
Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 6, pp. 5313–5321, DOI 10.12694/scpe.v25i6.3275

PERSONALIZED HEALTH MANAGEMENT STRATEGIES BASED ON DEEP REINFORCEMENT LEARNING IN THE NETWORK ENVIRONMENT

LILI WEI* AND JINDA WEI†

Abstract. In order to study the optimal personalized motion push target, the author proposes a personalized health management strategy based on deep reinforcement learning in the network environment. Firstly, the research problem is defined, and a real-time interactive personalized motion target decision-making model is constructed; Subsequently, in response to the uncertain characteristics of user behavior in the problem, a deep reinforcement learning algorithm was adopted, combined with the departure strategy temporal difference learning method and neural network nonlinear fitting method, in order to learn strategies from user historical data; Finally, the effectiveness of the proposed method was validated using a real dataset from Fitbit. The research results indicate that personalized motion goal push services based on deep reinforcement learning can help users cultivate a healthy lifestyle and improve their personal health management level by analyzing user behavior data in real-time, providing scientific guidance and timely incentives.

Key words: Mobile health information service management, Personalized sports goal optimization, Real time interaction model, Deep reinforcement learning

1. Introduction. The process of networking in modern society profoundly shapes the network characteristics of traditional culture, especially bringing tremendous changes to the lifestyle of young people [1]. The openness, dynamism, and virtuality of online cultural forms have expanded traditional receptive learning into experiential online learning, promoting active cognitive expansion; It also attracts students to stay away from classroom teaching, resulting in a weakening of the quality of formal learning and negative impacts on psychological, moral, and legal aspects. Combining the characteristics of the lifestyle of the network society in the era of information technology and knowledge economy, the author from the perspective of "online cultural education", links the prosperity and development of online culture with the healthy growth of young people through "education", and systematically examines and analyzes the healthy growth guidance strategies of young people towards "online cultural education".

The content construction and development of online culture have brought more diverse ways for young people to acquire knowledge, expanding their cognitive perspectives on nature, society, and thinking. The vast group of school students in their growth period, to a certain extent, suffer from lack of discernment and self-control leading to abnormal online behavior. Advocate for positive online cultural education among teenagers, establish a correct view of the internet, form a purified online environment and a good personal psychology, thereby cultivating a healthy online lifestyle. The following provides a method guide for the orderly development of online cultural education and the promotion of healthy growth of young people by exploring the connotation of the correct network view and the mechanism of educational innovation. Network culture originates from the spiritual reference of traditional culture and is also a physical reconstruction based on the network environment. The concept of network culture is the expression of the role of values in the field of network culture, which refers to people's overall understanding and basic views on network culture issues, including value goals, value evaluation, and value orientation. It reflects the attitudes of specific groups and individuals towards network culture, namely the cultural values, thinking patterns, and behavioral tendencies under the existing norms of the network society. Due to differences in ideological quality, moral cultivation, and other factors, teenagers inevitably exhibit different characteristics in their views on online culture [2,3].

*Department of Sports, Shangqiu Polytechnic, Shangqiu, 476000, China
†Legal Affairs Department, ZTE Corporation, Shenzhen, Guangdong, 518000, China (Corresponding author, Wei12345620240126.com)

The correct view of online culture should be to use online culture to enhance knowledge understanding, expand thinking horizons, promote social value convergence in learning progress and healthy growth. Therefore, establishing a correct view of online culture should adhere to traditional cultural values, pay attention to the harmonious atmosphere, moral standards, and legal compliance of online participation. The basic connotation of the healthy and upward network culture view is the overall understanding and external behavior of the network lifestyle in terms of culture, which conforms to the overall orientation of the socialist core value system. Require individuals and groups to reflect their beliefs, ideals, spirits, morals, and psychology in a rational adherence to social core values in their natural state, viewing the internet as an extension of the real society, and adhering to its essential constraints of technological tools, resource platforms, and knowledge education. The concept of online culture, as a cognitive view reflected in the perspective of online culture, requires young people to have "advanced culture, harmonious themes, scientific development, and key control", in order to form correct goals for group participation and individual intervention in online culture, establish scientific standards for measuring the value of online culture, and continuously regulate and guide the positive and innovative orientation of online culture education.

A healthy and upward view of online culture can help teenagers align their learning and growth goals, rely on online cultural education to focus on the main channel of knowledge learning, promote the improvement of thirst for knowledge and the cultivation of values.

