http://dx.doi.org/10.5755/j02.eie.28875 ELEKTRONIKA IR ELEKTROTECHNIKA, ISSN 1392-1215, VOL. 27, NO. 3, 2021

Optimizing of Q-Learning Day/Night Energy Strategy for Solar Harvesting Environmental Wireless Sensor Networks Nodes

Michal Prauzek*, Jaromir Konecny
Faculty of Electrical Engineering and Computer Science, VSB - Technical University of Ostrava, Ostrava, Czech Republic
michal.prauzek@vsb.cz

***Abstract*—This research article presents the application of the Q-learning algorithm in the operational duty cycle control of solar-powered environmental wireless sensor network (EWSN) nodes. Those nodes are commonly implemented as embedded devices using low-power and low-cost microcontrollers. Therefore, there is a significant need for an effective and easy way to implement a machine learning (ML) algorithm in terms of computer performance. This approach uses a Q-learning-based policy implementing a sleep/run switching algorithm driven by the state of charge. The presented algorithm is based on two modes: daylight and nighttime, which is a suitable solution for solar-powered systems. The study includes the complete process of design EWSN node strategy with an optimal reward policy. The presented algorithm was tested and verified on an EWSN node model and a 5-year data set of solar irradiance values was used for the learning process and its validation. As part of the study, we are also presenting the validation in terms of Q-learning parameters, which include the learning rate and discount factor. The result section shows that the overall performance of the presented solution is more suitable for solar-powered EWSN then state-of-the-art studies. Both day/night experiments reached 828 203 measurement/transmission cycles, which is 12.7 % more than in the previous studies using the strategy defined by the state of energy storage.**

***Index Terms*—Energy management; Microcontrollers; Semi-supervised learning; Wireless sensor networks.**

I. INTRODUCTION

The application of machine learning (ML) methods is becoming more popular, and the number of applications is increasing. Various ML approaches use a scalable level of computer performance. We are focusing on the implementation of ML methods in low-performance embedded systems. The field of embedded intelligence (EI) has several challenges [1], including small data, power and energy consumption, wireless communication constraints, etc. Each challenge represents an important research opportunity in embedded device operations.

Many kinds of embedded devices could be driven by ML, but in this contribution, we are focusing on environmental wireless networks (EWSNs) [2]. EWSNs are equipped with a low-performance and low-cost microcontroller, sensors, powering module [3], [4], data transmission interface, and data storage (see Fig. 1). Their main purpose lays in environmental data and parameter collection and transmission [5]. In this research paper, we are focusing on energy harvesting EWSN with local energy storage [6] operated by the Q-learning method.

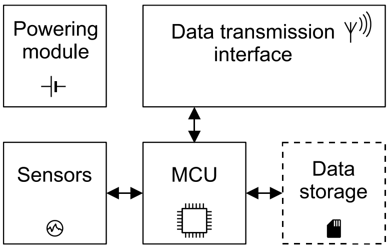

Fig. 1. Block diagram of EWSN node. The microcontroller (MCU) runs the control algorithm.

The Q-learning method can be used to control an EWSN node that obtains parameters from the environment and sends data to the Internet via a wireless interface [7]. The system obtains energy for its operation from a solar panel, and in this context, it must optimize its behavior according to the harvested energy [8]. In this experiment, Q-learning will be used to optimize the run/sleep duty cycle to maximize the obtained data.

In this contribution, we are presenting the follow-up research based on the previous paper [9]. This past research presented the design of a hybrid energy management strategy with Q-learning control during daylight and a linear discharging process at night. Now, we are presenting a method to optimize Q-learning parameters to achieve the best and stable performance results.

This article is organized into five sections. The first section introduces the article with a brief description of the state of the art. The background section brings important information about the EWNS node model, input data description, and the basic theory of the Q-learning method. Section III describes our controller solution and the

Manuscript received 5 December, 2020; accepted 22 May, 2021.

This work was supported by the project No. SP2021/29, “Development of algorithms and systems for control, measurement and safety applications VII” of the Student Grant System, VSB-TU Ostrava. This work was also supported by the European Regional Development Fund in the Research Centre of Advanced Mechatronic Systems’ project, project No. CZ.02.1.01/0.0/0.0/16_019/0000867 within the Operational Programme Research, Development and Education. This work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 856670.

designed experiment. The results section presents the data obtained from the experiment and details on the various settings of the learning procedure, and it also brings a technical discussion about the results. The final section concludes the article and shows outlines for future work.

II. BACKGROUND

In this section, we present the hardware model of the EWSN device and input data general description. In addition, we provide the basic terms of the Q-learning method here.

