https://doi.org/10.26160/2474-5901-2020-21-75-81

MATHEMATICAL MODEL OF CONTROLLED ELECTRIC SIGNAL FILTRATION BASED ON REINFORCEMENT LEARNING

Tao S., Smirnov A.S., Astashev M.G.

Keywords: machine reinforcement learning, Switch mode power supply, PID controller, Q-learning.
Abstract. With the development of machine learning, we can replace many commonly used control methods in industry, such as traditional PID controllers. This time we take the switch mode power supply as an example, using the popular reinforcement learning method in machine learning to compare its control effect with the PID controller and compare their control efficiency.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ УПРАВЛЯЕМОЙ ФИЛЬТРАЦИИ ЭЛЕКТРИЧЕСКОГО СИГНАЛА НА ОСНОВЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Тао С., Смирнов А.С., Асташев М.Г.

Ключевые слова: машинное обучение, импульсный источник питания, ПИД-регулятор, Q-обучение.
Аннотация. С развитием машинного обучения мы можем заменить многие общепринятые в отрасли методы управления, такие как традиционные ПИД-контроллеры. На этот раз мы возьмем в качестве примера импульсный источник питания, используя популярный метод обучения с усилением в машинном обучении, чтобы сравнить его эффект управления с ПИД-регулятором и сравнить их эффективность управления.

Introduction

Switch mode power supplies are one of the most important functional units of the secondary power sources of electronic and electrical equipment for various purposes. Currently, voltage stabilization of supply is provided, in the vast majority of cases, by means of proportional-integral-differential (PID) regulators.

Nowadays, machine learning methods are becoming increasingly effective. Reinforcement learning (RL) is one of the methods of machine learning, in which a learning algorithm called an agent interacts with a certain environment, exerting influences on it that can be considered as environmental management. In the process of interaction, the agent receives reinforcement signals from the environment (feedback on the effectiveness of the produced control actions from the point of view of achieving the control goal), analyzing which it produces training.

The purpose of this work is to research the effect of the reinforcement learning algorithm for stabilizing the output signals of a voltage regulator.

Reinforcement Learning

RL is one of the machine learning methods used to describe and solve the problem of maximizing returns or achieving specific goals by an agent through a learning strategy in interaction with the environment. The focus of RL is on finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge) [1]. *Q*-learning was introduced by Chris Watkins [2] in 1989 and is a

reinforcement learning method based on the value function. "Q" names the function that returns the reward used to provide the reinforcement and can be said to stand for the "quality" of an action taken in a given state [3]. During iteration *t*, we have the state s_t, when the agent performs the action, and then receives the reward r_t for the transition to the new state s_{t+1}, so that the *Q* value is updated, and its update formula:

$$Q(s_t, a_t) = (1-\alpha) \cdot Q(s_t, a_t) + \alpha \cdot r_t + \gamma \cdot \max_a Q(s_{t+1}, a_t). \quad (1)$$

where $Q(s_t, a_t)$ - old value; α - learning rate; r_t - reward; γ - discount value; $\max_a Q(s_{t+1}, a)$ - estimate of optimal future value.

From the above formula, it can be seen that updating the *Q* value is equivalent to using the learning speed to perform a weighted sum of the past *Q* value and the current *Q* value.

In the formula (1), let $\lambda = \gamma/\alpha$ and combine the similar terms on the right side to get:

$$\begin{aligned} Q(s_t, a_t) &= (1-\alpha) \cdot Q(s_t, a_t) + \alpha \cdot r_t + \gamma \cdot \max_a Q(s_{t+1}, a_t) \\ &= Q(s_t, a_t) + \alpha \cdot r_t + \gamma \cdot \max_a Q(s_{t+1}, a_t) - \alpha \cdot Q(s_t, a_t) \\ &= Q(s_t, a_t) + \alpha \cdot (r_t + \frac{\gamma}{\alpha} \cdot \max_a Q(s_{t+1}, a_t) - Q(s_t, a_t)) \\ &= Q(s_t, a_t) + \alpha \cdot (r_t + \lambda \cdot \max_a Q(s_{t+1}, a_t) - Q(s_t, a_t)). \end{aligned} \quad (2)$$

According to the definition of the Bellman equation in reinforcement learning, we have:

$$r_t + \lambda \cdot \max_a Q(s_{t+1}, a) = Q(s_{t+1}, a). \quad (3)$$

Substituting the above formula (3) in the formula (2), we obtain:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \cdot (Q(s_{t+1}, a_t) - Q(s_t, a_t)). \quad (4)$$

If the learning rate α is set to 1 (in practice, often a constant learning rate is used [4]), then $\lambda = \gamma$, the above formula (4) can be written as:

$$Q(s_t, a_t) = Q(s_t, a_t) = r_t + \gamma \cdot \max_a Q(s_{t+1}, a_t). \quad (5)$$

After that, we store the result value, and the next action a_{t+1} is selected by maximizing the *Q* function for the state s_{t+1} by *a*:

$$a_{t+1} = \arg\max_a Q(s_{t+1}, a). \quad (6)$$

This is the main implementation of our *Q*-learning algorithm.

