ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2020, № 6, с. 132–151

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

УДК 621.311

КОНЦЕПЦИЯ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ "ИСКУССТВЕННЫЙ ДИСПЕТЧЕР" ДЛЯ АВТОМАТИЧЕСКОЙ СИСТЕМЫ УПРАВЛЕНИЯ ЭЛЕКТРИЧЕСКИМИ СЕТЯМИ НА БАЗЕ ГЛУБОКОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ[1]

© 2020 г. Н. В. Томин

ИСЭМ СО РАН, Иркутск, Россия

e-mail: tomin.nv@gmail.com

Поступила в редакцию 29.08.2018 г.
После доработки 13.03.2020 г.
Принята к публикации 30.03.2020 г.

Предложена концепция автономной интеллектуальной системы управления "Искусственный диспетчер" для интеграции в современные автоматические системы управления электрическими сетями с целью повышения эффективности управления режимами электроэнергетических систем. Данная интеллектуальная система управления реализуется на базе технологии глубокого машинного обучения с подкреплением при совместном использовании метода Монте-Карло для поиска в дереве и глубоких искусственных нейронных сетей. Показано, что эффективное обучение интеллектуальной системы управления "Искусственный диспетчер" достижимо без применения больших баз данных по схемно-режимным ситуациям и экспертного опыта управления режимами электроэнергетических систем за счет инновационного алгоритма самостоятельной игры. Приведены расчетные примеры использования агентов, создаваемых на основе концепции интеллектуальной системы управления "Искусственный диспетчер", в задачах автоматизации технологическими процессами подстанций (управление регуляторами напряжения и реактивной мощности) и промышленных предприятий (управление серводвигателем постоянного тока).

DOI: 10.31857/S0002338820050121

Введение. Эффективное управление современными электроэнергетическими системами (ЭЭС) реализуется посредством организации машинного сопровождения действий диспетчера, координацией действий автоматики и работы оборудования ЭЭС на базе различных автоматизированных систем управления (АСУ) (англ. "industrial control system"), прежде всего с использованием АСУ технологическими процессами (АСУ ТП) и автоматизированных систем диспетчерского управления (АСДУ) [1]. Передача ряда функций управления машине позволяет ускорить процесс принятия решений, освободить персонал от рутинной работы, избежать ошибок при выработке управляющих воздействий (УВ). В настоящий момент на энергорынке представлены различные интегрированные АСУ для разных уровней напряжения электрических сетей, включающий программные модули для автоматизации различных режимных операций: мониторинг состояний, оптимизация и ведение режима, отключение оборудование, удаленное управление и пр. Такие системы объединены с традиционными комплексами сбора информации (типа оперативно-информационный комплекс/SCADA), а также различными геоинформационными системами (типа WAMS, WACS), что в итоге позволяет реализовывать целый комплекс согласованных решений по управлению электрическими сетями ЭЭС [2].

Однако все возрастающее насыщение современных ЭЭС новыми стохастическими компонентами (ветротурбины большой мощности, солнечные панели, системы управления спросом и пр.), внедрением новых технологий (распределенная генерация, оборудование гибких электропередач, системы накопления энергии, микро-ЭЭС, гибридные сети переменного/постоянного тока, цифровые подстанции и т.п.), а также действием рыночных принципов регулирова-

[1] Работа выполнена в рамках программы фундаментальных исследований СО РАН, рег. № АААА-А17-117030310438-1, научный проект III.17.4.2.

ния существенно усложняются задачи управления режимами современных ЭЭС [3]. В этих условиях существующие АСУ, как правило, использующие традиционные алгоритмы управления и оптимизации, не всегда могут обеспечить оптимальное и/или надежное регулирование режимов ЭЭС. Кроме того, ряд операций, выполняемых диспетчером по сей день, автоматизированы лишь частично, а отдельные системы автоматизации диспетчерского пункта чаще всего никак не интегрированы. В итоге на диспетчера ложится задача объединения информационных потоков всех систем, сопоставления хронологии событий, анализ и выработка УВ. Все это ведет к увеличению финансовых убытков энергокомпаний, а также повышает риски возникновения опасных и аварийных режимов в ЭЭС.

В работе предлагается подход к совершенствованию современных АСУ с применением технологии машинного обучения (МО) путем интеграции "умных" приложений в уже существующие программные платформы АСУ в качестве дополнительного базового модуля с использованием их функционала (расчет, анализ режимов, сбор/передача данных, телеуправление и пр.) для выработки и реализации УВ. Для реализации такого подхода в статье предложена концепция инновационного инструмента – интеллектуальной системы управления "Искусственный диспетчер" (ИСУ ИД) на базе моделей глубокого МО с подкреплением. ИСУ ИД будет представлять машинный интеллект, который во многих случаях будет способен заменить реального диспетчера по принципу "автопилота" с целью повышения эффективности управления ЭЭС.

1. Постановка задачи. Круг задач управления режимами ЭЭС можно разделить на задачи оперативно-диспетчерского (ОДУ) и противоаварийного управления (ПАУ). ОДУ осуществляется диспетчерами системного оператора. Одной из основных целей ОДУ является управление нормальными и послеаварийными режимами с целью предотвращения возникновения аварийного режима. В свою очередь аварийный режим ликвидируется главным образом автоматическими системами ПАУ – релейной защитой, различными локальными и централизованными устройствами противоаварийной автоматики, основной задачей которых является обеспечение перехода к послеаварийному режиму. Общий комплекс задач ОДУ и ПАУ в зависимости от режимов работы ЭЭС может быть проиллюстрирован диаграммой рис. 1. При этом данные задачи ОДУ и ПАУ лежат в своих определенных временных диапазонах (рис. 2). Все состояния ЭЭС могут быть определены ограничениями в форме равенства и неравенства, которые могут быть нарушены или не нарушены. Ограничения в виде равенства выражают баланс нагрузки и генерации, в то время как ограничения в виде неравенства выражают физические ограничения отдельных элементов ЭЭС, к примеру силового оборудования.

Автоматизация процессов управления решается применением различных типов специализированных АСУ, представляющих собой многоуровневые человеко-машинные системы управления. Отдельным направлением совершенствования ОДУ стало развитие "советчиков-диспетчеров", т.е. особых подсистем АСУ для поддержки принятия решений, помогающих диспетчеру ЭЭС эффективно управлять энергосистемой в условиях неопределенности и малого резерва времени. Исторически успехи в этом направлении были достигнуты при использовании теории автоматического управления, теории конечных автоматов, методов ситуационного управления, теории распознавания образов и полиномиальной аппроксимации, теории нечетких множеств [4–8].

Однако многие разработанные виды АСУ, такие, как АСУ ТП, АСДУ и пр., не оснащены программно-аналитическими модулями статистического управления процессами[2], а оперативный персонал, в том числе диспетчеры, далеко не всегда обучены этим методам управления. Кроме того, при реализации в АСУ классических экспертных систем для приложений советчика-диспетчера, например системы MIMIR-3 [9], выработка решений часто основана на относительно простых продукционных правилах, полученных с применением специально создаваемой базы знаний. В результате такие системы не всегда способны эффективно адаптироваться к многообразию видов электрической сети и условий их эксплуатации, особенно при появлении новых элементов системы.

Важно также отметить, что системы поддержки решений в АСУ имеют целый ряд ограничений, прежде всего, по времени вычислений, что обусловлено необходимостью постоянного решения сложных оптимизационных задач для поиска верных решений в виде советов диспетчеру и/или автоматических УВ. Обычно подобные системы используют традиционные методы анализа надежности, оптимизации режимов, основанные на алгоритмически "жестких" моделях с

[2] Statistical process control (англ.) – метод мониторинга производственного процесса с использованием статистических инструментов с целью управления качеством продукции "непосредственно в процессе производства".

Рис. 1. Режимы работы (состояния) ЭЭС и их переходы в зависимости от возмущений

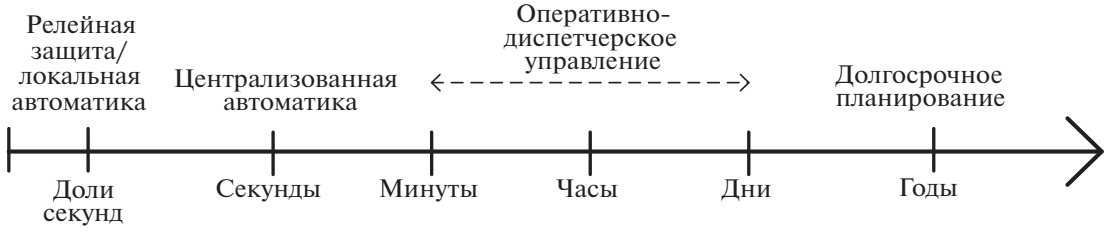

Рис. 2. Временная шкала задач ОДУ и ПАУ режимами энергосистем

недостаточно высокой робастностью и быстродействием. Например, АСУ "Network Manager" (компания "ABB") включает в себя функцию оптимизации реактивной мощности (volt-vAR optimization – VVO) [10] для автоматического поиска оптимальных УВ для достижения определенных целей при сохранении приемлемого напряжения и нагрузки. Однако для больших схем электрических сетей сложной структуры расчеты в рамках приложения VVO могут занимать достаточно большое время, что затрудняет использование АСУ в реальном времени.

В итоге отсутствие эффективных модулей поддержки принятия решений (или невозможность полагаться на них в реальном времени) приводит к тому, что диспетчеры в попытках управлять процессами постоянно совершают ошибки первого и второго рода, не понимая, например, как различить естественную вариабельность статистически управляемого процесса от неестественной (выход процесса из-под контроля под действием особых причин) [11]. Традиционное применение классического автоматического регулирования в АСУ пропорционально-интегрально-дифференцирующих (ПИД) регуляторов игнорирует знания о вариабельности любых изменений параметров режима ЭЭС и фактически не решает проблему ошибок первого и второго рода. Незнание, как минимизировать риск совершения ошибок первого и второго рода, приводит к

потерям, а иногда к серьезным последствиям. Примером может служить известная авария на Саяно-Шушенской ГЭС, когда при работающей АСУ ТП в критической ситуации не было предпринято своевременных действий.