A. *EWSN Node Model and Data*

In this work, we are using a previously described energy harvesting EWSN node [10] depicted in Fig. 2 as the reference for all model parameters. The EWSN node contains an ARM Cortex-M0+ microcontroller, a solar panel charging module, energy storage represented by two serial-connected super-capacitors, a data storage module (EEPROM and SD-card), and a low-lower wireless interface (IEEE 802.15.4). The complete hardware specification of this EWSN node is detailed in a previous publication [11].

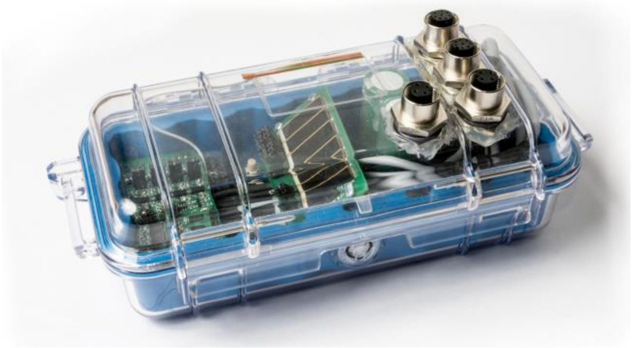

Fig. 2. Hardware implementation of EWSN node in IP68 waterproof box.

The proposed experiment uses 5 years of data from the Fairview Agricultural Drought Monitoring station (AGDM) located in Alberta, Canada [12], coordinates at 56.0815 ° latitude, -118.4395 ° longitude, and 655.00 m elevation. This data set contains the total incoming solar radiance in W/m^2 measured per each five-minute interval continuously from 2008 to 2012. We are using this time interval due to the easy quantitative comparation to previous studies.

B. *Q-Learning Method*

The Q-learning method belongs to the group of reinforcement learning (RL) methods [13]. The Q-learning algorithm is a suitable candidate, which can control an embedded system represented by an WSN node because it is not very extensive in terms of computer performance. This approach uses a properly defined policy, and it could be used to perform the optimal EWNS node management of data collection and wireless transmission.

The Q-learning algorithm is detailed in publication [14]. Q-learning is a RL algorithm and it belongs to the family of semi-supervised model-free methods. The mathematical formalization of the decision problem, which consists of the states of a system *S*, the performed actions *A*, and the rewards *R*, is known as the Markov decision process [15].

The implementation of the Q-learning method uses a Q-table implemented as a data array stored in a memory, where each column represents the quantitative value of the actions. The size of the Q-table is defined by the number of states and the number of actions [16]. The basic idea of the Q-learning algorithm is to estimate the future reward represented by $Q(S_t, A_t)$ for the performed action *A* in the state *S* and at the same time follow the optimal internal policy [15].

The Q-learning is generally described as follows

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \\ +\alpha\left[R + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\right], \tag{1}$$

where $Q(S_t, A_t)$ represents an estimate of the reward value in the Q-table for the currently performed action A_t and state S_t. The learning rate is represented by α parameter. If $\alpha = 0$, the algorithm uses only the knowledge gained from the previous steps; if $\alpha = 1$, the algorithm uses only new knowledge. *R* represents learning feedback in the form of the current reward. The γ parameter is the discount factor that determines whether the current reward ($\gamma = 0$) or cumulative reward ($\gamma = 1$) is preferred.

III. EXPERIMENT

In this simulation-based study, the Q-learning method is used to control the operation of the EWSN node model. The system obtains energy for its operation from the solar panel; therefore, it is appropriate to optimize node behavior according to the available energy. The aim of the optimization is to control duty cycle of the EWSN node and achieve the maximum amount of collected data. This optimization goal directly corresponds with the elimination of overcharging energy storage, in other words, the node should utilize the maximum available energy. At the same time, there is a requirement for the continuous operation of the device without failure due to a lack of energy. The device must have enough power for not failing during sleep or the operation itself. Moreover, EWSN node should have enough energy for nighttime proper function at the end of daytime.

When we are designing a control algorithm that uses Q-learning, it is necessary to design individual actions that will be controlled by the algorithm, the states in which the system can be, and a reward policy. Since there is no energy available at night to charge the supercapacitor, the optimization algorithm is divided into two different modes. The Q-learning algorithm is active in daylight mode and a different strategy is applied in nighttime mode.

A. *Daylight Control Strategy*

Daylight mode is active from sunrise to sunset. For this reason, it is necessary to determine when the sun rises and sets. This can be determined based on a real-time clock implementation and the geographical deployment location. A detailed calculation of sun position according to geographical position can be found on the website [17]. It is possible to implement this calculation directly in the microcontroller or the look-up table for a given day can be used to determine the sunrise and sunset.

