению, как в детерминированных, так и в стохастических условиях

Например, если целью является минимизация эксплуатационных расходов, то алгоритмом выбора, скорее всего, будет алгоритм управления с прогнозированием модели (Model Predictive Control, MPC). Это связано с тем, что алгоритмы MPC оптимизированы для генерации наилучшего решения по управлению, как в детерминированных, так и в стохастических условиях.

Кроме того, алгоритмы MPC имеют ряд преимуществ перед другими методами оптимизации — они могут работать с несколькими целями и ограничениями, а также могут применяться на нескольких временных шагах.

С другой стороны, если целью является, например, оптимизация использования возобновляемых источников энергии и минимизация углеродного следа сети, то лучшим выбором может быть стохастический алгоритм оптимизации, такой как алгоритм марковского процесса принятия решений (Markov Decision Process, MDP). Это связано с тем, что алгоритмы MDP специально разработаны для учёта неопределённости, что позволяет лучше оптимизировать и прогнозировать будущее использование энергетических ресурсов (рис. 3).

В конечном счёте выбор лучшего алгоритма для оптимизации микросети в значительной степени будет зависеть от желаемых целей и характеристик конкретной микросети.

Решить марковский процесс принятия решений означает найти оптимальную стратегию, максимизирующую «вознаграждение» (функцию ценности). Самая простая функция ценности — это математическое ожидание формального ряда:

$$E \left[\sum_{t=0}^{\infty} R_{A_t}(s_t, s_{t+1}) \right],$$

где $a_t = \pi(s_t)$, а математическое ожидание берётся в соответствии с $s_{t+1} \sim P_{A_t}(s_t)$.

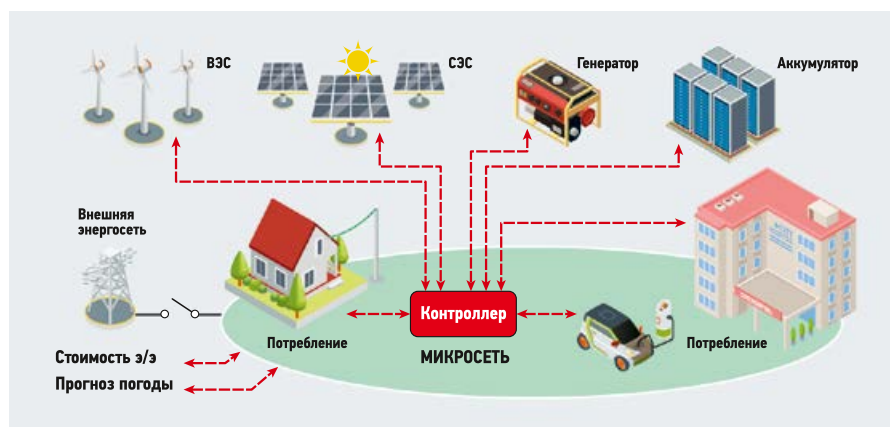


Рис. 4. Пример smart microgrid («умной» системы распределения электроэнергии)

Алгоритмы на основе искусственного интеллекта

Широкую популярность в оптимизации микросетей получили алгоритмы на основе искусственного интеллекта, в этом случае некоторые исследователи часто выделяют отдельный класс микросетей — smart microgrid (рис. 4).

Smart microgrid — это «умная» система распределения электроэнергии, которая соединяет между собой потребление электроэнергии, объекты распределённой энергетики и накопители электрической энергии в чётко определённых электрических границах, чтобы действовать как единый контролируемый объект по отношению к основной сети [9, 10].

Основными алгоритмами, которые могут быть использованы в smart microgrid, в данном случае могут быть:

- нейронные сети (neural networks);
- метод роя частиц (Particle Swarm Optimization, PSO);
- алгоритмы нечёткой логики (fuzzy logics), задействующие классическую логику и теорию множеств;
- метод имитации поведения бактерий (Bacterial Foraging Optimization, BFO);
- алгоритмы обучения с подкреплением.

Обучение с подкреплением в оптимизации микросетей

Обычно задача обучения с подкреплением (Reinforcement Learning, RL) заключается в нахождении оптимальной стратегии на основе собственного опыта. Модель обучения с подкреплением состоит из среды, множества действий и агента. Задача обучения с подкреплением основана на гипотезе о вознаграждении, которая утверждает возможность достижения любой цели (или решения любой задачи) через максимизацию вознаграждения.