Целью настоящего исследования является разработка концепции принципиально нового инструмента управления электрическими сетями ЭЭС, в том числе его алгоритмической основы, который позволит в рамках существующих АСУ объединить функции автоматического и оперативного управления режимами энергосистем, нивелировав тем самым слабые стороны существующей автоматики и действий человека-оператора (диспетчера ЭЭС). Для реализации этой задачи в статье предлагается использование современных технологий МО, а именно аппарата глубоких искусственных нейронных сетей (ИНС) и подходов на основе обучения с подкреплением.

2. Технология машинного обучения в задачах АСУ в электроэнергетике. Развитие методов искусственного интеллекта (ИИ) дало возможность существенно ускорить и автоматизировать решение целого комплекса задач для АСУ при управлении ЭЭС [12]. Одним из активно развиваемых направлений здесь является применение и внедрение технологии МО, включающей методы построения алгоритмов, способных обучаться. Применение различных видов обучения моделей МО: с учителем (англ. supervised learning), без учителя (англ. unsupervised learning), с подкреплением (англ. reinforcement learning), глубокое обучение (англ. deep learning) и др., позволило создавать отдельные адаптивные, обучаемые программные модули регулирования и управления как отдельными компонентами ЭЭС, так и ее режимом в целом. Их основными преимуществами стали – быстродействие, высокая адаптивность, способность к аппроксимации нелинейных функций и наличие определенного рода машинного интеллекта, что позволяет разрабатывать максимально автономные системы, способные к самостоятельному принятию решений на основе опыта и оригинальных свойств к обобщению.

Например, отдельные модели МО с учителем, такие, как ИНС, деревья принятия решений, получили развитие при разработке различных функций советчика-диспетчера, в частности, в АСУ “EMS-SCADA Hydro-Quebec” [13], энергокомпании “Energinet.dk” [14]. В этом случае задача оценки и управления устойчивостью системы решается через обучение моделей МО распознавать характерные индикаторы надежности ЭЭС, к примеру, коэффициенты чувствительности матрицы Якоби установившихся режимов [4]. При этом происходит трансформация классической оптимизационной проблемы в задачу восстановления регрессии или классификации, что позволяет значительно сократить время расчета с сохранением приемлемой точности [15].

Переход современных ЭЭС в эпоху больших объемов данных (англ. big data) открывает для этого широкие возможности для применения методов глубокого МО, прежде всего глубоких ИНС. Например, компания “Siemens” уже сейчас использует глубокие ИНС в различных проектах: улучшение работы многофункциональных устройств релейной защиты SIPROTEC [16], оптимизация работы газовых турбин для снижения выбросов токсичных оксидов азота [17], интеллектуальное регулирование положения роторов ветряных турбин в зависимости от направления ветра [18]. В другом крупном проекте [19] авторы показали, что глубокие ИНС позволяют прогнозировать электропотребление в крупных энергорайонах с высокой скоростью (несколько миллисекунд) и со значительно низкими ошибками. Этот факт дает возможность использовать такие модели для автоматического управления электропотреблением в рамках интеллектуальных АСУ.

В табл. 1 представлены некоторые задачи управления режимами ЭЭС, которые успешно решаются на базе МО, в том числе в сочетании с другими технологиями ИИ.

3. Машинное обучение с подкреплением при разработке автономных интеллектуальных систем управления ЭЭС. Значимые успехи в задачах управления сложными системами были получены на основе группы методов МО с подкреплением, таких, как методы Монте-Карло (управление по методу Монте-Карло (МКУ), Монте-Карло для поиска в дереве (МКПД)), динамическое программирование (ДП), обучение на основе временных различий (ВР) (SARSA, Q-обучение) и др. [30]. Эти методы подразумевают обучение тому, что надо делать, как следует отображать ситуации в действия, чтобы максимизировать некоторый сигнал поощрения (вознаграждения), принимающий числовые значения. Обучаемой модели (агенту) не говорят, какое действие следует предпринять, как это имеет место в большинстве подходов МО. Вместо этого они, пробуя выполнять различные действия, должны найти, какие из них принесут ему наибольшее вознаграждение.

При этом различные методы МО с подкреплением дают возможность агенту выбирать стратегию управления ЭЭС или ее отдельными компонентами не случайным образом, а учитывать

Таблица 1. Применение различных технологии МО в управлении режимами современных ЭЭС

Задача	Тип управления	Технология ИИ, ссылки
Управление возобновляемой генерацией	Управление в нормальных режимах	МО с подкреплением [20], МО с учителем, глубокое обучение [21]
Управление изолированными ЭЭС и микрогридами	Управление в нормальных режимах	Глубокое МО с подкреплением [22]
Регулирование напряжения и реактивной мощности	Управление в нормальных режимах	МО с учителем [23] и без учителя [24], МО с подкреплением [25]
Предотвращение крупных аварий	Противоаварийное управление	МО с подкреплением [26]
Восстановление энергосистемы	Управление в восстановительном режиме	МО с подкреплением [27]
Управление энергопотреблением	Управление в нормальных режимах	Глубокое МО [19]
Управление пиковой нагрузкой и перегрузками	Противоаварийное управление	МО с подкреплением [28]
Управление в переходных режимах	Противоаварийное управление	МО с подкреплением [29], МО с учителем [13, 14]

опыт предыдущего взаимодействия с окружением (средой) на основе оценки функции ценности состояния (англ. value function) (рис. 3). Таким образом, агент должен действовать в окружении, чтобы максимизировать некоторый долговременный выигрыш. Окружение обычно формулируется как марковский процесс принятия решений (МППР) с конечным множеством состояний.

Формально простейшая модель МО с подкреплением состоит из: множества состояний окружения S; множества действий A и множества вещественнозначных скалярных “выигрышей”. В произвольный момент времени t агент характеризуется состоянием $s_t \in S$ и множеством возможных действий $a \in A(s_t)$, он переходит в состояние s_{t+1} и получает выигрыш r_t. Основываясь на таком взаимодействии с окружающей средой, агент, обучающийся с подкреплением, должен выработать стратегию $\pi : S \times A \to [0, 1]$ (англ. policy function), где $\pi(s, a)$ – вероятность выбора действия $a \in A(s_t)$ в состоянии s. Данная стратегия максимизирует величину $R = r_0 + r_1 + \cdots + r_n$ в ММПР, имеющего терминальное состояние, или величину $R = \sum_t \gamma^t r_t$ для МППР без терминальных состояний (где $0 \le \gamma \ll 1$ – дисконтирующий множитель для “предстоящего выигрыша”).

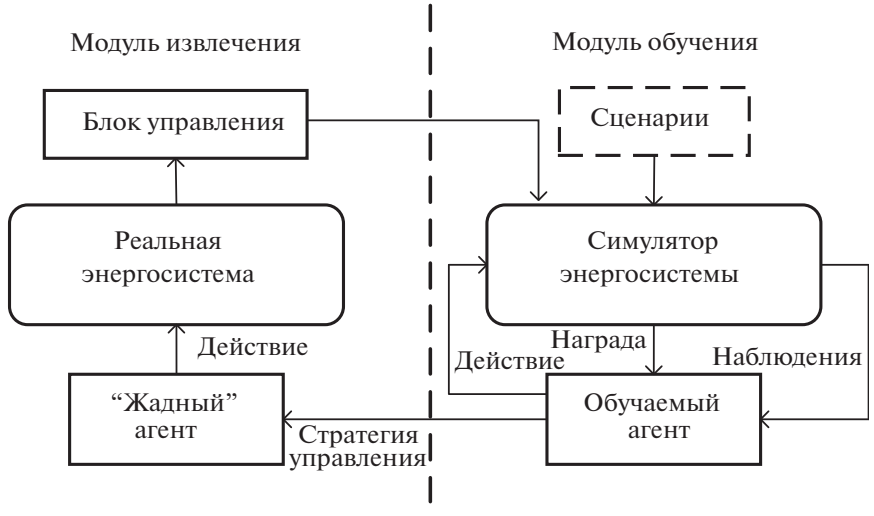

Рис. 3. Общая схема управления режимами работы ЭЭС с использованием МО с подкреплением (адаптировано из [31])

Когда определена функция выигрыша, нужно определить алгоритм, который будет использоваться для нахождения стратегии, обеспечивающей наилучший результат. Подход с применением функции полезности Q имеет множество оценок ожидаемого выигрыша только для одной стратегии π (либо текущей, либо оптимальной). При этом пытаются оценить либо ожидаемый выигрыш, начиная с состояния s, при дальнейшем следовании стратегии π на основе функции ценности состояния $V(s) = E[R \mid s, \pi]$, либо ожидаемый выигрыш при принятии решения a в состоянии s и дальнейшем соблюдении π на основе функции[3] состояния-действия $Q(s, a) = E[R \mid s, \pi, a]$. В случае выбора оптимальной стратегии используется функция полезности Q, оптимальные действия всегда можно выбрать как действия, максимизирующие полезность.

Проблема состоит в том, что нельзя только применить уже проверенные действия или только искать новые эффективные действия, поскольку это ведет к провалу попыток решения задачи. Определенные успехи для решения этой проблемы достигнуты с помощью методов Монте-Карло, который используется для нахождения стратегии управления, π в соответствии с идеей итерации по стратегиям (англ. general policy iteration). Она заключается в том, что стратегию π всегда можно улучшить относительно функции V, а функцию V – относительно функции стратегии π. Эти два вида изменения в некоторой степени воздействуют друг на друга, так как каждый создает движущуюся цель для другой, но вместе они заставляют как функции π и V приближаться к оптимальности.