The algorithm for the daylight mode is based on the intensity of sunlight over the past thirty minutes. This information can be obtained from the difference between the amount of energy in the supercapacitor represented by the state of energy storage (SoES) and the amount of energy consumed by the EWSN node activity. The consumed energy can be calculated as the energy required for one action times the number of actions plus the static consumption of the EWSN node. The solar ratio (SR) can be calculated by the following equation

$$\mathrm{SR} = \frac{\mathrm{pastEnergy}(t)}{\mathrm{energyMax}}, \tag{2}$$

where pastEnergy(t) is the total amount of energy harvested over the past 30 minutes, energyMax is the maximum solar energy that can be harvested in a 30-minute interval under ideal conditions (clear sky without any clouds) for the given deployment location. The selected state is then defined by the *SR* range (Table I).

TABLE I. EWSN NODE STATES DEFINED BY $\mathrm{SR}_{\mathrm{MIN}}$ AND $\mathrm{SR}_{\mathrm{MAX}}$.

State (*S*) No.	$\mathbf{SR_{min}}$	$\mathbf{SR_{max}}$
1	0	0.01
2	0.01	0.03
3	0.03	0.05
4	0.05	0.1
5	0.1	0.15
6	0.15	0.2
7	0.2	0.3
8	0.3	0.4
9	0.4	0.5
10	0.5	0.6
11	0.6	1

The defined of the *SR* parameter ranges (Table I) provide high resolution for very low solar irradiation (states 1 to 3), medium resolution for average solar irradiation (states 4 to 6), and low resolution for ideal sunlight conditions (states 7 to 10). State 11 is designed for ideal sunlight conditions (more than 60 % of the maximum solar irradiation).

Next, the reward policy should be established. The Q-learning agent will charge the supercapacitor in several steps until the energy storage reaches full charge; therefore, a charging step is defined for a 30-minute interval

$$\mathrm{chStp}(t) = \frac{\mathrm{SoES}_{\mathrm{max}} \times \mathrm{dayTarget}}{\mathrm{dayStepNumber}}, \tag{3}$$

where $\mathrm{SoES}_{\mathrm{max}}$ is the energy storage capacity in joules, dayTarget is the optimal daily charge ratio (60 %), and dayStepNumber is the number of control algorithm steps during daylight mode.

Due to the physical maximum of the SoES, the daily goal setting for the charge target *chT* is adjusted for values close to the total capacity of the energy storage $\mathrm{SoES}_{\mathrm{max}}$:

$$\mathrm{SoES}_{\mathrm{diff}}(t) = \mathrm{SoES}_{\mathrm{max}} - \mathrm{SoES}(t), \tag{4}$$

$$chT = \begin{cases} \mathrm{chStp} & : \ \mathrm{chStp} < \mathrm{SoES}_{\mathrm{diff}}(t), \\ \mathrm{SoES}_{\mathrm{diff}}(t) : & \ \mathrm{chStp} \geq \mathrm{SoES}_{\mathrm{diff}}(t). \end{cases} \tag{5}$$

The daily goal in this experiment is very closely related to the reward policy. The reward policy consists of three basic parts. The first part of the reward $\mathrm{R_A}(S, A)$ *r*epresents the reward for performing the action. The reward strategy considers the need to collect data from the environment by the EWSN node. The agent receives the highest reward for the shortest operational period (1 minute) and the lowest reward for the longest period (30 minutes). This component is described in Table II.

TABLE II. $\mathrm{R_A}$(S, A) REWARD COMPONENT CALCULATION.

Action (min)	1	2	3	5	10	15	30
Reward	1	0.9	0.8	0.7	0.6	0.5	0.4

The second part of the reward $R_{\mathrm{T}}(S, A)$ represents the fulfillment of the charging target. This reaches the maximum value of 1 when the agent fulfills the selected target. When the agent starts overcharging, this reward $R_{\mathrm{T}}(S, A)$ decreases. Decreasing the reward value $R_{\mathrm{T}}(S, A)$ on recharging is limited when recharging by more than 90 % of the charging target. In this case, the reward $R_{\mathrm{T}}(S, A)$ is set to 0.1. Overcharging is less of a problem than a lack of energy, so the reward $R_{\mathrm{T}}(S, A)$ for meeting the goal is always positive. If the agent does not meet the target due to slow charging, the reward $R_{\mathrm{T}}(S, A)$ is also less than one. The reward value $R_{\mathrm{T}}(S, A)$ for a charging goal is expressed by the following equation

$$R_{\mathrm{T}}(S, A) = 1 - \left| \frac{chT - [\mathrm{SoES}(t) - \mathrm{SoES}(t - T)]}{chStp} \right|. \tag{6}$$

