

Ю. Я. Лямец [и др.] // Известия вузов. Электромеханика. 2009. С. 24–30.

2. Дементий Ю. А. Активное обучение интеллектуальной релейной защиты. Противостоящие режимы // Электрические станции. 2021. № 9(1082). С. 45–53.

3. Estimating the Support of a High-Dimensional Distribution / V. Schölkopf [et al.] // Neural Computation, № 13. 2001. PP. 1443–1471.

4. Мартынов М. В. Исследование и разработка обучаемых модулей микропроцессорных защит линий электропередачи: дис. ... канд. техн. наук. Чебоксары, 2014. 226 с.

Маслов А.Н., Дементий Ю.А.
(Чебоксары, ООО «Релематика»)

АВТОМАТИКА УПРАВЛЕНИЯ ДУГОГАСЯЩИМ РЕАКТОРОМ НА ОСНОВЕ АЛГОРИТМА ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Алгоритмы с использованием машинного обучения не просто демонстрируют результаты, аналогичные результатам классических алгоритмов, но зачастую опережают их, позволяя добиться максимально возможных показателей качества для имеющейся информационной базы [1, 2]. В рассматриваемой работе будет показано применение алгоритма обучения с подкреплением (reinforcement learning) с целью создания автоматики управления дугогасящим реактором (ДГР) плунжерного типа. Работа является демонстрацией возможного принципа построения автоматики управления ДГР, поэтому рассматривается одна из наиболее простых моделей реактора.

В алгоритме обучения с подкреплением имеется агент (в данном случае автоматика управления ДГР) и среда, с которой он взаимодействует (модель ДГР). За один шаг агент может формировать одно из трёх действий – движение плунжера вверх, движение плунжера вниз, запрет движения плунжера. Среда реагирует на действие, выдаваемое агентом, и формирует состояние (расстройка компенсации ёмкостного тока сети), ха-

рактирующее среду, а также награду, что позволяет агенту определить правильность выбранного действия. Схема работы алгоритма обучения с подкреплением показана на рисунке.

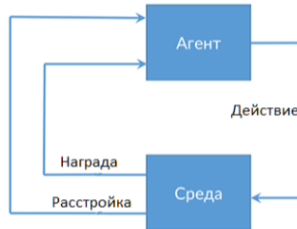


Схема работы алгоритма обучения с подкреплением

Целью обучения является выработка агентом оптимальной стратегии управления средой в результате анализа состояния среды и получаемой награды, в зависимости от совершаемых агентом действий [3].

В представленной работе среда включает в себя модель реактора с номинальным напряжением 6,3 кВ, минимальным током компенсации 25 А и максимальным током 125 А, а также модель сети с ёмкостным током 75 А.

Среда возвращает агенту текущее состояние (расстройка), а также награду. Расстройка компенсации определяется выражением

$$v = \left(1 - \frac{I_L}{I_C} \right) 100 \% ,$$

где I_L – ток реактора, А; I_C – ток сети, А.

Награда агента, формируемая средой, определяется следующим выражением:

$$r = \frac{1}{\lceil v \rceil + c} ,$$

где $\lceil v \rceil$ – модуль расстройки компенсации, %; c – коэффициент сглаживания функции, число.

Наибольшее значение, равное $1/c$, функция принимает при нулевой расстройке. Коэффициент c призван для ограничения награды при значениях расстройки, близких к нулю, и в текущей работе принимался равным единице.

Целью агента является максимизация получения награды. Таким образом, оптимальной стратегией действия агента (автоматики ДГР) является удержание расстройки вблизи нуля, так как именно в этой области получаемая награда максимальна.

В начале обучения агент не имеет никакой стратегии управления, а действует наугад. В процессе обучения агент формирует свою стратегию на основе анализа своих действий, а также состояния среды и получаемой награды.

В результате обучения агента в ходе 20000 игр, каждая из которых состояла из 100 шагов, удалось обучить агента, который выработал наиболее оптимальную стратегию – смещение плунжера дугогасящего реактора к положению, приводящему к минимально возможной расстройке (с учётом дискретизации шага). Одна игра представляет собой случай с произвольным начальным положением плунжера, в ходе которой агенту даётся 100 шагов. За один шаг агент может сдвинуть плунжер на 0,1 % вверх или вниз, либо не предпринимать никаких действий.

Обучение с подкреплением позволяет построить автоматику управления дугогасящим реактором, не уступающую классическим алгоритмам. Более того, усложнение среды и учёт реальных особенностей работы приводит к необходимости усложнения классических алгоритмов, в свою очередь алгоритм обучения с подкреплением остаётся неизменным, так как учёт любых особенностей будет произведён на этапе формирования стратегии агента.

В рассматриваемом примере среда включает в себя реактор с конкретными параметрами, однако, полученный агент применим для реакторов плунжерного типа любой мощности, так как обученный агент для своей работы использует только значение расстройки.