Совершенствование стратегий π в МКУ часто заключается просто в вычислении “жадной стратегии”, которая в каждом состоянии выбирает наилучшее действие s в соответствии с $Q^{\pi}(s, a)$. В итоге выполняются чередующиеся полные шаги оценки стратегии E, когда $Q(s, a) \to \to Q^{\pi}(s, a)$, и улучшения стратегии I, когда для каждого действия s детерминистически выбирает действие с максимальным значением $Q(s, a)$ $\pi_{i+1}(s) = \arg\max_a Q_i(s, a)$ [30]. Процедура МКУ начинается с произвольной стратегии π_0 и в результате взаимного воздействия двух изменений E и I приводит к оптимальной стратегией π^* и оптимальной функцией изучения-применения Q^*:

$$\pi_0 \xrightarrow{E} Q^{\pi_0} \xrightarrow{I} \pi_1 \xrightarrow{E} Q^{\pi_1} \xrightarrow{I} \pi_2 \xrightarrow{E} \cdots \xrightarrow{I} \pi^* \xrightarrow{E} Q^*. \tag{3.1}$$

Сходимость итеративного процесса (3.1) гарантируется теоремой улучшения стратегии [30], и остается только проверять, что новая стратегия будет одинаково лучше, когда мы имеем для всех π_k, π_{k+1} и $s \in S$:

$$Q^{\pi_k}(s, \pi_{k+1}(s)) = Q^{\pi_k}(s, \arg\max_a Q^{\pi_k}(s, a)), \tag{3.2}$$

$$Q^{\pi_k}(s, \pi_{k+1}(s)) = \max_d Q^{\pi_k}(s, a),$$

$$Q^{\pi_k}(s, \pi_{k+1}(s)) \geq Q^{\pi_k}(s, \pi_k(s)),$$

$$Q^{\pi_k}(s, \pi_{k+1}(s)) = V^{\pi_k}(s).$$

В итоге, согласно теореме улучшения стратегии, $Q^{\pi_{i+1}} \geq Q^{\pi_i}$.

Методы временного различия (ВР) (англ. TD-методы – temporal difference) совмещают в себе идеи методов Монте-Карло и ДП. Как и методы Монте-Карло, ВР-методы могут обучаться без модели динамики окружения, т.е. непосредственно из опыта. Преимущество ВР-методов над МКУ заключается в том, что первые являются интерактивными, полностью инкрементными методами. Если в МКУ необходимо дождаться завершения эпизода, чтобы узнать награду, то в ВР-методах необходимо дождаться лишь следующего временного шага.

Исследования показывают, что агенты, обучаемые офлайн на основе методов МО с подкреплением, успешно контролируют отдельные компоненты ЭЭС при управлении нормальными и аварийными режимами [32]. Например в [33], глобальный регулятор возбуждения генератора WASCCO, реализованный на базе МО с подкреплением для демпфирования электромеханиче-

[3] Ввиду того, что методы МО с подкреплением основаны на теории МППР, когда стратегия управления оптимальна, если она достигает наилучшего (максимального) ожидаемого выигрыша из любого состояния [30], то исходные статистические распределения для функций $V(s)$ или $Q(s, a)$ не играют никакой роли. Хотя формально для неизвестного начального распределения функция ценности состояния для максимизации была бы следующей: $V(s) = E[R(s_0, a_0) + \gamma R(s_1, a_1) + \ldots \mid \pi]$ [31].

ских колебаний и управления напряжением переходного процесса, позволяет системе возбуждения “учиться” при взаимодействии с ЭЭС и эффективно прогнозировать свои будущие состояния. Испытания показали, что такой подход дает лучшее демпфирование и переходную характеристику, чем традиционные системные стабилизаторы, и позволяет гасить межобластные колебания режима лучше, чем регулятор на основе эвристического программирования, действующий самостоятельно. Хорошие результаты также были получены, когда Q-агентами выступали динамический тормоз [34], тиристорно-управляемый последовательный конденсатор [35], синхронные генераторы, индивидуальные или агрегированные нагрузки [36] и пр. для реализации оптимальных стратегий управления. Более того применение МО с подкреплением показали хорошие результаты в целом спектре задач ОДУ и ПАУ (табл. 1).

Согласно разработкам последних лет, использование перспективной технологи глубокого МО позволяет создавать полностью автономные ИСУ. Глубокое МО представляет собой часть более широкого семейства методов МО – обучения представлениям, где векторы признаков располагаются сразу на множестве уровней. Как правило, речь идет о специальных ИНС с множеством скрытых слоев нейронов (уровней), в основе которых заложены специальные строительные блоки, к примеру ограниченные машины Больцмана, позволяющие провести предобучение сети, обучая каждый слой отдельно. Однако именно на базе соединения методов МО с подкреплением и глубоких ИНС были созданы технологии, ставшие прорывами последнего времени в области ИИ, такие, как AlphaGo [37], Atari Deep Q-learning [38] и др. Эти разработки показали, что такие модели способны решать задачи лучше самого человека (эксперта в предметной области) или полностью вытеснить классический, традиционный алгоритм решения задачи. По оценкам некоторых экспертов именно эта технология позволит в ближайшем будущем создавать полноценный машинный интеллект, подобный человеческому [39].

Основная идея здесь заключалась в том, чтобы использовать глубокие ИНС для представления так называемой Q-сети (англ. Q-network) и обучать эту сеть для прогнозирования общей награды [38]. В основе подхода лежит алгоритм Q-обучения, реализующий итерационное приближение функции Q посредством обучения по временным различиям, когда на каждом шаге мы стремимся минимизировать среднюю квадратичную ошибку между предиктором $Q(s, a)$ и целью $r + \gamma \max_{a'} Q(s', a')$. В итоге, так называемое табличное Q-обучение предполагает формирование таблицы, содержащей как старые, так и предварительно обновленные оценки функции Q, используя следующее правило обновления:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma \max_{a'} Q(s',a') VQ(s,a)), \tag{3.3}$$

где s' – конечное состояние после применения действия a для состояния s, r – непосредственное вознаграждение, полученное для действия a в состоянии s, γ – коэффициент дисконтирования, а α – скорость обучения.

Когда число состояний велико, сохранение таблицы поиска со всеми возможными значениями пар действия-состояния нецелесообразно. В [38] было предложено общее решение этой проблемы использованием параметризированной функции аппроксимации Θ, так чтобы $Q(s,a) \approx Q(s,a;\Theta)$. В качестве аппроксиматора было предложено применять глубокую нейросеть. В этом случае получаемая Q-сеть будет обучаться “приблизиться” к оптимальной функции $Q^*(s,a)$, минимизируя в общем варианте следующее выражение:

$$(r + \gamma \max_{a'} Q(s',a';\Theta) - Q(s,a;\Theta))^2. \tag{3.4}$$

В электроэнергетике подобная технология стала находить применение в области разработки автономных онлайн-систем управления различными параметрами режима гибридных ЭЭС. Например, в [22] рекуррентная ИНС с подкреплением была использована для поиска оптимальной стратегии управления $Q^*(s,a)$ зарядами водородных накопителей энергии в автономной микро-ЭЭС в Бельгии, содержащей источники солнечной генерации. В [19] авторы применяют глубокие ИНС с подкреплением для онлайн-оптимизации графиков управления электропотребления зданий, что позволяет находить оптимальные стратегии планирования потребления электрической энергии в реальном времени.

Несмотря на все преимущества, глубокие Q-сети не позволяют выполнять поиск действий (планирование) в более долгосрочной перспективе, когда необходимо выполнить оценку влияния каждого конкретного УВ на состояние системы в будущем. Однако решение этой задачи легко достижимо с помощью метода МКПД, который осуществляет глубокий поиск решений в

рамках концепции МО с подкреплением [40]. Феноменальный успех системы AlphaGo наглядно демонстрирует, что объединение технологии МКПД с глубокими ИНС позволяет разрабатывать системы машинного интеллекта для эффективного управления сложными системами, к которым относятся ЭЭС. Например, в 2017 г. началось сотрудничество компаний “Google DeepMind” с крупной британской энергокомпанией “National Grid UK” для применения системы AlphaGo в задачах управления ЭЭС Великобритании.

4. Концепция ИСУ ИД. В настоящей статье была разработана концепция интеллектуального инструмента нового типа – прототипа ИСУ ИД, который позволил объединить функции оперативного и автоматического управления режимами ЭЭС на основе технологии глубокого МО с подкреплением. Ключевым аспектом этой технологии является имитация человеческого принятия решений с существенно более высокой скоростью и точностью, что достигается оригинальными свойствами моделей глубокого МО с подкреплением и возможностью оперировать большими схемно-режимными ситуациями, которые моделируются методом МКПД, что недоступно человеку в силу ограничений его памяти. Предполагается, что создание системы полноценного “машинного интеллекта” позволит в большинстве случаев заменить диспетчера ЭЭС по принципу “автопилота”.

Предполагается разработать ИСУ ИД таким образом, чтобы иметь возможность интегрироваться в уже существующие программные платформы АСУ электрическими сетями, прежде всего в платформу СК-11 (ЗАО “Монитор-Электрик”) [41], которая является открытой для интеграции приложений различных производителей на базе международных протоколов, например, технология OPC Unified Architecture (IEC 62541). В этом случае ИСУ ИД будет рассматриваться как внешнее программное приложение, использующее функционал эксплуатируемой АСУ для выработки и реализаций УВ.

А л г о р и т м 1. При разработке любых обучаемых систем мы сталкиваемся с задачей формирования специальных выборок данных, содержащих возможные примеры поведения объекта. Особенно это актуально для моделей глубокого МО, когда для достижения высокой эффективности модели требуются большие базы данных (ББД) для обучения. В первоначальной концепции ИСУ ИД [42] мы рассматривали отдельную задачу формирования ББД, включающих как результаты моделирования, так и фактический экспертный опыт диспетчеров в области ОДУ. Предполагалось использовать три основных ББД: моделирование, реальное поведение и фактический опыт диспетчера. Однако, как показывает практика, такие данные могут быть слишком дорогими, ненадежными или просто недоступными, особенно с учетом коммерческой тайны.

Как отмечалось выше, последние исследования в области глубокого МО с подкреплением показывают, что высокая эффективность (в том числе и сверхчеловеческая[4]) обучаемых систем управления достижима без ББД и фактических экспертных знаний человека. Поэтому за основу разработки ИСУ ИД предлагается использование алгоритма самостоятельной игры (АСИ) (англ. self-play algorithm), основанный на новом методе глубокого МО с подкреплением разработанной компанией “Google DeepMind” для системы AlphaGo Zero [37]. Исследование показывает, что совместное применение этого алгоритма и метода МКПД позволяет достичь эффективности выше человеческой.