The last part of the reward is a penalty for a lack of energy in the EWSN node. If the state of the charge falls below the minimum amount, $R_{\mathrm{F}}(S, A)$ is set to -0.5. The mathematical formalization of this reward component is formulated as

$$R_{\mathrm{F}}(S, A) = \begin{cases} 0 & : \ \mathrm{fail}_{\mathrm{Status}} \text{ is False}, \\ -0.5 & : \ \mathrm{fail}_{\mathrm{Status}} \text{ is True}. \end{cases} \tag{7}$$

The total reward $R(S, A)$ consists of all three basic components

$$R(S, A) = R_{\mathrm{A}}(S, A) \times R_{\mathrm{T}}(S, A) + R_{\mathrm{F}}(S, A). \tag{8}$$

The reward for the action $R_{\mathrm{A}}(S, A)$ is multiplied by the reward for fulfilling the goal $R_{\mathrm{T}}(S, A)$, and finally, the penalty for energy failure $R_{\mathrm{F}}(S, A)$ is added. The reward for the action and the reward for meeting the goal are multiplied, as this balances the strategy of maximizing the number of measurements and meeting the charging goal.

B. *Nighttime Control Strategy*

At night, when solar energy is not available, the policy of the control algorithm is set differently than during the daylight period. The device is discharged linearly depending on the current state of the energy storage, the remaining length of the night, and the estimated energy required for one operational cycle.

The consumption estimate for one measuring cycle is determined as follows

$$E_{\text{cycle}} = \text{SoES}(t-T) - \text{SoES}(t). \tag{9}$$

The discharge target for the night is 20 % of the total storage capacity. This value is a reserve for the beginning of the next daylight period. The remaining energy for the night is calculated

$$E_{\text{rest}} = \text{SoES}(t) - \text{SoES}_{\max} \times 0,2. \tag{10}$$

If there is a lack of energy, the period is set at 30 minutes as a default. If there is more than 20 % of the energy remaining in the energy storage, the number of cycles until sunrise is determined as follows

$$\text{operationCount} = \frac{E_{\text{rest}}}{E_{\text{oneCycle}}}. \tag{11}$$

The resulting operational period at night is then determined as

$$T_{\text{night}} = \frac{\text{timeUntilSunrise}}{\text{operationCount}}. \tag{12}$$

The period upper limit is also set to 30 minutes to maintain the maximum measurement period. It is obvious that this calculation underestimates the night discharge target and the total discharge of the EWSN node.

IV. RESULTS

The experiment simulation was performed using 1 763 days of historical solar radiation data described in the background section. The Q-learning algorithm was learned in 30-minute iteration steps.

A. Control Strategy Results

The count of selected actions during the sample summer days is shown in Fig. 3. The controller selects a 1-minute period during short summer nights when sufficient energy is available in the energy storage. During the day, the algorithm considers various states of incoming solar energy to select the appropriate action that would lead to the greatest possible reward.

The total count during the winter period is shown in Fig. 4. In the nighttime mode, the controller has selected a suitable operational period so that the data collection is properly distributed during a night. During short winter days, the controller selects the appropriate actions, however, with a longer duty cycle than during the long daylight period in the summer.

Table III shows an overview of the simulation results and their comparison with previous state-of-the-art studies performed on the same environmental data [11], [18]. The current day/night-based simulation experiments use Q-learning with the following parameters: Learning rate 0.1 and 0.68, discount factor 0.1, and epsilon greedy strategy fixed to 0.05. The EWSN node model uses two types of actions - measurement and transmission (M and T). Both day/night experiments are shown in Table III, and the best configuration (alpha 0.1) reaches 828 203 measuring transmission cycles, which is 12.7 % more than in the previous study [18] using the SoES strategy. A simple timer-based controller designed in an earlier study [11] performed more measurement cycles, but in the presence of a significant transmission delay caused by low transmission frequency. The current approach fails in 161 cases. Although this number is higher than in the previous study [11], it should be noted that the study [11] allowed the Q-learning algorithm to use up to a 60-minute period (compared to the 30-minute operational period in this study).

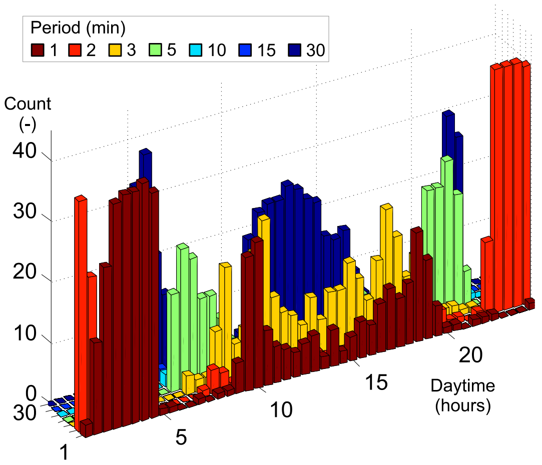

Fig. 3. Summer period: Total count of selected actions during 40 days in summer (0–24 hours).