Health management is an effective strategy in the context of industrialization, globalization, urbanization and aging, the implementation of hypertension, diabetes and other multiple diseases, is the internal requirements of the implementation of the healthy China strategy and the implementation of a new round of healthy Hubei national action. Health management is not only a technology, but also an integrated governance of the whole population, the whole life cycle and the whole society. Government leaders at all levels should take the lead in understanding the multi-level connotation of health management and guide the whole society to promote the deepening of health management reform. Health administrative departments at all levels should take the initiative to win the attention of the government and the support of various government departments, fully consider health factors in the economic and social development planning, integrate health into all policies, promote the connection between health management planning and macroeconomic policies and social policies, and create a good development environment for health management.

At present, with the development of artificial intelligence technology, mobile devices such as smartphones can collect health-related information and accumulate massive structured and unstructured health data, such as exercise steps, heart rate, etc. How to leverage advanced data analysis and machine learning methods to fully tap into the value of massive health data, guide users to participate in fitness activities more scientifically, and help users improve their personal health status is a hot topic of concern in today's management academia and related enterprises. On August 3, 2021, the State Council issued the "National Fitness Plan (2021-2025)", proposing to build a higher-level public service system for national fitness under the national strategy of national fitness, which can fully leverage the comprehensive value and diverse functions of national fitness in improving people's health levels [4]. The plan points out that improving the level of scientific and health guidance services is the main task that the country needs to complete at present. Mobile health information services refer to service providers providing health guidance services to users based on shared personal health data, promoting users to achieve health goals such as weight loss and body shaping. But users need to go through a period of exercise to experience the effects of health improvement. That is to say, the user's exercise behavior has delayed reward characteristics. Therefore, health information service providers usually provide guidance services for users on short-term goals to generate certain motivational effects, thereby helping users persist in exercising and achieve the ultimate goal of improving their health condition. Currently, most wearable device service providers encourage users to increase their exercise volume and improve their health by setting fixed exercise goals. For example, according to the recommendations of the Centers for Disease Control and Prevention in the United States, Fitbit bracelets set a target of 10000 steps for users to exercise. The 10000 step exercise goal is difficult to match the different health needs of different users. Therefore, it is particularly important for service providers to set targeted short-term goals based on the current state of users and maximize their exercise effectiveness.

On this basis, the author explores health information service providers providing users with sports goal push

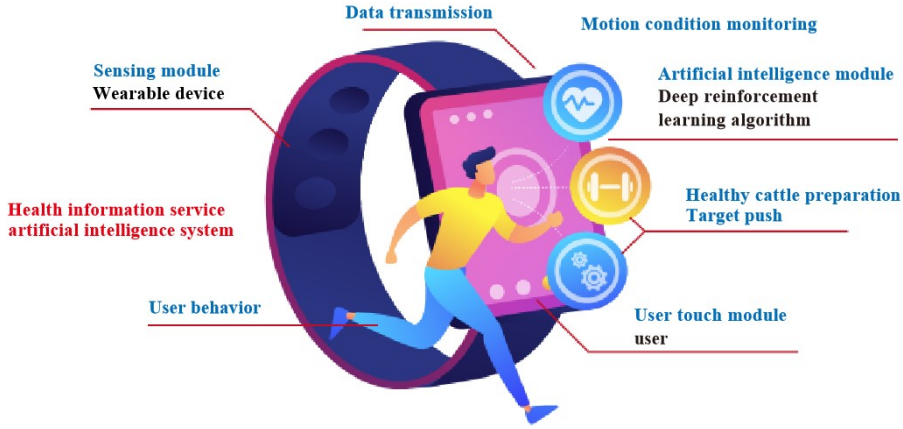

Fig. 1.1: Health Information Service Artificial Intelligence System Framework

services in a way similar to virtual coaches. This type of intelligent information service based on sports big data is a new trend in the development of the health service industry. For example, the sports social app Gudong has laid out intelligent sports and proposed using AI technology to empower and promote the development strategy of "Fitness 3.0". The author combines the concept of "Fitness 3.0" with the basic model of behavioral theory, and applies artificial intelligence technology to the service of pushing sports goals. On the one hand, compared to fixed sports goal setting, it can continuously track the user's exercise progress and make adaptive adjustments to the exercise plan. On the other hand, compared to users setting their own exercise goals, this service can provide users with more scientific exercise plans, helping them improve their health level. The author's core is to design deep reinforcement learning algorithms to achieve the functions of traditional fitness coaches, providing users with interactive and targeted one-on-one scientific guidance services. The artificial intelligence system framework for intelligent health information services is shown in Figure 1.1. The artificial intelligence module relies on cutting-edge deep reinforcement learning algorithms to collect daily exercise and other physiological data tracked by wearable devices, and learn recommendation rules for developing personalized exercise plans for users. It is worth noting that health information services can only recommend suitable exercise goals for users and cannot directly control their exercise behavior. Therefore, artificial intelligence systems need to continuously interact with users, adaptively adjust recommendation strategies based on their historical exercise data, and improve the quality of health services.