Метод МКПД заключается в формировании дерева состояний (например, режимов работы всей ЭЭС или ее отдельных компонент). Задача АСИ – выбрать наиболее выигрышный вариант развития событий. Дерево представляет из себя структуру, в которой помимо хода и указателей есть количество сыгранных и количество выигранных партий. На основе этих двух параметров метод выбирает следующий шаг. Работа алгоритма МКПД обычно представляет собой цикл, повторяющийся множество раз и включающий четыре шага.

Ш а г 1. Выбор. На этом шаге алгоритм выбирает ход своего противника. Если такой ход существует – он выбирается, если нет – добавляется.

Ш а г 2. Расширение. К выбранному узлу с ходом противника добавляется узел со своим ходом и с нулевыми результатами.

Ш а г 3. Моделирование. Отыгрывается партия от текущего состояния игрового поля до чейлибо победы. Отсюда берется только первый ход и результаты.

Ш а г 4. Обратное распространение. Результаты моделирования будут распространяться от текущего состояния до корня. Ко всем родительским узлам добавляется единица в количество

[4] “Superhuman performance” (англ.) – термин, введенный для оценки действий машинного интеллекта, которые приводят к эффективности решения задачи в предметной области лучшей, чем этого мог бы достичь эксперт [37].

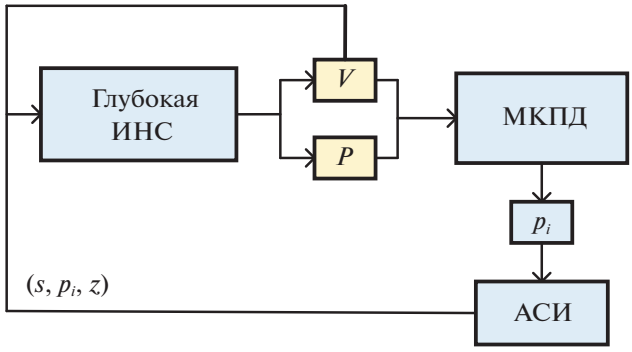

Рис. 4. Общая схема алгоритма работы ИСУ ИД

сыгранных партий, а если наталкиваются на узел победителя, то в этот узел добавляется единица в количество выигранных партий.

Для ИСУ ИД предлагается использовать модифицированный метод МКПД, предложенный в [37] для улучшения стратегии π. В этом случае стратегия управления π представляет собой вероятности каждого возможного действия условного противника для различных состояний. Кроме того, модификация метода позволяет сохранит вероятность каждого узла в соответствии со стратегией, которая впоследствии используется для корректировки оценки узла во время выбора действий. Согласно этому, на первом шаге выбирается дочерний узел (состояние) с максимальным значением вероятностной величины:

$$U_i = \frac{V_i}{N_i} + cP_i\sqrt{\frac{\ln N_p}{1+N_i}}, \tag{4.1}$$

где V_i – накопленная ценность состояния (или количество побед) i-го дочернего узла, N_i – количество посещений i-го дочернего узла; N_p – количество посещений родительского узла, P_i – вероятность выбора i-го действия и C – эмпирически задаваемый коэффициент ($C \geq 0$), необходимый для установки нужного баланса между шириной и глубиной поиска. Чем C больше, тем более глубокий будет поиск.

Высокая эффективность определяется тем, что с методом МКПД дерево решений растет асимметрично: более “интересные” узлы посещаются чаще, менее “интересные” – реже, а оценить отдельно взятый узел становится возможным без раскрытия всего дерева.

Предлагаемый подход для ИСУ ИД с применением АСИ подразумевает использование модели глубокой ИНС для оценки ценности состояния узла, добавленного на втором шаге МКПД (рис. 4). АСИ при проигрывании множества циклов МКПД дает все лучшие и лучшие экспертные стратегии управления π и приближенную функцию ценности состояния V, играя против себя с ускоренной работой МКПД. Для этого на вход глубокой ИНС f подается информация о текущем состоянии исследуемой ЭЭС (к примеру, по данным телеметрии или системы SCADA). На выходе ИНС мы получаем два результата в виде матриц с размерностью возможных состояний $s \in S$: стратегию π, представляющую вероятности каждого хода (действий) из текущего состояния p_i, и приближенную оценку текущего состояния V (т.е. вероятность выиграть партию из текущего состояния):

$$f(s) \to [p, V]. \tag{4.2}$$

Основная идея алгоритма ИСУ ИД будет заключаться в том, что предсказания ИНС могут быть улучшены, а игра, созданная МКПД, может использоваться для предоставления данных обучения. В этом случае экспертная стратегия глубокой ИНС будет улучшена посредством обучения предсказывать вероятности p, для базового состояния s_0 (родительского узла) в соответствии с улучшенной стратегией вероятностей π, полученной на основе запуска МКПД также для s_0. После запуска МКПД, улучшенные предсказания стратегии будут пропорциональны количеству посещений i-го дочернего узла с учетом задаваемой постоянной τ:

$$\pi_i \propto N_i^{1/\tau}. \tag{4.3}$$

Согласно [40], значения τ, близкие к нулю, формируют стратегию, соответствующую выбору лучшего действия, согласно оценке МКПД.

В свою очередь параметр ценности состояния V будет улучшаться путем обучения предсказанной оценки соответствовать конечному результату игры Z (выигрышу, проигрышу или ничьей). Тогда общую функцию потерь l можно будет рассчитать по следующему выражению:

$$l = (V - Z)^2 + \pi^T \ln p + \lambda \|\Theta\|_2^2. \tag{4.4}$$

Выражение (4.4) состоит из трех частей, которые требуют пояснения. Первая часть $(V - Z)^2$ – потеря оценки, показывающая, насколько хорошо глубокая ИНС умеет предсказывать результат игры (т.е. Z (результат игры) не должно отличаться от V (предсказанная оценка игры)); вторая часть $\pi^T \ln p$ – потеря стратегии, показывающая, насколько точно предсказываются те вероятности, которые будут получены при проходе по дереву (аналог функции потерь перекрестной энтропии в обучении с учителем, когда обучение стремится приблизить прогнозируемое распределение к истинному распределению); третья часть $\lambda \|\Theta\|_2^2$ – дополнительный член регуляризации с параметрами $\lambda \geq 0$ и Θ, представляющими параметры глубокой ИНС.

Обучение осуществляется исключительно в режиме самостоятельной игры (т.е. без внешних обучающих выборок). Первоначально набор параметров Θ глубокой ИНС инициализируется случайным образом. Затем нейросеть используется в нескольких играх, в которых она играет сама. В каждой из этих игр для каждого хода МКПД применяется для вычисления π. Конечный результат каждой игры определяется оценкой игры Z. Параметры Θ затем улучшаются с помощью градиентного спуска на функцию потерь для случайного выбора воспроизводимых состояний.

Разработка окружения (среды) для ИСУ ИД. Обучение и тестирование агентов ИСУ ИД должны проходить с помощью специальных сред, представляющих собой симуляторы поведения как отдельных компонент ЭЭС, так и всей системы в целом. Для создания таких симуляторов первоначально могут быть использованы как стандартные тестовые среды с последующей оригинальной модификацией, например из платформы OpenAI Gym, что позволит тестировать ИСУ ИД на решение простых задач управления отдельными компонентами ЭЭС, так и полностью оригинальные среды с привлечением таких программных платформ, как Matlab (power system analysis toolbox [43]) и АНАРЭС (разработка Института систем энергетики СО РАН [44]), для тестирования ИСУ ИД управлять всей ЭЭС. В этом отношении приоритетом является система АНАРЭС, которая имеет интеграцию с языком Python, что даст возможность обмена с внешними системами автоматизации управления ЭЭС через стандартные протоколы и механизмы: CIM/XML для модели и исходных данных ЭЭС; IEC-60870-5-104, IEC-61850 для обмена оперативными данными (измерения, состояния коммутационных аппаратов и устройств регулирования), а также применения библиотек МО с интерфейсом Python (tensorflow, keras).

При этом для создания таких симуляторов предполагается использование двух подходов: динамического и квазидинамического [45]. В первом подходе применяется полный набор дифференциально-алгебраических уравнений ЭЭС во времени с учетом различных видов возмущений и различных случайных внешних факторов (короткие замыкания, отключения нагрузки и генерации и т.д.). При этом динамические модели, являющиеся основой создаваемой среды для ИСУ ИД, должны отражать лишь наиболее важные свойства системы, упрощая тем самым решаемую задачу. В этом случае динамика электромеханического движения ЭЭС может быть представлена в следующем общем виде:

$$\begin{aligned} \frac{d\xi_S}{dt} &= \varphi_S(\xi_S, \xi_i, \xi_f, u, t); \\ \frac{d\xi_i}{dt} &= \varphi_i(\xi_S, \xi_i, \xi_f, u, t); \\ \frac{d\xi_f}{dt} &= \varphi_f(\xi_S, \xi_i, \xi_f, u, t), \end{aligned} \tag{4.5}$$

где ξ_S – переменные, описывающие поведение систем с большими постоянными времени (например, систем вторичного регулирования, которые должна реализовать УВ в течение 15 минут); ξ_i – переменные, характер изменения которых определяет свойства переходного процесса; ξ_f – переменные, характеризующие процессы с малыми постоянными времени (например, электро-

магнитные процессы при коммутациях в сети), которые одновременно учитывают быструю и медленную динамику; u, t – соответственно вектор дискретных переменных и время.

Второй упрощенный подход предполагает замену моделирования динамики ЭЭС серией установившихся режимов и обеспечивает точность моделирования динамики, достаточную для обучения агентов ИСУ ИД управлять устойчивостью по напряжению, локальными устройствами автоматики, предотвращать токовую перегрузки связей и т.д.