В результате взаимодействия агента со средой агент может получать награду (рис. 5), то есть переходит в состояние s_{i+1} , получая вознаграждение r_i . Задачей в об-



Рис. 5. Взаимодействие агента со средой

учении с подкреплением является максимизация математического ожидания вознаграждения $E\pi[G] \rightarrow \max$.

Имплементация многих алгоритмов обучения с подкреплением основана на дилемме «исследование или эксплуатация» (exploration vs exploitation), то есть баланса между максимизацией функции вознаграждения и исследования среды.

В свою очередь, алгоритмы с подкреплением можно разделить на model-free и model-based. Так, в model-free алгоритмам отсутствует модель, описывающая

Алгоритмы model-free делятся на Q-learning и Policy optimization. Они могут показать одинаковые результаты в простых задачах, при этом в Q-обучении нужно максимизировать Q-value во всём дискретном пространстве действий, а алгоритмы Policy optimization нацелены на поиск взаимосвязи между пространством состояний и действий

среду, а в model-based алгоритмах такая модель присутствует — $p(s_0 | s, a)$. Классификация алгоритмов с подкреплением представлена на рис. 6.

Алгоритмы типа model-free можно разделить на алгоритмы Q-learning и Policy optimization. Данные алгоритмы могут показать одинаковые результаты в простых задачах, при этом в Q-обучении нужно максимизировать Q-value во всём дискретном пространстве действий, а алгоритмы типа Policy optimization нацелены на поиск взаимосвязи между пространством состояний и действий, который могут быть непрерывными.

Алгоритмы Policy optimization, в отличие от алгоритмов Q-learning, применяются в задачах с большим количеством возможных действий.

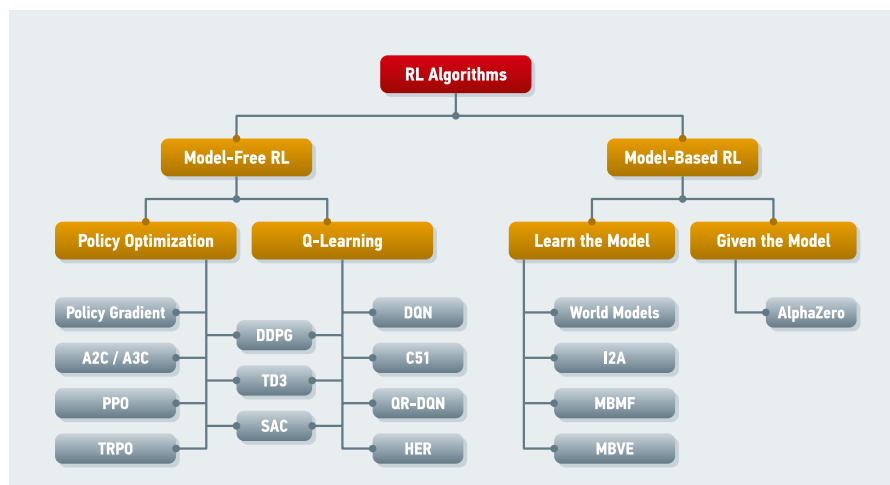


Рис. 6. Классификация алгоритмов с подкреплением



Алгоритмы Q-learning

Большой популярностью пользуются алгоритмы Q-learning, которые направлены на определение функции полезности Q. Данные алгоритмы построены на принципе оптимальности Беллмана.

Q-обучение — это обучение с подкреплением без модели, без политики, которое находит наилучший курс действий, учитывая текущее состояние агента. В зависимости от того, где агент находится в окружающей среде, он будет решать, какое следующее действие предпринять.

Цель модели — найти наилучший курс действий с учётом текущего состояния. Для этого он может придумать собственные правила или действовать вне рамок заданной ему политики. Отсутствие модели означает, что агент использует предсказания ожидаемой реакции среды для продвижения вперёд. Он не использует систему вознаграждения для обучения, а скорее метод проб и ошибок:

$$V(s) = \max E[r_0 + \gamma V(s')]; \pi.$$

Глубинное Q-обучение

Главным недостатком алгоритма Q-learning является возможность его применения только для очень маленьких сред, поскольку алгоритм быстро теряет свою целесообразность при увеличении числа состояний и действий в среде. Решение вышеупомянутой проблемы приходит от осознания того, что значения в матрице имеют только относительную ценность, то есть значения имеют ценность только по отношению к другим значениям.