2. Methods.

2.1. Problem Description. In real-time interactive mobile health information services, service providers learn service operation strategies from historical interactive data [5]. During each interaction cycle, service providers use wearable devices to monitor user data, including physiological status, exercise status, etc. Then, based on a decision rule and the monitoring results of the user's current state, select an appropriate motion target scheme for the user. After receiving the exercise target, the user responds and executes the exercise activity. Subsequently, the service provider collects feedback information on user behavior based on wearable devices and updates the decision rules for the next cycle. Due to the fact that service providers can only determine the exercise goals to be pushed to users, they are unable to control their behavior in executing exercise activities. Therefore, the problem that health information service providers need to solve is how to set personalized optimization exercise goals for users with the goal of maximizing their long-term utility in an environment of uncertain user behavior, and improve the quality of health information services.

2.2. Model construction.

(1) Personalized Sports Objective Optimization Decision Model. The author studies the problem of health information service providers pushing personalized exercise goals to users at each decision-making stage. The author is based on behavioral economics theory and considers the impact of exercise goal setting on user utility.

At time t, t=0, 1, T. The service provider observes the user's health status ot at time t-1 through wearable devices, including the amount of exercise and calorie consumption generated. Based on this information, the service provider pushes the exercise target gt for stage t to users, which is the amount of exercise that needs to be completed. After receiving the motion target gt, the user performs exercise and generates a new health state ot+1. In this process, the user's utility includes three parts: the health benefits brought by burning calories, the cost of exercise, and the motivational value of exercise goals. The decision-making goal of service providers is to maximize the total revenue of users during the service cycle.

(2) Optimal strategy. The goal of the Markov decision process constructed by the author is to find an optimal strategy π° [6]. From the decision-making process, it can be seen that when the system is at decision time t, the trajectory from time 0 to time t is a deterministic trajectory that has already occurred, denoted as $h_t = (o_0, g_0, o_1, g_1, \cdots, o_{t-1}, g_{t-1}, o_t), t < T$. The random event of user movement from time t to service termination time T did not occur, and the utility of users recommended based on motion targets is also random. Assuming that the motion target gt is selected at time t, a profit sequence of $R_{t+1}, R_{t+2}, \cdots, R_T$ is obtained in the subsequent interaction process. For the discounted return $Y_t = R_{t+1} + \delta R_{t+2} + \cdots + \delta^{T-t-1} R_T$ obtained by mobile health service providers, δ is the discount factor. In addition, the user's state changes follow Markov properties, and the random variables after time t are only related to t. Introduce the value function v_π and the state action function q_t^π to measure the total expected utility of the policy user from time t to time T.

2.3. Deep Reinforcement Learning Methods. Mobile health information service providers need to find the optimal strategy for personalized exercise goals. Due to the influence of many random factors on user behavior during exercise activities, service providers find it difficult to predict the impact of exercise goal decisions on user health benefits. That is to say, service providers are unable to obtain accurate information on the probability of user state transition for processing the state transition process. Furthermore, it is not possible to fully model the problem of personalized motion goal decision-making. The following adopts a data-driven approach to learn strategies from the real interactive environment of health services. The author used a neural network-based algorithm for temporal differential learning of derailment strategies to solve personalized motion target optimization problems. The departure strategy temporal differential learning method, also known as Q-learning algorithm. The Q-learning algorithm combines the sampling method of valuation updates with the Bellman optimal equation of the optimal strategy to solve personalized motion target decision-making problems. However, due to the large state space of users in personalized motion goal decision-making problems, each state action pair (o, g) corresponds to a value function q (o, g) that needs to be learned, so the process of estimating the state action value function is slow. It is very important to use neural network approximation to estimate the value function q (o, g). By approximating the function, a small number of parameters θ can be used to fit the state action value function, $\widetilde{q}(o, g; \theta) \approx q_\pi(o, g)$.