Игровая интерпретация окружения для ИСУ ИД. Как отмечалось выше, окружение или среда действия агентов для МКПД будет формироваться таким образом, чтобы учитывать возникающие непредсказуемые события в ЭЭС (неопределенность возобновляемых источников энергии, перегрузка оборудования, отключения, короткие замыкания и др.) и целенаправленные вредоносные действия (терроризм, кибератаки и т.п.). Такого рода факторы будут рассматриваться как определенные ограничения на действия агентов ИСУ ИД, например, как установленные лимиты по запасам режимной надежности, рыночные ограничения, ограничения управляющих способностей системы (недостаток регуляторов мощности, наличие прерывистой генерации, внезапные отключения оборудования и т.д.) и т.п. Таким образом заданные ограничения в разработанной симуляции среды будут рассматриваться как своего рода противники (виртуальные игроки) для каждого узла или ветки дерева в рамках МКПД. Например, в случае рыночных ограничений при управлении ЭЭС необходимо учитывать интересы различных субъектов рынка электроэнергии и мощности, а именно генерирующих компаний, энергосбытовых компаний, крупных потребителей. В инженерной теории игр считается, что такие игроки действуют как высокоавтономные субъекты и взаимодействуют с внешними объектами (к примеру, с Системным оператором ЕЭС, в рамках которой может работать ИСУ ИД) способами, которые соответствуют их собственным интересам. Обычно основой такой задачи является некооперативная игра, в которой основное внимание уделяется конкуренции между отдельными, рациональными игроками с глобальным равновесием Нэша [46].

В этих случаях агенты будут обучаться обходить возникающие или заданные ограничения, выполняя поиск оптимальной стратегии или обновляя текущую принятую стратегию с учетом возникших ограничений. Общим выигрышем в такой игре для ИСУ ИД будет приведение электрической сети в оптимальный нормальный или послеаварийный режим работы посредством необходимых УВ (ходов). Более того, при возникновении новых непредсказуемых ситуаций дерево будет инкрементально достраивается, т.е. возникать новая ветка событий с множественным проигрыванием возможных УВ, которые необходимы для победы в игре.

5. Экспериментальные расчеты. В данной работе предложена только концепция ИСУ ИД. При этом сама система автономного интеллектуального диспетчера находится в стадии разработки и тестирования простых примеров управления элементами ЭЭС (прежде всего, автоматическими устройствами, регуляторами), которые позволяют реализовывать УВ для регулирования режима работы энергосистемы. На этом этапе в настоящий момент были протестированы основные алгоритмы МО с подкреплением: ДП, МКУ, Q-обучение, SARSA и МКПД для следующих задач: управление напряжением и реактивной мощностью на подстанциях и управление серводвигателем постоянного тока. При этом для дискретизации состояний системы были использованы элементы нечеткой логики. Алгоритм и модели управления были построены в программных средах Julia (библиотека POMDPs [47]) и Python (библиотека Dissecting reinforcement learning) [48]. Предполагается, что разработанные реализации моделей являются первыми, простейшими реализациями концепции ИСУ ИД.

Управление напряжением и реактивной мощностью на подстанциях. Для эффективного управления напряжением и реактивной мощностью на подстанциях в литературе были предложены методы N-областей [49, 50], которые позволяют разделять область регулирования на N-областей с использованием фиксированного верхнего и нижнего пределов напряжения и реактивной мощности или коэффициента потребляемой мощности. В рассматриваемом примере мы использовали метод 17 областей в сочетании с нечеткой стратегией управления, предложенный в [50] (рис. 5).

Для регулирования напряжения был представлен регулятор напряжения трансформатора под нагрузкой (РПН), а для регулирования реактивной мощности – управляемый конденсаторной установки (УКРМ) (рис. 6). В качестве входных нечетких переменных выступали: отклонения реактивной мощности e_q и отклонения напряжения на вторичной обмотке трансформатора e_v. Выходными переменными выступали два УВ: направление переключения отпаек РПН –

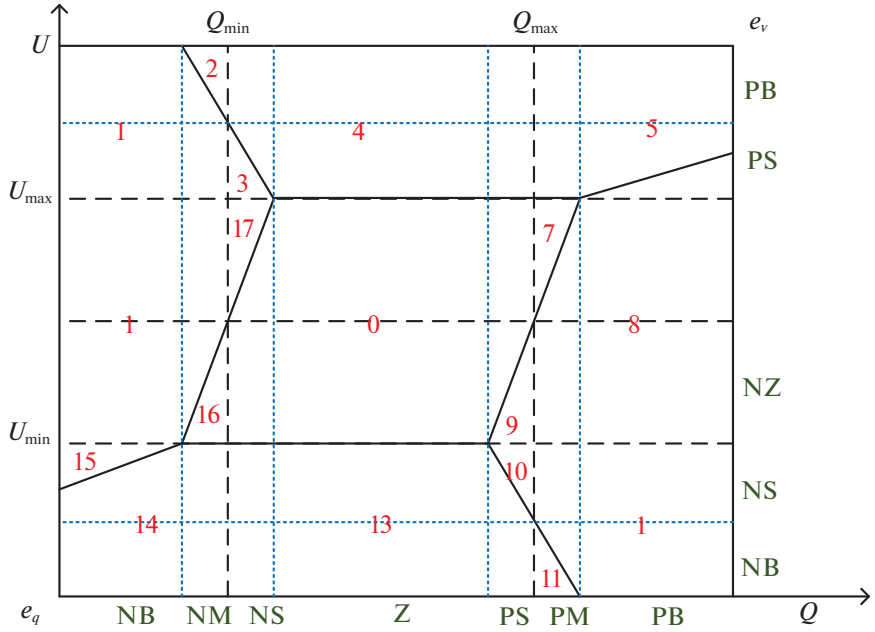

Рис. 5. Области регулирования напряжения и реактивной мощности на подстанции, разделенные на 17 областей с нанесением нечетких входных переменных e_v и e_q

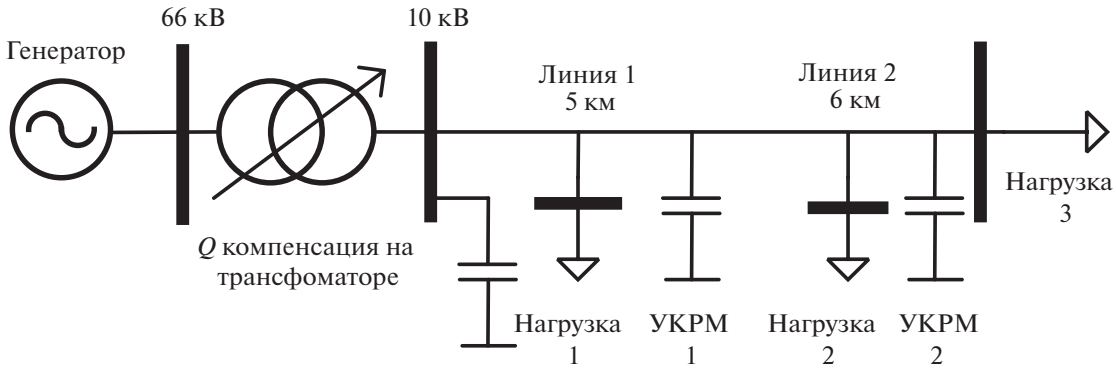

Рис. 6. Тестовая схема подстанции

“вверх” (up); “вниз” (down); “не переключать” (none) и регулирование подключаемых и отключаемых конденсаторов УКРМ – “подключить” (left), “отключить” (right), “ничего не делать” (none).

В соответствии с реальной ситуацией общая стратегия управления должна быть следующей. Границы регулирования напряжения $\{U_{min}, U_{max}\}$ должны быть относительно фиксированными, а границы регулирования реактивной мощности $\{Q_{min}, Q_{max}\}$ ограничены условием по изменениям напряжения. Когда отклонение напряжения выше установленной границы, то емкость в УКРМ либо не добавляется (при условии нехватки реактивной мощности), либо конденсаторы отключаются в небольшом количестве (при условии, что реактивная мощность не превышает слишком большие пределы). В свою очередь когда напряжение ниже, емкость добавляется в УКРМ или конденсаторы не отключаются.

Для обучения агента действовать в этих условиях было использовано стандартное окружение “grid world”, представляющее собой двумерную прямоугольную сетку размера (Ny, Nx) с агентом, начинающимся с одного квадрата сетки и пытающимся перейти на другой квадрат сетки, расположенный в другом месте. В рассматриваемом примере была создана сетка 6 × 7, где строки –

Таблица 2. Результаты обучения различных моделей МО с подкреплением в задаче управления напряжением и реактивной мощностью на подстанциях

Метод МО с подкреплением	Количество итераций
ДП	200
МКПД	20000
Q-обучение	5000
SARSA	5000

Таблица 3. Оптимальная стратегия управления π*, полученная ДП- и ВР-методами (Q-обучение, SARSA), где ← = влево, → = вправо, ↓ = вниз, ↑ = вверх, 0 = не двигать

Регулирование количества УКРМ	Переключение отпаек РПН						
	NB	NM	NS	Z	PS	PM	PB
NB	↓	↓	↓	↓	↓	← (↓*,**)	↓
NS	↓ (→*,**)	↓ (→*)	↓	↓	↓	↓	↓
NZ	→	0	0	0	←	←	←
PZ	→	↑ (→*,**)	→	0	0	0	← (↑*)
PS	↑	↑ (→**)	↑ (→**)	↑	←	↑ (←*,**)	↑
PB	↑	→ (↑*)	↑	↑	↑	↑	↑

* Действие на основе метода Q-обучения.

** Действие на основе метода SARSA.

параметр e_v, а столбцы – параметр e_q. Дискретность получаемых ячеек сетки определяли значения входных нечетких переменных, как показано на рис. 5. Агенту были заданы пространство действий в соответствии с выходными переменными, где движения “down”, “up”, “none” – переключения отпаек РПН, а движения “right”, “up”, “none” – регулирование количества конденсаторов УКРМ. При этом была использована так называемая разреженная награда, когда вознаграждение дается только по достижении конечного состояния цели, а в других состояниях отсутствует. В нашем случае основной целью агента являлась достижение области 0, где наблюдается оптимальный баланс напряжения и реактивной мощности (рис. 5).

Для сравнения эффективности в поиске оптимальной стратегии управления π* были задействованы различные методы МО с подкреплением: Q-обучение, SARSA, динамическое программирование, а также МКПД. При этом каждому из методов потребовалось разное количество итераций для решения этой задачи (табл. 2). В итоге, все методы пришли к схожей стратегии управления (табл. 3), что соотносится с принципами управления, предложенными в [46]. В табл. 3, направление стрелки указывает стратегию движения агента, совпадающую для трех методов – ДП, Q-обучение и SARSA. В случае различия в скобках были указаны другие направления для соответствующего метода.