Таким образом, данное мышление приводит нас к глубинному Q-обучению (Deep Q-learning, DQL), который использует глубинную нейронную сеть (Deep Neural Network, DNN) для аппроксимации значений. Эта аппроксимация значений не вредит, пока сохраняется относительная важность. Основной рабочий шаг для Deep Q-learning заключается в том, что начальное состояние подаётся в нейронную сеть, которая возвращает Q-значение всех возможных действий в качестве выхода.

Deep Q-learning — это вариант Q-learning, который использует глубинную нейронную сеть для представления Q-функции, а не простую таблицу значений. Это позволяет алгоритму обрабатывать среды с большим количеством состояний, а также обучаться на основе высокоразмерных данных, таких как изображения или данные датчиков.

Одна из основных проблем при реализации глубинного Q-обучения заключается в том, что Q-функция обычно нели-

Главным недостатком алгоритма Q-learning является возможность его применения для очень маленьких сред, поскольку алгоритм быстро теряет свою целесообразность при увеличении числа состояний и действий в среде. Решение здесь в том, что значения в матрице имеют только относительную ценность

нейна и может иметь множество локальных минимумов. Это может затруднить сходимость нейронной сети к правильной Q-функции. Для решения этой проблемы было предложено несколько методов, таких как воспроизведение опыта и целевые сети. Воспроизведение опыта является непосредственно методом, при котором агент хранит подмножество своего опыта (состояние, действие, вознаграждение, следующее состояние) в буфере памяти и делает выборку из него для обновления Q-функции. Это помогает декоррелировать данные и сделать процесс обучения более стабильным.

Целевые нейронные сети, с другой стороны, используются для стабилизации обновлений Q-функции. В этой технике отдельная нейросеть используется для вычисления целевых Q-значений, которые затем используются для обновления сети Q-функции.

Глубинное Q-обучение применяется для решения широкого круга задач, включая игры, робототехнику и автономные транспортные средства. Например, оно использовалось для обучения агентов, которые могут играть в Atari и Go, и для управления роботами, выполняющими такие задачи, как захват и навигация.

Компромиссы между оптимизацией политики и Q-learning

Основная сила методов оптимизации политики заключается в том, что они «принципиальны» — в том смысле, что вы напрямую оптимизируете то, что вам нужно. Это, как правило, делает их стабильными и надёжными, в то время как методы Q-обучения лишь косвенно оптимизируют работу агента, обучая удовлетворять уравнению самосогласованности. Для такого обучения существует множество способов отказа, поэтому оно, как правило, менее стабильно.

Однако методы Q-обучения имеют то преимущество, что они значительно более эффективны, когда они работают, потому что они могут повторно использовать данные более эффективно, чем методы оптимизации политик. По счастли-

вой случайности, оптимизация политики и Q-обучение не являются несовместимыми (а при некоторых обстоятельствах, как выясняется, эквивалентными), и существует целый ряд алгоритмов, которые находятся между этими двумя крайностями. Подходящие алгоритмы, находящиеся в этом спектре, способны тщательно искать компромисс между сильными и слабыми сторонами.

Примерами могут послужить:

- Deep Deterministic Policy Gradient (DDPG) — алгоритм, который одновременно обучается детерминированной политике и Q-функции, используя каждую из них для улучшения другой;
- Soft Actor Critic (SAC) — вариант, который использует стохастические политики, энтропийную регуляризацию и несколько других методов для стабилизации обучения и получения более высоких результатов, чем DDPG на стандартных эталонах.

Policy optimization

В обучении с подкреплением существует множество различных вариантов того, что аппроксимировать, — политики, функции ценности, модели динамики или их комбинации. Это контрастирует с контролируемым обучением, где обычно изучается отображение от входов к выходам.

В обучении с подкреплением есть два ортогональных выбора: какую цель оптимизировать (политика, функции ценности, модели динамики) и какой тип аппроксимационной функции выбрать.

На самом «верхнем» уровне мы имеем два различных подхода для создания алгоритмов RL: оптимизация политики и динамическое программирование.

Методы оптимизации политики сосредоточены вокруг политики — функции, которая отображает состояние агента к его следующему действию. Эти методы рассматривают обучение с подкреплением как задачу численной оптимизации, где мы оптимизируем ожидаемое вознаграждение в зависимости от параметров политики. Существует два способа оптимизации политики.

Во-первых, это алгоритмы оптимизации без производных (Derivative-Free Optimization, DFO), включая эволюционные алгоритмы. Данные алгоритмы работают путём возмущения параметров политики множеством различных способов, измеряя их производительность, а затем двигаются в направлении хорошей производительности. Они просты в реализации и очень хорошо работают для политик с небольшим количеством параметров, но они плохо масштабируются с увеличением числа параметров.