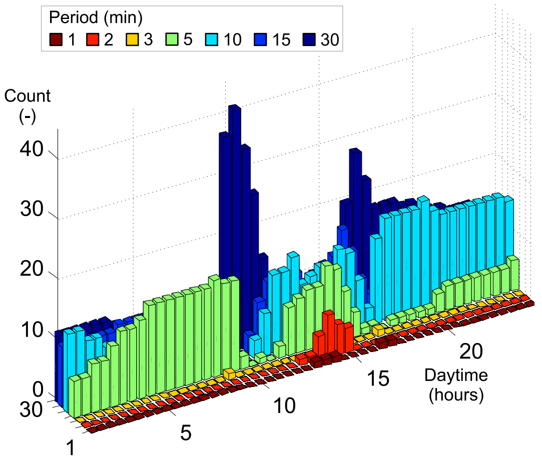

Fig. 4. Winter period: Total count of selected actions during 40 days in winter (0–24 hours).

TABLE III. RESULT COMPARISON WITH STATE-OF-THE-ART METHODS.

Method	Day/Night		SoES Strategy		Timer
Learning	0.1	0.68	0.68	0.9	-
Step	1 800	1 800	60	60	60
Fails	161	151	0	0	1 669
ABS M	828 203	733 653	735 674	712 126	899 391
ABS T	828 203	733 653	735 674	712 126	17 143
AVG M	469.8	438.8	417.3	403.9	510.1
AVG T	469.8	438.8	417.3	403.9	9.7

Figure 5 shows the internal states of the EWNS node in the summer period, including solar irrandiance, SoES, and operational periods selected by Q-learning. In the summer, there is sufficient energy income, therefore a short operational interval is selected. There are also significant decreases in incoming solar energy due to cloudy weather, where the algorithm selects longer operational periods. In winter days (see Fig. 6), the lack of incoming energy is significant and the algorithm selects longer operational periods rather than shorter. Unfortunately, the ESWN node can fail at night due to the fact that at the end of the day it does not have enough energy for the algorithm to be able to work even for the longest operational period.

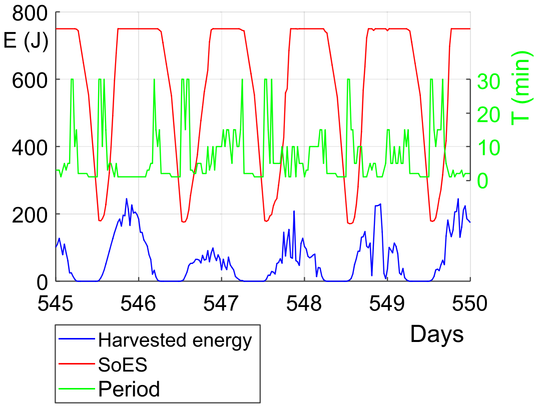

Fig. 5. Summer period (5 days): Obtained energy, state of energy storage (SoES), EWSN operational period.

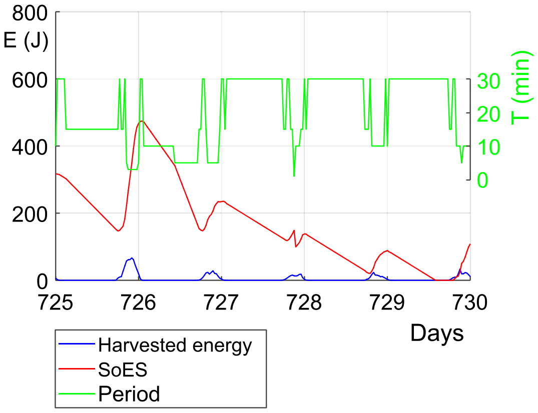

Fig. 6. Winter period (5 days): Obtained energy, state of energy storage (SoES), EWSN operational period.

B. *Optimization Results*

To find the optimal parameters of the learning settings, an extensive set of experiments was performed. The algorithm has been tested for various α and γ settings. The results of the individual simulations for the parameters α and γ are variable due to the stochastic nature of the learning process (epsilon greedy policy).

This variability can be suppressed by a multiple performance simulation. Thus, the following experiments were performed 100 times for each of the α and γ settings, and the resulting figures show the average results of these experiments.