В данном расчетном примере определенные расхождения в стратегиях для некоторых ситуаций являются корректными вариантами решения задачи, так как это соответствует реальным условиям регулирования напряжения и реактивной мощности, когда для достижения оптимальных режимных условий могут быть задействованы различные сценарии регулирования РПН и УКРМ. Изучение численных значений функции $Q^{\pi}(s,a)$ в матрице стратегии π* показывает, что в ряде случаев получаемые награды для некоторых пар действий близки по значению. Например, в методе SARSA для режимной ситуации $e_v = PS$ (“малое положительное”), $e_q = NS$ (“малое отрицательное”) выполняется действие “вправо” с наградой 7.00914, соответствующее отключению конденсатора УКРМ для стабилизации напряжения до номинальных значений. При этом действие “вверх”, соответствующее переключению РПН, имеет близкую по значению награду 6.96883.

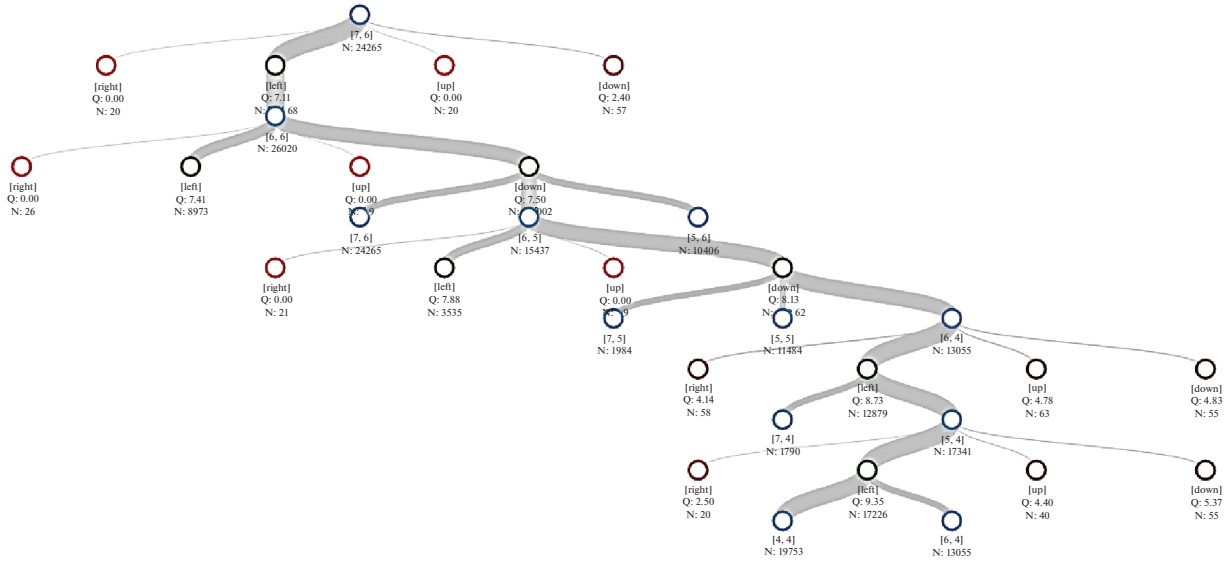

Рис. 7. Пример дерева, построенных методом МКПД для исходного состояния $e_V = NB$ (“*большое отрицательное*”), $e_q = PB$ (“большое положительное”)

Как было отмечено выше, метод МКПД является расширенной версией МКУ, выполняющий более глубокое моделирование возможных ситуаций и эффективно реализующий поиск решений в долгосрочной перспективе. Для обучения агента по методу МКПД было выполнено 20 тыс. итераций, в результате чего было построено дерево, с помощью которого можно наглядно отследить последовательность УВ для различных режимных состояний на подстанции. На рис. 7 показан фрагмент полученного дерева с последовательностью УВ для исходного состояния с максимальным снижением напряжения ниже $U_{\min}$ ($e_V = NB$) и максимально высокой реактивной мощностью выше $Q_{\max}$ ($e_q = PB$). При этом голубыми кружками показаны узлы дерева, характеризующие возможные состояния режима подстанции s, где указаны количество посещений узла N и координаты агента e_V, e_q в окружении. Обученный МКПД-агент верно выбирает стратегию коррекции коэффициента мощности на первом и заключительных этапах действий (реализация через УВ “влево”, соответствующие подключениям конденсаторов УКРМ) с последующим повышением напряжения на втором и третьем этапах (реализация УВ “вниз” на отпайке РПН). Это соответствует реальной оптимальной стратегии управления, отмеченной в [50], и стратегиям, полученным на базе ДП, SARSA и Q-обучения.

Интересно отметить, что в определенных узлах действия “влево” и “вправо” близки по показателям функции полезности, $Q^{\pi}(s,a)$ и количеству посещений узла N_i, что соответствует реальной ситуации, когда на фоне сниженного напряжения необходимо определять баланс УВ между повышением напряжения и корректировкой коэффициента мощности.

Управление серводвигателем постоянного тока. Серводвигатель – это вращающийся двигатель с датчиком обратной связи, который позволяет точно контролировать угловое положение, скорость и ускорение исполнительного механизма. Это устройство применяется в составе сервомеханизма и для работы требует относительно сложную систему управления, которая обычно разрабатывается специально для использования с серводвигателем. В ряде работ были предложены нечеткие системы управления серводвигателями [51, 52], которые показывают более высокую эффективность по сравнению с ПИД-регулированием.

В данном расчетном примере управление серводвигателем реализуется на основе решения задачи управления поворотным обратным маятником (рис. 8, 9) с помощью нечетких правил регулирования. Модель обратного маятника была реализована в программной среде Julia как окружение для создаваемого агента серводвигателя. В этом случае, на основе угла поворота маятника $\theta_2 \in [-\pi/2, \pi/2]$ (рад) и его угловой скорости $\dot{\theta}_2 \in [-\pi, \pi]$ (рад/с) реализуется нечеткое регулирование углового положения и скорости вращающегося рычага серводвигателя. Таким образом параметры θ_2 и $\dot{\theta}_2$ выступают как входные лингвистические переменные нечеткой системы, принимающие 12 значений, под которыми понимаются дискретные диапазоны поворота (табл. 4).

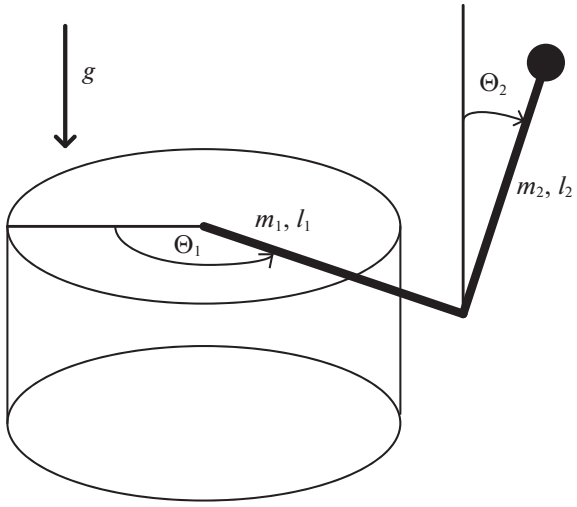

Рис. 8. Иллюстрация к построению упрощенной модели серводвигателя в виде поворотного обратного маятника

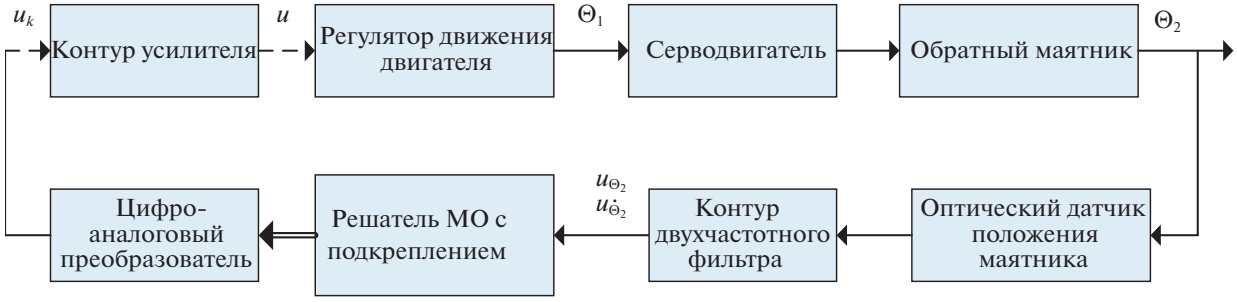

Рис. 9. Принципиальная схема управления серводвигателем на основе обратного маятника с применением решателя МО с подкреплением, адаптированная из [47]

Кроме того, представленной системе задаются массы вращающего рычага серводвигателя m_1 и обратного маятника m_1, их длины соответственно l_1 и l_2, а также ускорение свободного падения g. В пространстве действий мы имеем три силы $[-50, 0, 50]H$, которые могут быть приложены к вращающемуся рычагу для раскачивания маятника. Это соответствует категориям действий: “влево”, “вправо”, “не двигать”.

Учитывая принятые параметры и модель (рис. 5), угол θ_2 и угловая скорость $\dot{\theta}_2$ в момент $t + 1$ обновляются по следующему выражению:

$$\theta_{2,t+1} = \theta_{2,t} + \frac{g \sin \theta_{2,t} - \alpha m_2 l_2 \dot{\theta}_{2,t}^2 \sin 2\theta_{2,t}/2 + \alpha \cos \theta_{2,t} a_t}{4l_2/3 - \alpha m_2 l_2 \cos^2 \theta_{2,t}} \Delta t, \tag{5.1}$$

где $\alpha = 1/m_1 + m_2$; a_t – действие в момент времени t; $\Delta t = 0.1$ с – шаг времени.

Выражение (5.1) определяло окружение для обучения агента. При этом, награда для модели обновлялась с учетом косинус θ_2, что реализовало правило “больше угол – меньше награда”. Например, награда равна 0.0, когда положение маятника горизонтально, и 1 – при вертикальной позиции. Когда маятник полностью горизонтален, эпизод заканчивается. Оптимальная стратегия заключается в компенсации угла и скорости изменения, сохраняя маятник в максимально возможной степени вертикального положения. Использование нечеткой логики позволяет дискретизировать задачу управления и представить полученную стратегию управления, π в виде квадратной матрицы.