Литература

1. *Дементий Ю. А.* Активное обучение классификатора режимов работы объекта с использованием имитационной модели // Современные тенденции развития цифровых систем релейной защиты и автоматики: материалы науч.-техн. конф. молодых специалистов форума «РЕЛАВЭКСПО-2021». Чебоксары: Изд-во Чуваш. ун-та, 2021. С. 157–162.

проведена дискретизация осей наблюдаемых значений;
получены прямоугольные ячейки наблюдаемых значений;
шаг дискретизации наблюдаемых значений позволил разделить каждую ось наблюдаемых значений на 100 ячеек;
алгоритм определил размер матрицы \mathbf{M} как $2 \times 100 \times 100$;
пустая матрица \mathbf{M} приняла первый набор объектных параметров в соответствующие индексы положения в матрице;
во время каждой итерации алгоритм генерирует новые точки, сравнивая их с предыдущими. По результату сравнения происходит перезапись значений объектных параметров матрицы \mathbf{M} .

Алгоритм завершил работу по критерию остановки, который гарантирует получение границ интервала объектных параметров, стремящихся к истинным значениям. Результат работы алгоритма показан на рис. 3. Визуализация соответствует значениям матрицы \mathbf{M} .

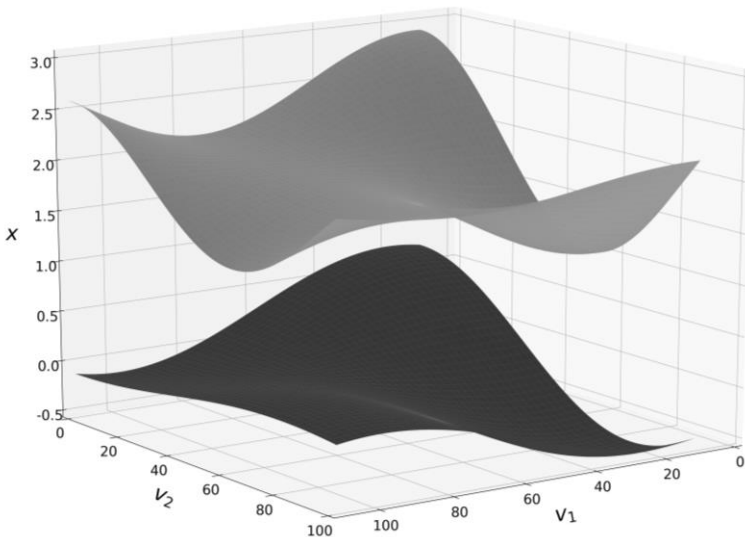


Рис. 3. Пример работы алгоритма активного обучения моделей оценок верхней и нижней границ интервала

Таким образом, предлагаемый метод позволяет построить интервалы оценок объектных параметров наблюдаемого объекта; предлагаемый метод обучения алгоритмов РЗА является

универсальным инструментом построения РЗА для объекта с любым набором параметров; предлагаемый метод позволяет реализовать защиту, основанную на оценке интервала объектных параметров на любой современной платформе РЗА.

Литература

1. *Мартынов М. В.* Исследование и разработка обучаемых модулей микропроцессорных защит линий электропередачи: диссертация ... кандидата технических наук: 05.14.02 / Мартынов М.В.; Чебоксары, Чуваш. гос. ун-т им. И.Н. Ульянова, 2014.

2. Об информационной теории релейной защиты / Ю. Я. Лямец [и др.] // Известия Академии электротехнических наук РФ. 2009. С. 32–44.

3. *Дементий Ю. А.* Активное обучение классификатора режимов работы объекта с использованием имитационной модели // Современные тенденции развития цифровых систем релейной защиты и автоматики: материалы науч.-техн. конф. молодых специалистов форума «РЕЛАВЭКСПО-2021». Чебоксары: Издво Чуваш. ун-та, 2021. С. 157–162.

4. *Дементий Ю. А., Шорников Е. В.* Машинное обучение для интервальной оценки параметров объекта // Современные тенденции развития цифровых систем релейной защиты и автоматики: материалы науч.-техн. конф. молодых специалистов форума «РЕЛАВЭКСПО – 2021». Чебоксары: Изд-во Чуваш. ун-та, 2021. С. 153–157.

5. *Дементий Ю. А., Шорников Е. В.* Применение активного обучения для построения моделей параметрической идентификации // Динамика нелинейных дискретных электротехнических и электронных систем: материалы XIV Всерос. науч.-техн. конф. Чебоксары: Изд-во Чуваш. ун-та, 2021. С. 446–450.

6. *Дементий Ю. А., Шорников Е. В.* Регуляризация в задаче обучения интеллектуальных алгоритмов параметрической идентификации // Динамика нелинейных дискретных электротехнических и электронных систем: материалы XIV Всерос. науч.-техн. конф. Чебоксары: Изд-во Чуваш. ун-та, 2021. С. 450–454.