Figure 7 shows the dependence of the average total number of operational cycles on the parameters α and γ. This figure shows just minor changes of the operational cycle number and the best results are obtained for α = 0.1. The limit value α = 0 leads to a significant decrease in the number of data collecting cycles. The algorithm is not learning in this case.

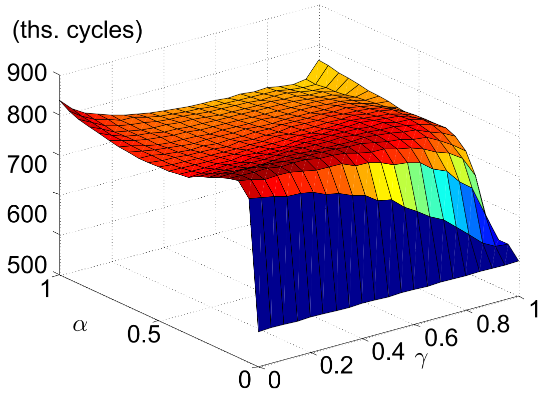

Fig. 7. Average of operational cycle count in various settings α and γ (epsilon is fixed to 5 %).

Figure 8 shows the dependence of the average total number of EWSN node failures based on parameters α and γ. It can be seen that for the discount factor γ = 1 and α = 1, the cumulative reward is maximized. Unfortunatelly it leads to node failure in most cases.

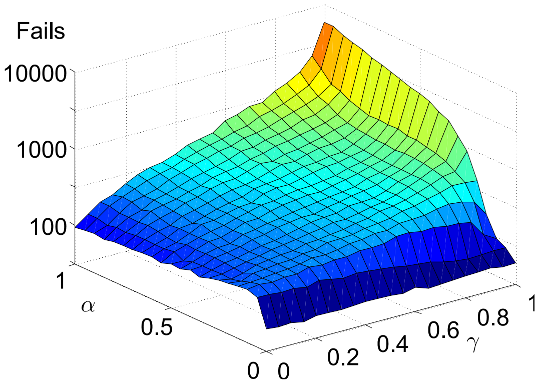

Fig. 8. Average of failure count in various settings α and γ (epsilon is fixed to 5 %).

Figure 9 shows the dependence of the average total number of overcharges on the parameters α and γ. Overcharging is generally described as not utilizing incoming solar energy (energy storage is full). For the limit state α = 0 (the algorithm is not learning and behaves only randomly), the number of overcharges increases. At the same time, if the algorithm considers cumulative rewards, it is not able to utilize the available energy. If the algorithm uses only new knowledge and adapts more quickly to the environment, the number of overcharges can be reduced, and the available energy is used more efficiently.

Figure 8 and Figure 9 show the dependence of the evaluation parameters on the learning parameters α and γ. To compile the overall evaluation criterion, it is necessary to combine and weigh these parameters appropriately. In addition, it should be noted that the total number of operational cycles needs to be maximized and the total number of failures and the total number of overcharges need to be minimized.

The total number of cycles can be normalized according to the following formula

$$\mathrm{CC}_{\mathrm{norm}}(\alpha,\gamma) = \frac{\mathrm{countCycles}(\alpha,\gamma)}{\max\limits_{\alpha,\gamma}\{\mathrm{countCycles}(\alpha,\gamma)\}}. \qquad (13)$$

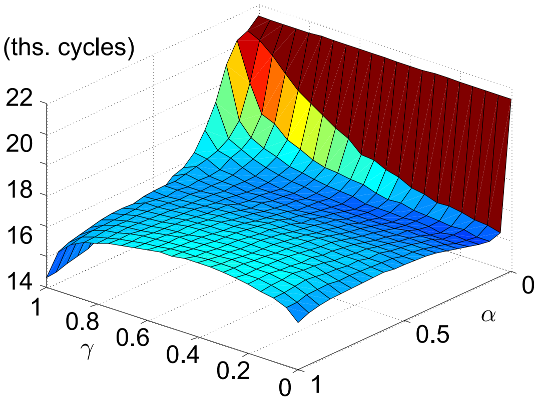

Fig. 9. Average of overcharge count in various settings α and γ (epsilon is fixed to 5 %).