❖ **Микросети** предлагают надёжное энергообеспечение и возможность использовать другие виды энергии, что помогает снизить расходы на коммунальные услуги. Микросеть, подключённая к более крупной энергосистеме, может работать независимо от неё в полностью автономном режиме, снабжая электроэнергией только своих потребителей.

Некоторые алгоритмы DFO, используемые для оптимизации политики, включают метод перекрёстной энтропии, адаптацию ковариационной матрицы и стратегии естественной эволюции (все три используют распределение Гаусса), а также метод генеративного кодирования Нурег-NEAT, который также задействует эволюцию топологии сети.

Во-вторых, существуют градиентные методы политики. Эти алгоритмы могут оценить направление улучшения политики, используя различные величины, измеренные агентом, причём, в отличие от алгоритмов DFO, им не нужно «возмущать» параметры, чтобы измерить направление улучшения. Градиентные методы политики немного сложнее в реализации, и они испытывают определённые трудности при оптимизации поведения, которое разворачивается в течение очень длительного времени, но они способны оптимизировать гораздо более крупные политики, чем алгоритмы оптимизации без производных DFO.

Второй подход к разработке алгоритмов RL — это адаптивное (приближённое) динамическое программирование (Adaptive Dynamic Programming, ADP). Эти методы сосредоточены на изучении функций ценности, которые предсказывают, сколько вознаграждения получит агент. Истинные функции ценности подчиняются определённым уравнениям согласованности, и алгоритмы ADP работают, пытаясь удовлетворить этим уравнениям. Существуют два известных алгоритма для точного решения задач RL, которые имеют конечное число состояний и действий: итерация политики и итерация значений (оба эти алгоритма являются частными случаями общего алгоритма, называемого модифицированной итерацией политики).

Эти алгоритмы могут быть объединены с аппроксимацией функций различными способами, и в настоящее время ведущие потомки итерации значений работают путём аппроксимации Q-функций.

Стоит также упомянуть, что существуют методы вида «актор — критик» (Actor Critic, AC), которые сочетают в себе элементы как оптимизации политики, так и динамического программирования. Эти методы оптимизируют политику, но они используют функции ценности для ускорения этой оптимизации, и часто задействуют идеи из приближённого динамического программирования для подгонки функций ценности.

Обучение с подкреплением в оптимизации микросетей

Обучение с подкреплением (Reinforcement Learning, RL) является мощным инструментом, который может быть использован для планирования энергосистем в высоко стохастических средах, таких как микросети с возобновляемыми источниками энергии. RL использовалось для планирования режима работы аккумуляторных батарей, диспетчеризации энергии микросети, а также управления HVAC, энергопотреблением в «умных домах», многоуровневыми накопителями энергии, общими накопителями энергии в кластерах зданий и т.д. Было замечено, что эффективность метода на основе RL зависит от того, как сформулирована проблема. Необходимо разработать такую формулировку, которая обеспечит лучшую оптимальность при значительной эффективности.

Одной из основных проблем в применении RL является разработка марковских процессов принятия решений (MDP), которые наилучшим образом представляет

среду оптимизации. При разработке детерминированных MDP стохастичность моделируемой среды игнорируется. Было установлено, что такие допущения отрицательно влияют на точность получаемых решений. Кроме того, детерминированные MDP занимают большой объём памяти. Более того, в случае частично наблюдаемого марковского процесса принятия решений (Partially Observable Markov Decision Process, PO-MDP) трудно обеспечить адекватные наблюдения, которые позволили бы обучающемуся агенту правильно оценить состояния. Современные методы включают в себя аспекты контролируемого обучения, в котором моделируемая среда используется для обучения искусственной нейронной сети (ИНС), лучше приближающих состояния, чем простые цепи Маркова, и занимающих меньше места в памяти, чем MDP.

Обзор литературы показал, что традиционные проблемы, с которыми сталкиваются алгоритмы RL, такие как «проклятие размерности» и «проблема исследования или эксплуатации», были грамотно решены. Было отмечено, что методы, используемые для решения этих проблем, вводят новые проблемы, такие как нестабильность в глубинных нейросетях Q-обучения (Deep Q-Network, DQN) и методы градиента чистой политики. Хотя воспроизведение опыта было применено для улучшения стабильности в методах, основанных на ценности, оно вносит дисперсию в методы, основанные на политике.