Q-learning algorithm is implemented using the following steps:

1) Receive the output voltage and its derivative, as well as feedback from the environment.

2) Variable sampling.

3) Update the *Q* matrix according to formula (5).

4) Then select the next action according to formula (6).

The above processing algorithm will be executed once for each time step, and then the matrix $Q_i(s, a)$ will gradually converge to $Q^*(s, a)$.

PID Controller

The PID controller (proportional-integrating-differentiating controller) consists of a proportional term *P*, an integrating term *I* and a differentiating term *D*. The purpose of control is to make the process variable *y* follow the set-point value *r*. The error *err*(*t*) is defined by $e = r - y$. Term *P* is proportional to the error at the instant *t*. Term *I* is proportional to the integral of the error up to the instant *t*. Term *D* is proportional to the derivative of the error at the instant *t* [5]. Three proportional parameters are configured - K_p, K_i and K_d.

A PID controller is a common feedback component in industrial control systems. The feedback value is determined by the following formula:

$$U(t) = K_p(err(t) + \frac{1}{T_I}\int err(t)dt + \frac{T_D derr(t)}{dt}). \quad (7)$$

Diagram of a voltage regulator and their mathematical description:

Switch mode power supply can have two work conditions, according to the conditions of switch. Circuit diagram of Switch mode power supply when the circuit is closed is shown in Fig. 1.

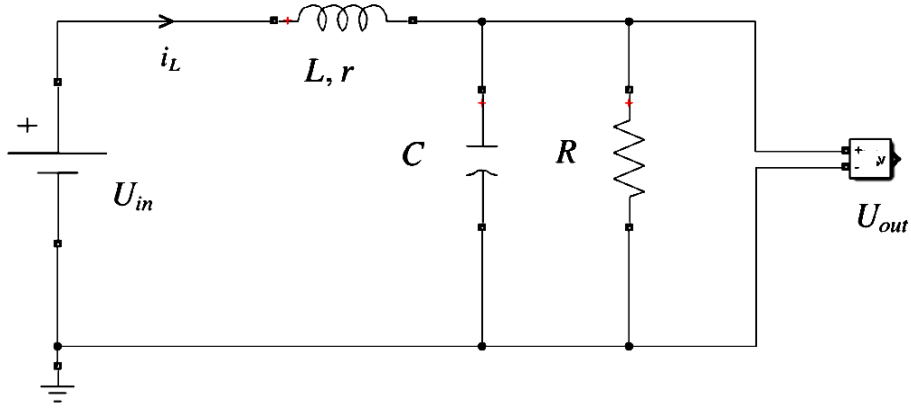

Fig. 1. Switch mode power supply

The work of this circuit can be described by the following differential equations:

$$\frac{di_L}{dt} = -\frac{r}{L}i_L - \frac{1}{L}U_{out} + \frac{1}{L}U_{in}, \frac{dU_{out}}{dt} = \frac{1}{C}i_L - \frac{1}{RC}U_{out}. \quad (8)$$

When the circuit is open, it can be converted to the following two cases (see fig. 2).

The left figure in Fig. 2 presents an equivalent circuit diagram, when the voltage of the inductor $i_L > 0$, the equation is as follows:

$$\frac{di_L}{dt} = -\frac{r}{L}i_L - \frac{1}{L}U_{out}, \frac{dU_{out}}{dt} = \frac{1}{C}i_L - \frac{1}{RC}U_{out}. \quad (9)$$

The right figure in Fig. 2 is the equivalent circuit diagram when $i_L = 0$. Accordingly, we can describe this with the equation:

$$\frac{dU_{out}}{dt} = -\frac{1}{RC}U_{out}. \tag{10}$$

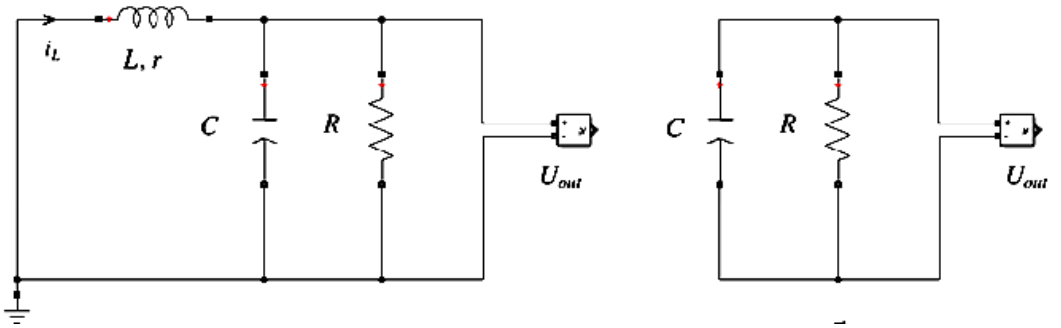

Fig. 2. Equivalent schemes of two cases of closed operation

At the same time, the current i_L can be eliminated from the two circuit equations (8), (9) and (10), and we can obtain the output voltage equations of the voltage regulator in Fig.1 and Fig.2:

When switch closed:

$$\frac{d^2U_{out}}{dt^2} + a_1\frac{dU_{out}}{d_t} + a_0U_{out} = b_0U_{in}. \tag{11}$$

When switch open:

$$\frac{d^2U_{out}}{dt^2} + a_1\frac{dU_{out}}{d_t} + a_0U_{out} = 0. \tag{12}$$

In which:

$$a_1 = \frac{L+RCr}{RLC}; a_0 = \frac{R+r}{RLC}; b_0 = \frac{1}{LC}.$$

Parameters in the circuit:

– capacitor capacity: C=0.0001 F;

– load resistance: R=10 Ω;

– coil inductance: L=10 H;

– coil internal resistance: r=0Ω.

Now after solving the differential equations (11) and (12) we can get the output voltage, which is out target control parameter. The target parameter for controlling the output voltage is set to 50 V, and the input voltage is set to a square wave, which varies between 100 ± 25 V. The total sampling time is set to 25 seconds, and the sampling step is 0.001 s. After modeling in Python, set other parameters and then analyze in the appropriate situation.

Results

After adjusting the input parameters, we got 3 groups of results, which are shown in fig. 3-5. Some comparations of control effect between PID and RL, and result data are in the tables. 1-3.

- Input signal range: 100 ± 25 *V*; Target value of the output signal: 50 *V*.

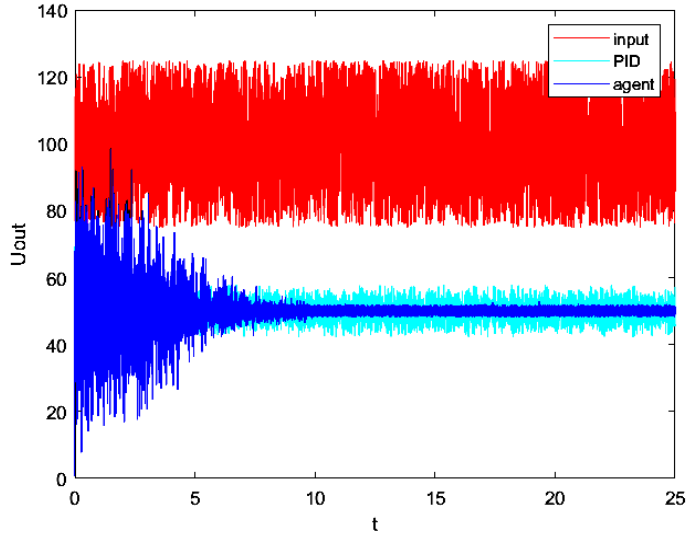

Fig. 3

- Input signal range: 100 ± 50 *V*; Target value of the output signal: 50 *V*:

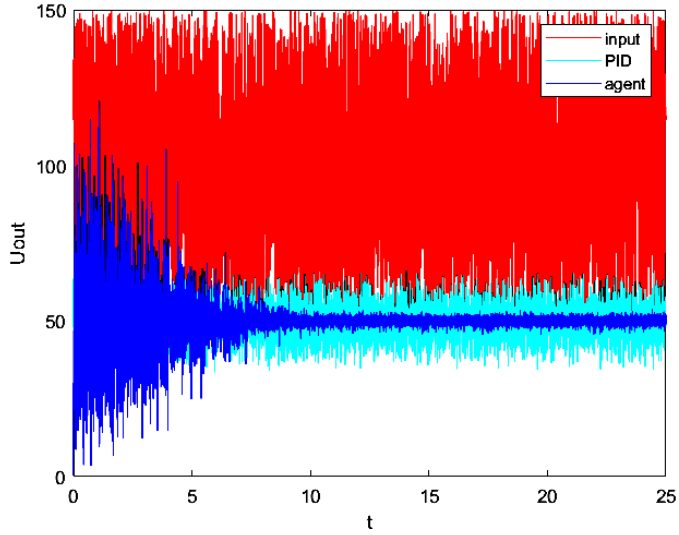

Fig. 4

Tab. 1

	Agent	PID-regulator
Average	49.79	50.01
Dispersion	0.4356	0.928
Standard deviation	0.66	0.9633
The number of times the signal exceeded $\|U_{out}-50\|>5$	0	42
The time when the signal went beyond $\|U_{out}-50\|>5$	0.0	0.0042