The number of recharges can be normalized in a similar way, but the meaning needs to be reversed

$$\mathrm{CO}(\alpha,\gamma) = 1 - \frac{\mathrm{countOverchrg}(\alpha,\gamma)}{\max\limits_{\alpha,\gamma}\{\mathrm{countOverchrg}(\alpha,\gamma)\}}. \tag{14}$$

The calculation for the number of errors was chosen by the logarithm because the difference between the minimum and maximum error values is too large. The meaning must be reversed because the number of errors needs to be minimized

$$\mathrm{CF}(\alpha,\gamma) = 1 - \frac{\log[\mathrm{countFail}(\alpha,\gamma)]}{\max\limits_{\alpha,\gamma},\{\log[\mathrm{countFail}(\alpha,\gamma)]\}}. \tag{15}$$

The overall rating can then be calculated as

$$\begin{aligned}\mathrm{E}(\alpha,\gamma) &= k_1 \times \mathrm{CF}(\alpha,\gamma) + \\ &+ k_2 \times \mathrm{CO}(\alpha,\gamma) + k_3 \times \mathrm{CC}(\alpha,\gamma),\end{aligned} \tag{16}$$

where k_1, k_2, and k_3 are the weight coefficients of the overall rating, where $k_1 + k_2 + k_3 = 1$.

Figure 10 and Figure 11 show the overall evaluation of the algorithm efficiency depending on the parameters α and γ. The only difference between these figures is the choice of weighting coefficients. The graphs basically show the two areas in which the algorithm works most efficiently.

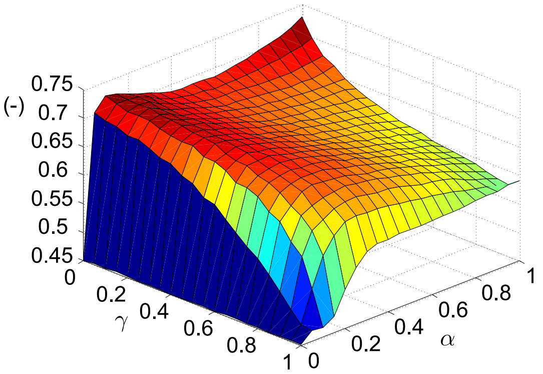

Fig. 10. Graphical representation of the overall evaluation of the algorithm efficiency depending on the parameters α and γ. Weight configuration k_1 = 0.2, k_2 = 0.2, and k_3 = 0.6.

The first area is the limit state of a fast learning algorithm, which is the most interested in the instant reward α = 1 and γ = 0. The second area is a more conservative approach, which learns more slowly, and at the same time, favors immediate rewards α = 0 and γ = 0. The stable area of the optimal algorithm operation is therefore the area near α = 0.1 and γ = 0.1.

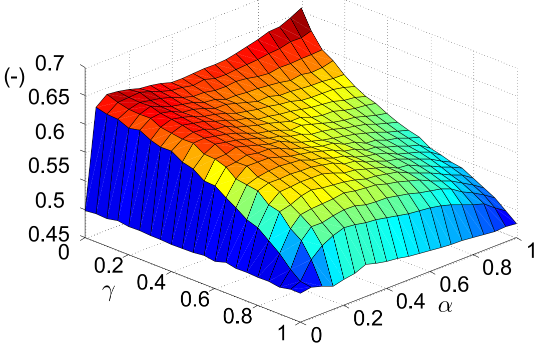

Fig. 11. Graphical representation of the overall evaluation of the algorithm efficiency depending on the parameters α and γ. Weight configuration k_1 = 0.4, k_2 = 0.1, and k_3 = 0.5.

V. DISCUSSION

In this contribution, we presented an algorithm based on the Q-learning method for an EWSN node. This principle can control the energy management of a solar-powered system, and it could work as a self-learning system directly on a deployment site. This study brings a novel algorithm based on two modes for daylight and nighttime. Such an approach allows a suitable solution for energy harvesting EWNS deployed at various locations.

VI. CONCLUSIONS

The experimental result based on the hardware model shows that this solution can work properly. Compared to previous studies, this approach shows a better result in terms of the number of operational cycles. The best configuration (alpha 0.1) reaches 828 203 measuring transmission cycles, which is 12.7 % more than the previous study. In addition, the overcharge and failure stability are sufficient, and the application can work without a significant loss of collected data. This study presents results for various settings of Q-learning parameters (learning rate and discount factor) with a fixed epsilon-greedy policy. The performed test proves that the configuration of the learning rate of 0.1 and a discount factor of 0.1 leads to the best performance result with good stability in terms of the unutilized energy and count of failure state.

VII. FUTURE WORK

There are several research opportunities for future work. The first opportunity includes extensive testing in various locations. The system may behave differently when the weather is more stable or unstable. Therefore, the future testing should be performed with data from various climatic regions of the world. The second research challenge lays in the modification of the presented algorithm to automatically detect the transition from day to night. Such solution could make it possible with elimination of the need for a real-time clock implementation including time synchronization. Also, the information about the location will be not needed in this case. The final challenge, which could replace the second one, includes an algorithm modification as a single approach by Q-learning for all day. This challenge needs a major revision of the reward policy published in this article.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] R. P. Dick, L. Shang, M. Wolf, and S.-W. Yang, "Embedded intelligence in the Internet-of-Things", *IEEE Design & Test*, vol. 37, no. 1, pp. 7–27, Feb. 2020. DOI: 10.1109/MDAT.2019.2957352.