В имеющейся литературе отмечается, что разделение сетей политик и сетей функций ценности даёт лучшие результаты. Это разделение является основной причиной успеха алгоритмов Actor Critic (AC). Тот факт, что архитектура AC гибридирует градиенты политики с методами, основанными на ценности, является основной причиной того, что она более надёжна в управлении питанием микросети. Было отмечено, что современные алгоритмы RL так или иначе принимают форму этой архитектуры.

В последнее время некоторые микросети оснащаются новыми технологиями, такими как децентрализованное распределение электроэнергии, распределённая система хранения энергии, интеллектуальное планирование нагрузки и реагирование на спрос в реальном времени, объединение микросетей и т.д. Такие технологии требуют более совершенных алгоритмов управления. Методы обучения с подкреплением были усовершенствованы для адаптации к новым задачам в области управления питанием микросетей.

Мультиагентное обучение с подкреплением (Multi-Agent Reinforcement Learning, MARL) рассматривается как одна из самых мощных технологий обучения для управления электроэнергией в распределённой схеме диспетчеризации и взаимосвязанных микросетях.

Кроме того, с увеличением размерности в сетевых микросетях могут справиться методы оптимизации политики, такие как Trust Region Policy Optimization (TRPO) и Proximal Policy Optimization (PPO), благодаря их способности плавно оптимизировать цели более эффективно в условиях высокой неопределённости, непрерывности и многомерности. Было замечено, что алгоритмы оптимизации также проще в реализации и настройке. Более того, применение трансфертного обучения в методах Deep Reinforcement Learning (DRL) может помочь эффективно перенести успех методов DRL в играх на современные среды управления питанием. А проблемы, такие как уязвимость интеллектуальных сетей к киберфизическим атакам и стихийным бедствиям, могут быть решены с помощью приоритетного воспроизведения опыта. Воспроизведение предыдущих событий с высокой степенью воздействия и низкой вероятностью для агента может лучше адаптировать его к более адекватной реакции на них, когда они повторяются.

Кроме того, для максимизации производительности агента в случаях, когда задача имеет большое сходство с той, которая уже решалась агентом ранее, может быть реализовано обучение с переносом. Методы обучения на основе внутренней мотивации, такие как обучение на основе любопытства и воспроизведение опыта «задним числом», способны уменьшить сложность разработки вознаграждений в сложных системах. Это связано с тем, что эти методы отделяют переходы среды от функций подкрепления.

Основной целью внедрения внутренней мотивации в RL является улучшение масштабируемости алгоритмов за счёт отделения динамики среды от функции вознаграждения. Весьма важно использовать существующие знания о динамике системы для улучшения производительности алгоритмов. Поэтому гибридизация методов внешней и внутренней мотивации может привести к более стабильному обучению в сложных условиях энергосистем.

Подобная гибридная схема вознаграждения может извлечь выгоду как из знаний разработчика о динамике среды, так и из собственного опыта агента в процессе обучения.



По оценке Wood Mackenzie, благодаря растущему спросу на услуги бесперебойного энергоснабжения, целям устойчивого развития крупных корпораций и планам Вооружённых сил США по обеспечению отказоустойчивости, американский рынок микросетей в 2022 году достиг объёма 10 ГВт.

Существующие решения: краткий обзор научных статей

В работе [11] рассматривалась микросеть для распределения энергии с местным потребителем, с ветроэнергетической установкой и накопителем (аккумулятор), подключённым к внешней сети через трансформатор. В ней предлагается алгоритм обучения с опережением в два шага для планирования работы аккумулятора, который играет ключевую роль в достижении целей потребителя. В основе лежит многокритериальное принятие решений отдельным потребителем, который ставит перед собой цели увеличить коэффициент использования батареи во время высокого спроса на электроэнергию (чтобы снизить закупки электроэнергии из внешней сети) и увеличить коэффициент использования ветряной турбины для местного использования (чтобы повысить независимость потребителя от внешней сети).

Прогнозы доступной мощности ветроэлектрических установок подаются в алгоритм обучения с подкреплением для выбора оптимальных действий по планированию батареи. Встроенный механизм обучения позволяет расширить знания потребителя об оптимальных действиях по планированию заряда батареи в различных условиях окружающей среды, зависящих от времени.

Разработанная схема даёт возможность интеллектуальным потребителям изучать стохастическую среду и использовать полученный опыт для выбора оптимальных действий по управлению энергией.