Таблица 4. Множества значений двух лингвистических переменных, характеризующих движение маятника

Значения переменной	Кодировка	Диапазон для угла θ_2, рад	Диапазон для угловой скорости $\dot{\theta}_2$, рад/с
Значительный отрицательный	NH	$[-\pi/2, -5\pi/12]$	$[-\pi, -5\pi/6]$
Большой отрицательный	NL	$[-5\pi/12, -4\pi/12]$	$[-5\pi/6, -4\pi/6]$
Средний отрицательный	NM	$[-4\pi/12, -3\pi/12]$	$[-4\pi/6, -3\pi/6]$
Малый отрицательный	NS	$[-3\pi/12, -3\pi/12]$	$[-3\pi/6, -3\pi/6]$
Минимальный отрицательный	NK	$[-2\pi/12, -\pi/12]$	$[-2\pi/6, -\pi/6]$
Отрицательный ноль	NZ	$[-\pi/12, 0]$	$[-\pi/6, 0]$
Положительный ноль	PZ	$[0, \pi/12]$	$[0, \pi/6]$
Минимальный положительный	PK	$[\pi/12, 2\pi/12]$	$[\pi/6, 2\pi/6]$
Малый положительный	PS	$[2\pi/12, -3\pi/12]$	$[2\pi/6, -3\pi/6]$
Средний положительный	PM	$[3\pi/12, 4/12]$	$[3\pi/6, 4/6]$
Большой положительный	PL	$[4\pi/12, 5/12]$	$[4\pi/6, 5/6]$
Значительный положительный	PH	$[5\pi/12, \pi/12]$	$[5\pi/6, \pi]$

Для моделирования были использован методы МКУ и МКПД для поиска оптимальной стратегии π^*. В случае МКУ, для улучшения процесса поиска была использована жадная стратегия ε. Каждый эпизод составлял 100 шагов (10 с). Максимальная награда в Q-функции, которая может быть получена, составляла 100 (т.е. когда маятник вертикален для всех этапов). Результаты обучения показали, что предложенный алгоритм может достаточно быстро найти хорошее решение, достигнув среднего значения награды 58 (рис. 10) без ее последующего накопления, что и означает нахождение окончательной стратегии π^*. Такая стратегия имеет очень хорошую

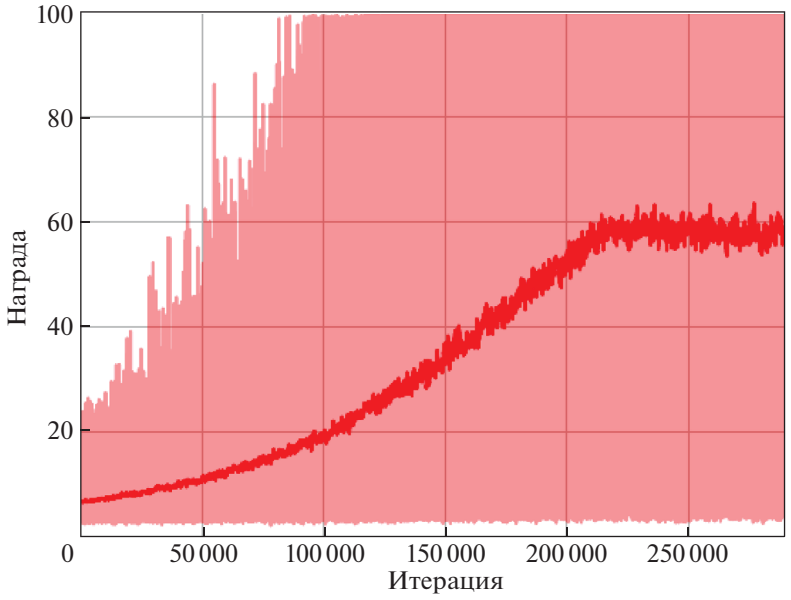

Рис. 10. График изменения накопленной награды Q-функции для каждого эпизода (светло-красная линия) и среднее скользящее для 500 эпизодов (темно-красная линия) на протяжении итеративного процесса МКУ

Таблица 5. Оптимальная стратегии управления π^* серводвигателем постоянного тока, полученная методом МКУ (где ← = влево, → = вправо, 0 = не двигать)

Угловая скорость $\dot{\theta}_2$	Положение маятника, θ_2											
	NH	NL	NM	NS	NM	NZ	PZ	PM	PS	PM	PL	PH
NH	→	←	→	→	→	0	←	←	→	←	0	→
NL	←	←	→	→	→	→	←	←	←	→	←	→
NM	→	0	→	→	→	→	→	←	→	→	0	←
NS	→	→	→	→	→	→	→	←	→	0	→	0
NM	→	→	→	→	→	→	→	0	→	←	→	→
NZ	0	←	0	→	→	→	→	←	0	0	→	←
PZ	0	0	0	<	→	→	0	←	←	→	←	→
PM	←	←	←	0	→	→	←	←	←	←	←	→
PS	←	→	←	←	→	0	←	←	←	←	←	0
PM	→	←	→	←	→	→	←	←	←	←	←	←
PL	0	0	0	←	→	→	0	←	←	←	←	←
PH	0	0	0	←	0	←	0	←	←	←	←	←

эффективность и легко справляется с сохранением маятника в вертикальном положении для всего эпизода (10 с) (табл. 5).

Использование метода МКПД позволило найти оптимальную стратегию π^* после 120 тыс. итераций, что значительно быстрее (на более чем 100 тыс. итераций), чем в МКУ. На рис. 11 приведен пример дерева по методу МКПД для случайного состояния маятника. Спуск по дереву показывает, что после серии УВ (движений) в итоге маятник остается в вертикальном положении.

Таким образом, согласно расчетным примерам, методы МО с подкреплением позволяют находить оптимальные стратегии управления системой в случаях, когда сам алгоритм управления не задан явно. Более того, для решения задачи не требуется привлечения ББД, не требуется аккумулирование экспертного опыта. Метод МКПД позволяет формировать многовариантную модель поиска оптимальных УВ в долгосрочной перспективе, что дает возможность выполнить оценку влияния выбранных УВ на состояние системы в будущем.

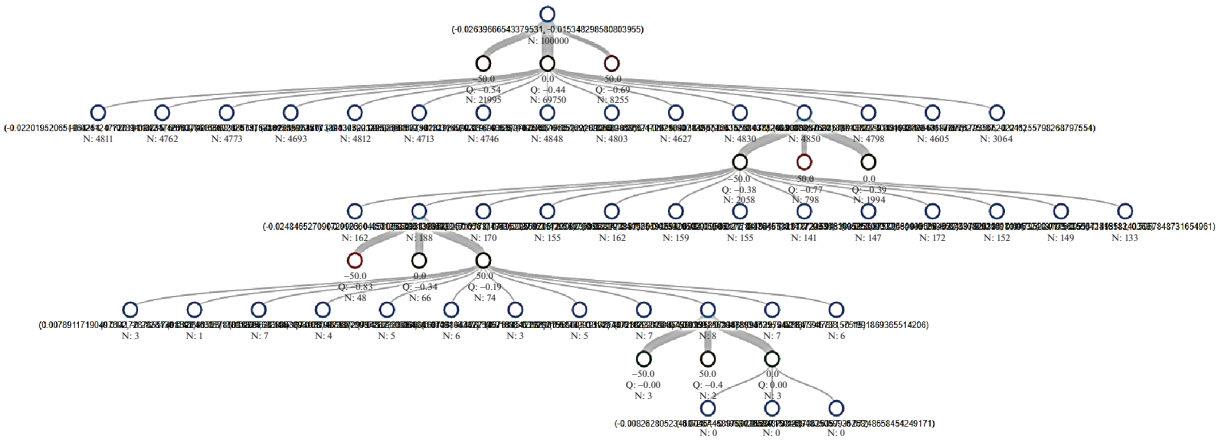

Рис. 11. Пример дерева, построенного методом МКПД для случайного состояния маятника

Заключение. Все возрастающая сложность управления современными ЭЭС с одновременным внедрением различных “умных” устройств регулирования приводит к необходимости разработки инструментов управления нового типа. В статье предложена ИСУ ИД, построенная на базе технологии глубокого МО с подкреплением при совместном использовании МКПД и глубоких ИНС. Применение АСИ позволяет обучать ИСУ ИД для управления режимами работы электрическими сетями без привлечения ББД и экспертного опыта диспетчеров. Этого удается добиться за счет того, что с помощью метода МКПД моделируется множество возможных ситуаций, в которых ИСУ ИД играет сама с собой большое количество циклов, тем самым формируя оптимальную экспертную стратегию управления ЭЭС. Полученные результаты наглядно демонстрируют возможность создания автономной интеллектуальной системы, способной выполнять функции диспетчера лучше человека в рамках существующих АСУ.