[2] A. de la Piedra, F. Benitez-Capistros, F. Dominguez, and A. Touhafi, "Wireless sensor networks for environmental research: A survey on limitations and challenges", in *Proc. of Eurocon 2013*, Zagreb, 2013, pp. 267–274. DOI: 10.1109/EUROCON.2013.6624996.

[3] V. Markevicius, D. Navikas, D. Andriukaitis, M. Cepenas, A. Valinevicius, M. Zilys, R. Malekian, A. Janeliauskas, W. Walendziuk, A. Idzkowski, "Two thermocouples low power wireless sensors network", *AEU - International Journal of Electronics and Communications*, vol. 84, pp. 242–250, 2018. DOI: 10.1016/j.aeue.2017.11.032.

[4] D. K. Sah and T. Amgoth, "Renewable energy harvesting schemes in wireless sensor networks: A survey", *Information Fusion*, vol. 63, pp. 223–247, 2020. DOI: 10.1016/j.inffus.2020.07.005.

[5] L. Liang, X. Yang, L. Zhang, D. Gao, and H. Zhang, "Issues for event monitoring in event-driven wireless sensor networks", in *Proc. of 2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*, Wuhan, 2011, pp. 1–4. DOI: 10.1109/wicom.2011.6040345.

[6] M. Prauzek, J. Konecny, M. Borova, K. Janosova, J. Hlavica, and P. Musilek, "Energy harvesting sources, storage devices and system topologies for environmental wireless sensor networks: A review", *Sensors*, vol. 18, no. 8, p. 2446, 2018. DOI: 10.3390/s18082446.

[7] K.-L. A. Yau, H. G. Goh, D. Chieng, and K. H. Kwong, "Application of reinforcement learning to wireless sensor networks: Models and algorithms", *Computing*, vol. 97, no. 11, pp. 1045–1075, 2015. DOI: 10.1007/s00607-014-0438-1.

[8] Y. Li, E. A. Hamed, X. Zhang, D. Luna, J.-S. Lin, X. Liang, and I. Lee, "Feasibility of harvesting solar energy for self-powered environmental wireless sensor nodes", *Electronics*, vol. 9, no. 12, p. 2058, 2020. DOI: 10.3390/electronics9122058.

[9] M. Prauzek, J. Konecny, J. Hlavica, and P. Musilek, "Self-learning for day-night mode energy strategy for solar powered environmental WSN nodes", in *Proc. of 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, London, ON, Canada, 2020, pp. 1–5. DOI: 10.1109/CCECE47787.2020.9255790.

[10] M. Prauzek, P. Musilek, and A. G. Watts, "Fuzzy algorithm for intelligent wireless sensors with solar harvesting", in *Proc. of 2014 IEEE Symposium on Intelligent Embedded Systems (IES)*, 2014, pp. 1–7. DOI: 10.1109/INTELES.2014.7008978.

[11] J. Konecny, M. Prauzek, M. Borova, K. Janosova, and P. Musilek, "A simulation framework for energy harvesting in wireless sensor networks: Single node architecture perspective", in *Proc. of 2019 23rd International Conference Electronics*, 2019, pp. 1–4. DOI: 10.1109/ELECTRONICS.2019.8765580.

[12] Alberta Agriculture and Rural Development, AgroClimatic Information Service, 2013. [Online]. Available: http://agriculture.alberta.ca/acis/

[13] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. USA: Cambridge University Press, 2014. DOI: 10.1017/CBO9781107298019.

[14] C. J. C. H. Watkins, "Learning from delayed rewards", Ph.D. dissertation, Cambridge University, 1989.

[15] R. S. Sutton and A. G. Barto, *Reinforcement Learning*, 2nd ed. Cambridge, Massachusetts: The MIT Press, 2018.

[16] C. J. C. H. Watkins and P. Dayan, "Technical note: Q-learning", *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992. DOI: 10.1023/A:1022676722315.

[17] "Solar calculation details", 2020. [Online]. Available: https://www.esrl.noaa.gov/gmd/grad/solcalc/calcdetails.html

[18] J. Konecny, M. Prauzek, J. Hlavica, J. Novak, and P. Musilek, "Simulation of a daytime-based Q-learning control strategy for environmental harvesting WSN nodes", in *Proc. of the Fourth International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'19). IITI 2019. Advances in Intelligent Systems and Computing*, vol. 1156. Springer, Cham, 2020. DOI: 10.1007/978-3-030-50097-9_44.