В статье [12] исследовалась эффективность различных алгоритмов глубинного обучения с подкреплением для улучшения системы управления энергией в микросети. Авторами была предложена новая модель микросети, которая состоит из ветрогенератора, системы хранения энергии,

набора термостатически управляемых нагрузок, набора нагрузок, реагирующих на цену, и подключения к основной сети. Предлагаемая система управления энергией предназначена для координации между различными гибкими источниками путём определения приоритетных ресурсов, прямых сигналов управления спросом и цен на электроэнергию. В данной работе были реализованы и эмпирически сравнены семь алгоритмов глубинного обучения с подкреплением. Численные результаты показали, что алгоритмы глубинного обучения с подкреплением сильно различаются по своей способности сходиться к оптимальным политикам. Добавив повторение опыта и полудетерминированную фазу обучения к известному асинхронному алгоритму «актор — критик» с преимуществами, была достигнута наивысшая производительность модели, а также сходимости к почти оптимальным политикам.

В работе [13] поднимается тема использования обучения с подкреплением в реальном времени. Благодаря последним достижениям и применению технологий интеллектуальных сетей, энергосистемы подвергаются радикальной модернизации. Микросеть играет важную роль в процессе модернизации, обеспечивая гибкий способ интеграции распределённых возобновляемых источников энергии в энергосистему. Однако распределённые ВИЭ, такие как солнечная и ветровая энергия, могут быть очень прерывистыми и стохастическими. Эти неопределённые ресурсы в сочетании со спросом на нагрузку приводят к случайным колебаниям как спроса, так и предложения, что затрудняет эффективное управление системой. Сфокусировавшись на этой проблеме, в данной работе предложен новый подход к управлению энергией для пла-

нирования работы микросети в реальном времени с учётом неопределённости спроса на нагрузку, возобновляемой энергии и цены на электроэнергию. В отличие от традиционных подходов, основанных на моделировании и требующих предиктора для оценки неопределённости, предлагаемое решение основано на обучении и не требует явной модели неопределённости. В частности, управление энергией сети моделируется как марковский процесс принятия решений (MDP) с целью минимизации ежедневных эксплуатационных расходов.

Для решения MDP разработан подход глубинного обучения с подкреплением (Deep Reinforcement Learning, DRL). В подходе DRL для аппроксимации оптимальной функции «действие — значение» разработана глубинная нейронная сеть с прямой передачей, а для обучения нейронной сети используется алгоритм глубинной Q-сети (DQN). Предложенный подход принимает в качестве входных данных состояние микросети и выдаёт непосредственно графики генерации в реальном времени. Наконец, используя реальные данные электросетей «Калифорнийского независимого системного оператора» (CAISO, США), были проведены тематические исследования для демонстрации эффективности предложенного подхода.

В последней из рассматриваемых статей [14] авторы пишут о том, что использование крупномасштабной распределённой возобновляемой энергии способствует развитию мульти-микросети (Multi-Microgrid, MMG), а это повышает необходимость разработки эффективного метода управления энергией для минимизации экономических затрат и поддержания собственной энергоэффективности. Мультиагентное глубинное обучение с подкреплением (Multi-Agent Deep Reinforcement Learning, MADRL) широко используется для решения проблемы управления энергией благодаря

В связи с необходимостью снижения затрат на энергоснабжение изолированных объектов и поселений, повышения надёжности энергоснабжения и сокращения выбросов парниковых газов, в автономных энергосистемах изолированных объектов и поселений будет увеличиваться доля ВИЭ

своей способностью планировать работу в режиме реального времени. Однако для его обучения требуются массивные данные о работе микросетей (Microgrid, MG), а сбор этих данных от разных MG ставит под угрозу их конфиденциальность и безопасность данных. Поэтому в данной работе предлагается решение этой сложной проблемы федеративным (federated) мультиагентным алгоритмом глубинного обучения с подкреплением (F-MADRL) с использованием физически обоснованного вознаграждения. В этом алгоритме для обучения алгоритма MADRL используется механизм федеративного обучения Federated Learning (FL), что обеспечивает конфиденциальность и безопасность данных. Кроме того, построена децентрализованная модель MMG, и энергия каждого участвующего MG управляется агентом, целью которого является минимизация экономических затрат и поддержание самодостаточности энергии в соответствии с физически информированным вознаграждением. Сначала MG индивидуально выполняют самообучение на основе местных данных о работе энергосистемы, чтобы обучить свои локальные модели агентов. Затем эти локальные модели периодически загружаются на сервер, а их параметры объединяются для создания глобального агента, который будет передан MG и заменит их локальных агентов. Таким образом, опыт каждого агента MG может быть совместным, а данные о работе энергосистемы не

передаются в явном виде, что позволяет защитить конфиденциальность и обеспечить безопасность данных. Наконец, ведутся эксперименты на тестовой системе микросети лаборатории распределённого управления энергией (ORNL-MG) Ок-Риджской национальной лаборатории (Oak Ridge National Laboratory, ORNL) в США, в которых проводится сравнение для проверки эффективности внедрения механизма FL и превосходства предложенного F-MADRL.