(1) Algorithm design ideas. The author adopted the prioritized replayDQN method, a neural network based on priority experience replay, for temporal differential learning of the derailment strategy. During sampling, the priority of the samples was taken into consideration, which resulted in faster convergence speed of the algorithm. Firstly, construct an experience pool with an accumulated binary tree structure to store sample data, which is the historical data of users used for learning by mobile health information service providers, and normalize the data. Then, using the target strategy $\pi(g|o)$ and the behavior strategy $\mu(g|o)$, the target network DQN1 and behavior network DQN2 neural network models are constructed for the two strategies, respectively. Among them, the target strategy $\pi(g|o)$ is the strategy that the service provider needs to learn, and the state action value function fitted by the target network DQN1 is $Q(o, g; \theta)$. Behavioral strategy $\mu(g|o)$ is the behavior strategy chosen by the service provider, and the state action value function fitted by the behavioral network DQN2 is $Q(o, g; \theta')$. During the learning process, the service provider selects motion targets based on behavioral strategies $\mu(g|o)$, obtains personalized interaction data of user motion targets, and assigns weights to each sample through TD error calculation, which is the probability of each sample appearing. Store sample data and priority indicators in the experience pool.

(2) Neural network structure. There is a non-linear relationship between the estimated state action value generated by the pushed motion target and the user's state. Health information service providers obtain user feedback on the pushed exercise target service by observing the user's status. Therefore, the artificial neural network method is used to fit the state action value function, $\widetilde{q}(o, g; \theta) \approx q_\pi(o, g)$. Neural networks are divided

into input layer, hidden layer, and output layer. The neurons in each layer of the input layer and hidden layer, hidden layer and hidden layer, and hidden layer and output layer are all fully connected, with each line corresponding to a parameter. In the DQN neural network model, a two-dimensional input vector consisting of the user's exercise amount m and calorie consumption f is set in the input layer. After being calculated by the neurons in the input layer, the output features are used as input data for the next layer of neurons, ultimately outputting the state action value estimation for each selectable motion target push action. Assuming there are l hidden layers in the middle layer, where hidden layer i has hi neurons, $i = 1, 2, L.\theta$ represents the parameters of the neural network, b represents the deviation coefficient, and $\sigma(.)$ is called the activation function.

The output of the first hidden layer neuron is:

$$s_j^1 = \sigma(\theta_{1,1}^0 m + \theta_{1,1}^0 f + b_j^0), j = 1, 2, \cdots, h_1 \tag{2.1}$$

The output of neurons in hidden layer i (i=2,.., l-1) is:

$$s_j^i = \sigma(\sum_{k=1}^{h_{i-1}} \theta_{i,k}^{i-1} s_k^{i-1}), j = 1, 2, \cdots, n \tag{2.2}$$

The final output of the neural network is:

$$\widetilde{q}(G_k) = \sum_{j=1}^{h_l} \theta_j^l s_j^l + b_j^l, k = 1, 2, \cdots, n \tag{2.3}$$

Using the modified linear unit function as the activation function, the expression is as follows:

$$\sigma(x) = max\{0, x\} \tag{2.4}$$

The advantage of using a modified linear unit function is that the function is linear, with low computational complexity and fast speed. In addition, when the input is a positive number, the derivative is 1 to avoid the problem of vanishing gradients; On the other hand, modifying the linear unit function to make some neurons output 0 reduces the dependency between parameters, which helps alleviate the problem of overfitting.

(3) Design of Deep Reinforcement Algorithm. The input information of the deep reinforcement learning algorithm is the state vector $\varphi(o)$ corresponding to the user state o, and then the neural network outputs the state action value function $Q(o, g; \theta')$ of all personalized motion targets in that state [7]. Meanwhile, the cumulative binary tree structure is used to store the information obtained from each interaction with the user. Among them, the target network DQN1 obtained through experience replay is used as a label for deep learning to calculate the error of the loss function of the behavioral network DQN2, update the parameter θ' of the behavioral network DQN2 through gradient backpropagation. After θ' converges, an approximate $Q(o, g; \theta') \approx q_\pi(o, g)$. can be obtained. Finally, the optimal strategy π_* for personalized motion goal decision-making can be obtained using the greedy strategy.