СПИСОК ЛИТЕРАТУРЫ

1. *Дроздов А.Д., Засыпкин А.С., Аллилуев А.А. и др.* Автоматизация энергетических систем. Учебное пособие для студентов электроэнергетических специальностей вузов. М.: Энергия, 1977. 440 с.
2. The World Market Study of SCADA, Energy Management Systems, Distribution Management Systems and Outage Management Systems in Electric Utilities: 2017–2019. A Four-Volume Report by Newton-Evans Research Company. Dublin, 2017.
3. *Voropai N.I., Tomin N.V., Sidorov D.N. et al.* A Suite of Intelligent Tools for Early Detection and Prevention of Blackouts in Power Interconnections // Autom Remote Control. 2018. № 79. P. 1741–1755.
4. *Панасецкий Д.А.* Совершенствование структуры и алгоритмов противоаварийного управления ЭЭС для предотвращения лавины напряжения и каскадного отключения линий: Дис. ... канд. техн. наук: 05.14.02. Иркутск: ИСЭМ СО РАН, 2015. 192 с.
5. *Сулейманова А.М.* Интеллектуальный советчик диспетчера по управлению электроэнергетической системой в аварийных ситуациях: Дис. ... канд. техн. наук: 05.13.06. Уфа: УГАТУ, 1993. 188 с.
6. *Поспелов Д.А.* Принципы ситуационного управления // Изв. АН СССР. Техн. кибернетика. 1971. № 2. С. 3–10.
7. *Кирилин И.В.* Классификация состояний электрических сетей промышленных предприятий для управления компенсацией реактивной мощности: Дис. ... канд. техн. наук: 05.14.02. Красноярск: Сибирский федеральный ун-т, 2011, 218 с.
8. *Hiyama T., Hubbi W., Ortmeyer T.H.* Fuzzy Logic Control Scheme with Variable Gain For Static Var Compensator to Enhance Power System Stability // IEEE Trans. on Power Systems. 1999. V. 4. P. 186–191.
9. Комплекс Каскад-НТ 2.0 [электронный документ], 2016. http://www.cascadent.ru/cascade2016.pdf.
10. Volt-VAr Management Solutions for Smart Grid Distribution Automation Applications [электронный документ] // Product Brochure of the ABB Smart Grid Center of Excellence. https://library.e.abb.com/public/d9e73a8d3d91161bc1257b6a006cc340/VVMS%20brochure_final_v4.pdf .
11. *Григорьев С.П.* АСУ ТП: Ошибки первого и второго рода [электронный документ] // Deming Pro. https://www.deming.pro/spc-cases-apcs.html.
12. *Negnevitsky M.* Artificial Intelligence: A Guide to Intelligent Systems. 3rd edn. Harlow, England: Addison Wesley. 2011. 504 p.
13. *Kaci A, Kamwa I., Dessaint L. et al.* Synchrophasor Data Baselining and Mining for Online Monitoring of Dynamic Security Limits // IEEE Trans. on Power Systems. 2014. V. 29. № 6. P. 2681–2695.
14. *Liu C., Bak C.L., Chen Z. et al.* Dynamic security assessment of western Danish power system based on ensemble decision trees // Proc. 12th IET Intern. Conf. on Developments in Power System Protection. Copenhagen, 2014.
15. *Tomin N.V., Kurbatsky V.G., Reutsky I.S.* Hybrid Intelligent Technique for Voltage/VAR Control In Power Systems // IET Generation, Transmission and Distribution. 2019. V. 13. № 20. P. 4724–4732.
16. *Kogan I., Boehme K., Herrmann H.-J.* Внедрение передовых технологий машинного обучения в реле защиты // Сб. тр. Междунар. научно-техническая конф. и выставки “Релейная защита и автоматика энергосистем 2017”. Санкт-Петербург. 2017.
17. *Busch R.* The Future of Manufacturing. Artificial Intelligence: Optimizing Industrial Operations [электронный документ] // Siemens Web Page. 2018. https://www.siemens.com/innovation/en/home/pictures-of-the-future/industry-and-automation/the-future-of-manufacturing-ai-in-industry.html.
18. *Webel S., Nikolaus K., Pease A.F.* Autonomous Systems. Getting Machines to Mimic Intuition [электронный документ] // Siemens Web Page. 2016. https://www.siemens.com/innovation/en/home/pictures-of-the-future/digitalization-and-software/autonomous-systems-machine-learning.html
19. *Mocanu E., Nguyen P.H., Gibescu M.* Deep Learning for Power System Data Analysis, In Big Data Application in Power Systems / Eds R. Arghandeh, Y. Zhou. Elsevier, 2018. P. 125–158.

20. *Tang Y., He H., Wen J., Liu J.* Power System Stability Control for a Wind Farm Based on Adaptive Dynamic Programming // IEEE Trans. Smart Grid. 2015. V. 6. 2015. P. 166–177
21. *Liu Y.* Machine Learning for Wind Power Prediction. MCS Thesis: Canada. University of New Brunswick, 2016. 88 p.
22. *Francois-Lavet V., Taralla D., Ernst D. et al.* Deep Reinforcement Learning Solutions for Energy Microgrids Management // European Workshop on Reinforcement Learning. Barselona. 2016
23. *Zhukov A., Tomin N., Sidorov D., Kurbatsky V., Panasetsky D.* On-Line Power Systems Security Assessment Using Data Stream Random Forest Algorithm // Innovative Computing, Optimization and Its Applications. Studies in Computational Intelligence. V. 741 / Eds I. Zelinka, P. Vasant, V. Duy, T. Dao. Springer, 2018.
24. *Tomin N., Negnevitsky M., Rehtanz Ch.* Preventing Large-Scale Emergencies in Modern Power Systems: AI Approach // Advanced Computational Intelligence and Intelligent Informatics. 2014. V. 18. № 5. P. 714–727.
25. *Xu Y., Zhang W., Liu W., Ferrese F.* Multiagent-Based Reinforcement Learning for Optimal Reactive Power Dispatch // IEEE Trans. Syst., Man, Cyber. 2012. V. 42. P. 1742–1751.
26. *Belkacemi R., Abdulrasheed Babalola A., Zarrabian S.* Real-Time Cascading Failures Prevention Through MAS Algorithm and Immune System Reinforcement Learning // Electric Power Components and Systems. 2017. V. 45. № 5. P. 505–519.
27. *Ye D., Zhang M., Sutanto D.* A Hybrid Multiagent Framework with Q-Learning For Power Grid Systems Restoration // IEEE Trans. Power Syst. 2011. V. 26. P. 2434–2441.
28. *Zarabbian S., Belkacemi R., Babalola A.A.* Reinforcement Learning Approach for Congestion Management and Cascading Failure Prevention with Experimental Application // Elec. Power Syst. Research. 2016. V. 141. P. 179–190.
29. *Glavic M., Ernst D., Wehenkel L.* A Reinforcement Learning Based Discrete Supplementary Control for Power System Transient Stability Enhancement // Engineering Intelligent Systems for Electrical Engineering and Communications. 2005. V. 13. P. 81–88.
30. *Саттон Р.С., Барто Э.Г.* Обучение с подкреплением / Пер. с англ. 2-е изд. М.: БИНОМ, 2014. 402 с.
31. *El Chamie M., Acikmese B.* Finite-Horizon Markov Decision Processes with State Constraints // arXiv:1507.01585 [math.OC], 2015.
32. *Glavic M., Fonteneau R., Ernst D.* Reinforcement Learning for Electric Power System Decision and Control: Past Considerations and Perspectives // IFAC-PapersOnLine. 2017. V. 50. № 1. P. 6918–6927.
33. *Yousefian R., Kamalasadan K.* Design and Real-Time Implementation of Optimal Power System Wide-Area System-Centric Controller Based on Temporal Difference Learning // IEEE Trans. on Industry Applications. 2015. V. 1. P. 395–401.
34. *Ernst D., Wehenkel L., Glavic M.* Power systems stability control: Reinforcement learning framework // IEEE Trans. Power Syst. 2004. V. 19. P. 427–435.
35. *Ernst D., Glavic M., Capitanescu F. et al.* Reinforcement Learning Versus Model Predictive Control: A Comparison on a Power System Problem // IEEE Trans. Syst., Man, Cyber. 2009. V. 39. P. 517–529.
36. *Vandael S., Claessens B., Ernst D. et al.* Reinforcement Learning of Heuristic EV Fleet Charging in A Day-Ahead Electricity Market // IEEE Trans. Smart Grid. 2015. V. 6. P. 1795–1805.
37. *Silver D., Schrittwieser J., Simonyan K. et al.* Mastering the Game of Go without Human Knowledge // Nature. 2017. V. 550. P. 354–359.
38. *Mnih V., Kavukcuoglu K., Silver D. et al.* Playing Atari with Deep Reinforcement Learning // in arXiv, abs/1312.5602. http://arxiv.org/abs/1312.5602.
39. *Irpan A.* Deep Reinforcement Learning Doesn't Work Yet // Sorta Insightful. 2018. https://www.alexirpan.com/2018/02/14/rl-hard.html.
40. *Silver D., Huang A., Maddison Ch. et al.* Mastering The Game of Go with Deep Neural Networks And Tree Search // Nature. 2016. V. 529. № 7587. P. 484–489.
41. Программный комплекс для создания автоматизированных систем оперативно-технологического управления в сетевых компаниях СК-11. http://www.monitel.ru/files/downloads/products/Broshyura%20-%20CK-11.pdf?201711.
42. *Tomin N.V., Kurbatsky V.G., Negnevitsky M.* The Concept of the Deep Learning-Based System Artificial Dispatcher to Power System Control and Dispatch // arXiv:1805.05408v1 [cs.CY] 7 May, 2018.
43. *Milano F.* Power System Modelling and Scripting. London.: Springer, 2010. 558 p.
44. Программно-вычислительный комплекс АНАРЭС. http://anares.ru.
45. *Воропай Н.И., Томин Н.В., Курбацкий В.Г. и др.* Комплекс интеллектуальных средств для предотвращения крупных аварий в энергосистемах. Новосибирск: Наука, 2016. 332 с.
46. *Fangxing Li, Yan Du.* From AlphaGo to Power System AI: What Engineers Can Learn from Solving the Most Complex Board Game // IEEE Power and Energy Magazine. 2018. V. 16. № 2. P. 76–84.
47. *Egorov M., Sunberg Z.H., Balaban E. et al.* POMDPs.jl: A Framework for Sequential Decision Making under Uncertainty // Machine Learning Research. 2017. V. 18(26). P. 1–5.

48. *Patacchiola M.* Dissecting Reinforcement Learning // Github. https://github.com/mpatacchiola/dissecting-reinforcement-learning.
49. *Sheng L.* Strategy Research of Substation Voltage Reactive Control Basing on The Ninth Region Plot // Qinghai Electric Power. 2005. V. 24. P. 1–4.
50. *Wu X., Wang J.-Ch., Yang P. et al.* Fuzzy Control on Voltage/Reactive Power in Electric Power Substation // Eds. B. Cao, T.-F. Li, C.-Y. Zhang. Fuzzy Info. and Eng. 2018. V. 2. P. 1083–1091.
51. *Manikandan R., Arulmozhiyal R.* Position Control of DC Servo Drive Using Fuzzy Logic Controller // Proc. Intern. Conf. on Advances in Electrical Engineering (ICAEE). India, 2014.
52. *Yunhai H., Bingfeng X., Gen Ping L.* A Power Control Method for Inverted Pendulum Based on Fuzzy Control // Proc. Intern. Conf. on Computer, Mechatronics, Control and Electronic Engineering. China, 2010.