Выводы

В силу больших территорий Российской Федерации при их неравномерном заселении проблема автономного энергоснабжения изолированных объектов и поселений играет значительную роль в развитии нашего государства. В связи с необходимостью снижения затрат на энергоснабжение изолированных объектов и поселений, повышения надёжности энергоснабжения и сокращения выбросов парниковых газов в автономных энергосистемах изолированных объектов и поселений будет увеличиваться доля ВИЭ. Это делает энергосистемы более волатильными и менее предсказуемыми и требует дополнительных механизмов снижения неопределённости их работы. Одним из таких механизмов, повышающих надёжность работы энергосистемы, является использование алгоритмов оптимизации на краткосрочном и среднесрочном горизонте планирования.

В настоящий момент компаниями «Смартрен» и «Дельта П» ведётся работа по разработке и внедрению аппарата прогнозирования для автономных энергосистем. Создан и испытан пилотный аппаратный метеорологический комплекс, ведутся совместные тесты и работы по оптимизации энергосистем и прогнозированию выработки ветровой и солнечной энергии на собственном стенде и для компаний, генерирующих электроэнергию ВИЭ по программе ДПМ. ●

1. Объекты генерации в изолированных и труднодоступных территориях в России: Доклад АЦ при Правительстве РФ. — М., 2020. 78 р.
2. Дальний Восток и регионы Арктики получают инвестиционную карту объектов промышленной распределённой генерации [Электр. текст]. РЭА Минэнерго России от 07.07.2023. Режим доступа: rosenargo.gov.ru. Дата обращ.: 08.07.2023.
3. Khan A.A., Naeem M., Iqbal M., Qaisar S., Anpalagan A. A compendium of optimization objectives, constraints, tools and algorithms for energy management in microgrids. Renewable and Sustainable Energy Reviews. 2016.
4. Wouters C., Fraga E., James A. MILP approach for the design of residential microgrids with energy interactions restrictions. Computer Aided Chemical Engineering. 2015. Vol. 37. Pp. 2357–2362.
5. Delgado C., Dominguez-Navarro J.A. Optimal design of a hybrid renewable energy system. Proc. of the 9th International Conference on Ecological Vehicles and Renew-

- able Energies (EVER). Monte-Carlo, Monaco. March 25–27, 2014.
6. Khan M.R.B., Jidin R., Pasupuleti J. Multi-agent based distributed control architecture for microgrid energy management and optimization. Energy Conversion and Management. 2016. Vol. 112. Pp. 288–307.
7. Khan M.W., Jie W. The research on multi-agent system for microgrid control and optimization. Renewable and Sustainable Energy Review. 2017. Vol. 80. Pp. 1399–1411.
8. Chouikhi S., Merghem-Boulahia L., Esseghir M. Energy demand scheduling based on game theory for microgrids. Proc. of the 2018 IEEE International Conference on Communications (ICC). May 20–24, 2018. Kansas City, KCMO, US.
9. Tavakoli A., Saha S., Arif M.T. Haque M.E., Mendis N. Impacts of grid integration of solar PV and electric vehicle on grid stability, power quality and energy economics: A review. The Institution of Engineering and Technology (IET) Digital Library. 2020. Vol. 2. Issue 3. Pp. 243–260.

10. Nair D.R., Nair M.G., Thakur T. A smart microgrid system with artificial intelligence for power-sharing and power quality improvement. Energies. 2022. Vol. 15. Issue 15.
11. Kuznetsova E.L., Li Y.-F., Ruiz C., Zio E., Ault G., Bell K. Reinforcement learning for microgrid energy management. Energy. 2013. Vol. 59. Pp. 133–146.
12. Nakabi T.A., Toivanen P. Deep reinforcement learning for energy management in a microgrid with flexible demand. Sustainable Energy, Grids and Networks. 2021. Vol. 25.
13. Ji Y., Wang J., Xu J., Fang X., Zhang H. Real-time energy management of a microgrid using deep reinforcement learning. Energies. 2019. Vol. 12. Issue 12.
14. Li Y., He S., Li Y., Shi Y., Zeng Z. Federated multi-agent deep reinforcement learning approach via physics-informed reward for multi-microgrid energy management. IEEE Transactions on Neural Networks and Learning Systems. January 3, 2023. 12 p.