3. Results and Analysis.

3.1. Data sources. The author analyzes the real data of Fitbit smart bracelets as an example. The dataset mainly consists of four parts. The Fitbit dataset in the first part is sourced from PMData, consisting of 16 users from November 2020 to March 2021[8]. The experiment utilizes FitbitVersa second-generation smart wristbands to track and record health data automatically. The PMData dataset was publicly presented at the 11th ACM International Multimedia Conference in 2021 for scientific research on mobile health management. The second part of the data comes from the "FitbitConnection" project initiated by the publicly available data sharing platform "Openhumansfoundation", which reads data generated by wearable devices by connecting users to their Fitbit accounts. From the start of the project in 2012, as of December 1, 2022, 37 individuals have publicly disclosed their personal health data. The author will preprocess the personal Fitbit data shared by the participants obtained. Due to the collection of all personal data from 7 years, the time span is too long, and user behavior habits have changed significantly. Therefore, the author divided the Fitbit data of participants

Table 3.1: Basic Information of Fitbit Sample Data

Data set	User ID	Range of motion steps	Calorie consumption range	Time frame	Sample data size
Dataset 1	P01-P16	[6,45342]	[1028,6491]	2020.11- 2021.03	2120
Dataset 2	P17-P132	[4,72417]	[118,9867]	annually(2012-2022)	33827
Dataset 3	P133-P162	[4,36019]	[50,4900]	2017.03-2017.05	1260
Dataset 4	P163-P165	[1683,39000]	[1801,4851]	2022.01- 2022.11	962

into different research subjects by year, resulting in a total of 116 sets of data. The third part of the data was obtained from the Zenodo knowledge database, which collected personal data submitted by 30 Fitbit users between March 12, 2017 and May 12, 2017 through a crowdsourcing task published by Amazon MechanicalTurk. The fourth part of the dataset sources and the experimental team's use of Fitbitarge3 generation smart bracelets to track and record the daily activities of the experimental personnel. A total of three experimental personnel's health data were collected from January 2021 to November 2021. Due to the author's research on personalized sports goal decision-making problems, although the time series data collected from 165 groups of experimental subjects varied in scope, the author optimized the decision rules for pushing motion goals to users by learning their personal historical data. The author validated the applicability of the algorithm through 165 repeated experiments. Firstly, the sample data of 165 experiments were analyzed. The daily exercise steps of the user in the data were selected as the amount of exercise completed by the user through their own efforts, m, and the daily calorie consumption f was obtained by combining Fitbit's built-in intelligent algorithm with user preprocessing such as heart rate, physical activity, and sleep.

The decision variable for personalized exercise goals is the daily exercise goal g pushed to the user, and the number of steps the user needs to complete each day is selected as the user's exercise goal. After preprocessing the raw data and removing noise, such as samples where the user did not equip a Fitbit smart bracelet, resulting in empty calorie consumption, a total of 38165 sample data were obtained. The basic information of the samples is shown in Table 3.1.

3.2. Experimental setup. Due to the fact that personalized motion goal optimization is the optimal decision strategy learned through interaction with the environment when running deep reinforcement learning algorithms. For specific users, first generate a random environment using real historical data generated by Fitbit smart bracelets [9]. Then, learn personalized sports goal decision-making. Assuming that the health information service provider pushes a set of exercise goals for users with selectable exercise steps as follows: {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}, among them, 0 indicates that motion targets are not recommended, while others indicate that motion targets increase proportionally by 10% of the highest historical exercise. Drawing on Jain's research on the impact of optimal goal setting on user psychology and behavior, it is assumed that the health benefits $\varphi(f_t) = \sqrt{f_t}$ brought by calorie consumption and the effort cost $h(m_t) = m_t^2/2$ put in by user exercise are assumed. The experimental parameters are shown in Tables 3.2 and 3.3. The exploration rate $\in$, TD error index α, and importance sampling index β, the number of samples extracted in a single instance m, the number of training iterations M, the number of neurons in the first layer p1, the number of neurons in the second layer p2, and the learning rate r are related to the convergence speed of deep reinforcement learning algorithms. By adjusting the parameters, good convergence results can already be achieved. Marginal loss of failure to achieve motion goals, benefits of achieving motion goals s, and discount factor δ, it is an exogenous parameter provided by mobile health information service providers to users for adaptive motion target decision-making, reflecting their sensitivity to target incentives and the time preference of enterprise decision-making. The author explores the impact of exogenous parameters on decision systems through parameter sensitivity analysis. The experimental environment is an Intel (R) Core (TM) i7-7700CPU3.60GHz (32GRAM) desktop computer running on the CentOS operating system, and the program is developed using PyCharm (version 2019.2) software.