Tab. 2

	Agent	PID-regulator
Average	49.95	50.0
Dispersion	0.929	3.296
Standard deviation	0.9638	1.815
The number of times the signal exceeded $\|U_{out}-50\|>5$	0	360
The time when the signal went beyond $\|U_{out}-50\|>5$	0.0	0.036

-Input signal range: 150 ± 50 *V*; Target value of the output signal: 50 *V*:

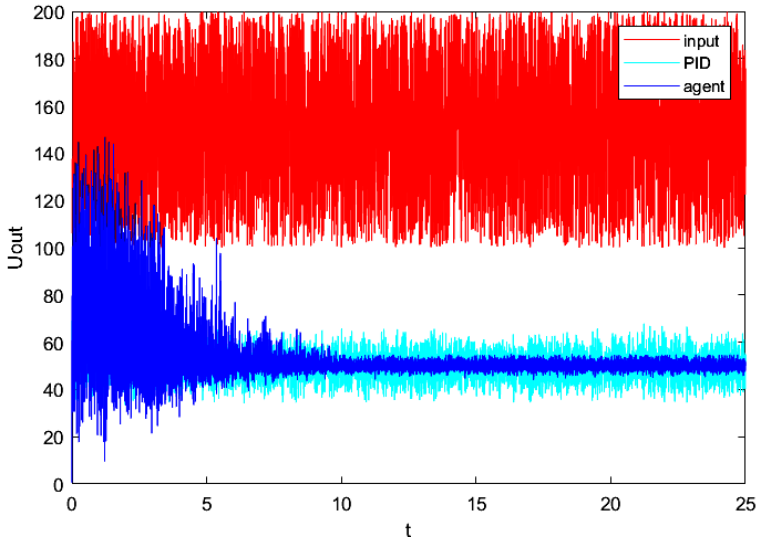

Fig. 5

Tab. 3

	Agent	PID-regulator
Average	49.78	50.0
Dispersion	1.432	4.275
Standard deviation	1.197	2.068
The number of times the signal exceeded $\|U_{out}-50\|>5$	0	498
The time when the signal went beyond $\|U_{out}-50\|>5$	0.0	0.05

We can see that when the input voltage is farther from our target value of the output voltage, the reinforcement learning control method enhances the excellent ability to output a smooth waveform. The dispersion is only one third of the PID control method. After the initial learning process, the waveform under the control of reinforcement learning will no longer fluctuate by more than ± 5 *V*, and the area outside of ± 5 *V* remains in the PID controller.

Conclusions

1. Developed models of controlled filtering of the electrical signal in AC voltage regulators based on the reinforcement learning method and on the basis of the PID regulator.

2. On the basis of the developed models, a comparative analysis of the quality indicators for controlling the voltage regulator using the reinforcement learning algorithm and the PID controller was carried out.

3. As a result of the comparative analysis, it was found that for a number of qualitative indicators, the reinforcement learning algorithm when solving the stabilization problem of the output voltage of a voltage regulator significantly exceeds the traditionally PID controllers, which indicates significant prospects for the application of a reinforcement learning algorithm in the management of technical objects.

References / Список литературы

1. Kaelbling L.P., Littman M.L., Moore A.W. Reinforcement Learning: A Survey, Journal of Artificial Intelligence Research, 1996, 4: 237-285.
2. Watkins C.J.C.H. Learning from Delayed Rewards, 1989, Cambridge University.
3. Matiisen T. Demystifying Deep Reinforcement Learning, 2015, Computational Neuroscience Lab.
4. Barto A.G., Sutton R.S. Reinforcement Learning: An Introduction, 1988. Bradford Book. Boston.
5. Araki M. PID Control, 2003.

Тао Сынянь – магистр, Московский государственный университет им. М.В. Ломоносова, Москва, Россия, 86taosinian@gmail.com	**Tao Sinian** – Master, M.V. Lomonosov Moscow State University, Moscow, Russia, 86taosinian@gmail.com
Смирнов Алексей Сергеевич – кандидат физико-математических наук, Московский государственный Университет, им. М.В. Ломоносова, Москва, Россия, Alexeysmirnov@yandex.ru	**Smirnov Alexey Sergeevich** – candidate of physical and math sciences, M.V. Lomonosov Moscow State University, Moscow, Russia, Alexeysmirnov@yandex.ru
Асташев Михаил Георгиевич - кандидат технических наук, доцент, Национальный исследовательский университет «МЭИ», Москва, Россия. AstashevMG@mpei.ru	**Astashev Michael Georgievich** – candidate of technical sciences, associate professor, National research university “MPEI”, AstashevMG@mpei.ru

Received 20.09.2020