References — see page 78.

Application of algorithms to ensure the reliability of power systems. Pp. 66–73.

Arsenij I. Velichko, general director of “Smartren”, LLC; **Mikhail S. Povernov**, technical director of “Delta P”, LLC; **Nikolaj Shuvalov**, Master’s Degree, [Peter the Great St. Petersburg Polytechnic University \(SPbPU\)](#)

1. *Ob#ekty generatsii v izolirovannykh i trudnodostupnykh territoriyakh v Rossii: Analiticheskij doklad ATs pri Pravitel#stve RF* [Generation facilities in isolated and hard-to-reach areas in Russia: The Report of the Analytical Center under the Government of the Russian Federation]. Moscow. 2020. 78 p. [In Russian]
2. *Dal’nij Vostok i regiony Arktiki poluchat investitsionnuju kartu ob#ektov promyshlennoj raspredelennoj generatsii* [The Far East of Russia and the Russian regions of the Arctic will receive an investment map of industrial distributed generation facilities]. *RJeA Minjenergo Rossii ot 07.07.2023* [Russian Energy Agency of the Ministry of Energy of the Russian Federation of July 7, 2023]. Web-source: rosenergo.gov.ru. Access date: July 8, 2023. [In Russian]
3. A.A. Khan, M. Naeem, M. Iqbal, S. Qaisar, A. Anpalagan. A compendium of optimization objectives, constraints, tools and algorithms for energy management in micro-grids. *Renewable and Sustainable Energy Reviews*. 2016.
4. C. Wouters, E. Fraga, A. James. MILP approach for the design of residential microgrids with energy interactions restrictions. *Computer Aided Chemical Engineering*. 2015. Vol. 37. Pp. 2357–2362.
5. C. Delgado, J.A. Domínguez-Navarro. Optimal design of a hybrid renewable energy system. *Proc. of the 9th International Conference on Ecological Vehicles and Renewable Energies (EVER)*. Monte-Carlo, Monaco. March 25–27, 2014.
6. M.R.B. Khan, R. Jidin, J. Pasupuleti. Multi-agent based distributed control architecture for microgrid energy management and optimization. *Energy Conversion and Management*. 2016. Vol. 112. Pp. 288–307.
7. M.W. Khan, W. Jie. The research on multi-agent system for microgrid control and optimization. *Renewable and Sustainable Energy Review*. 2017. Vol. 80. Pp. 1399–1411.
8. S. Chouikhi, L. Merghem-Boulahia, M. Esseghir. Energy demand scheduling based on game theory for microgrids. *Proc. of the 2018 IEEE International Conference on Communications (ICC)*. May 20–24, 2018. Kansas City, KCMO, USA.
9. A. Tavakoli, S. Saha, M.T. Arif, M.E. Haque, N. Mendis. Impacts of grid integration of solar PV and electric vehicle on grid stability, power quality and energy economics: A review. *The Institution of Engineering and Technology (IET) Digital Library*. 2020. Vol. 2. Issue 3. Pp. 243–260.
10. D.R. Nair, M.G. Nair, T. Thakur. A smart microgrid system with artificial intelligence for power-sharing and power quality improvement. *Energies*. 2022. Vol. 15. Issue 15.
11. El. Kuznetsova, Y.-F. Li, C. Ruiz, E. Zio, G. Ault, K. Bell. Reinforcement learning for microgrid energy management. *Energy*. 2013. Vol. 59. Pp. 133–146.
12. T.A. Nakabi, P. Toivanen. Deep reinforcement learning for energy management in a microgrid with flexible demand. *Sustainable Energy, Grids and Networks*. 2021. Vol. 25.
13. Y. Ji, J. Wang, J. Xu, X. Fang, H. Zhang. Real-time energy management of a microgrid using deep reinforcement learning. *Energies*. 2019. Vol. 12. Issue 12.
14. Y. Li, S. He, Y. Li, Y. Shi, Z. Zeng. Federated multi-agent deep reinforcement learning approach via physics-informed reward for multi-microgrid energy management. *IEEE Transactions on Neural Networks and Learning Systems*. January 3, 2023. 12 p.