Table 3.2: Relevant Parameters of the Experiment 1

Name	Parameter	Value
Marginal loss of failure to achieve sports goals	l	0
Benefits of achieving sports goals	s	1
discount factor	δ	0.9
Exploration rate	$\in$	0.9
Number of first layer neurons	P1	64
Number of second layer neurons	P2	64
Learning rate	r	0.0005

Table 3.3: Relevant Parameters of the Experiment 2

Name	Parameter	Value
The index of TD error	α	0.6
The index of importance sampling	β	0.4
The number of samples extracted in a single instance	m	32
Training frequency	M	20000

3.3. Analysis. In order to verify the effectiveness of the deep reinforcement learning algorithm proposed by the author in solving personalized motion objective optimization problems, the author presents experimental results [10]. Firstly, fifty users were randomly selected for training, with each user trained ten times. Each experiment can stabilize within 10000 training sessions, and the calculation time is less than two minutes. Figure 3.1 shows the variation of the error value of the loss function generated by training neural network parameters in deep reinforcement learning models. The horizontal axis represents the value of training times, and the vertical axis represents the error loss value of the loss function. From the graph, it can be seen that after training ten thousand times, the curve of the loss value tends to stabilize, and a stable sampling estimation value can be obtained. Next, the author takes the single training situation of P06 users as an example, with a decision cycle of 30 days, and each experiment is repeated 20 times for testing. We compared the user utility of optimization algorithms using personalized motion goal decision-making with Fitbit's built-in 10000 step motion goal, as well as three scenarios without motion goal motivation. Figure 3.2 shows the changes in the benefits obtained by users from participating in health information services every day during a single training process of the model. The horizontal axis represents the date, and the vertical axis represents the user's exercise reward value. After observing 20 sets of experiments, it was found that in most cases, the lower values in daily exercise benefits generated by adaptive motion goal decision-making using deep reinforcement learning algorithms appear less frequently. This means that compared to non personalized fixed motion goal decision-making, deep reinforcement learning algorithms can better adapt to the constantly changing exercise preferences of users and have a better incentive effect on their exercise behavior.

4. Conclusion. The author studied the optimization problem of personalized exercise goals in mobile health information services. Due to the influence of user physiological attributes, health needs, and exercise preferences, it is difficult to predict the motivational effect of exercise goals on users. The uncertainty of user behavior increases the difficulty of making motion target decisions. Therefore, the author proposes a decision-making problem for mobile health information services based on real-time information exchange, which can dynamically and adaptively adjust the decision-making of personalized exercise goals, achieving the goal of motivating users to complete fitness activities and maximizing long-term benefits. Firstly, the user's physiological attributes, exercise status, and other characteristics are taken as environmental factors and monitored in real-time through wearable devices. Train an intelligent agent with the goal of maximizing user health benefits and learn the optimal decision criteria for motion objectives. Subsequently, a deep reinforcement learning

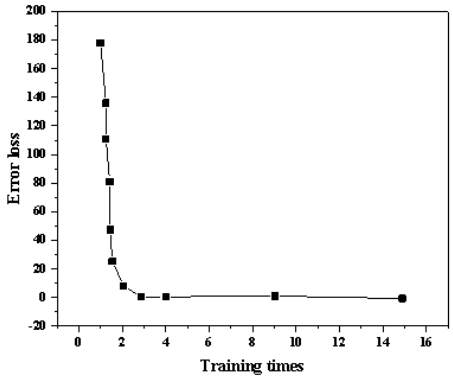

Fig. 3.1: Changes in error loss of personalized motion target decision model

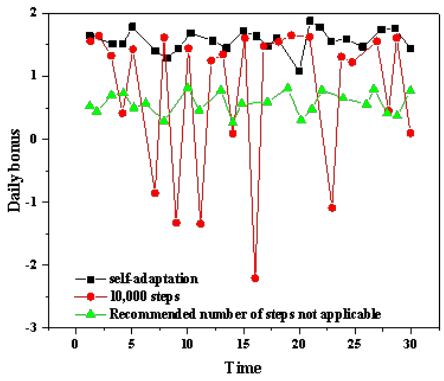

Fig. 3.2: Daily Changes in User Utility

algorithm was designed by combining techniques such as neural networks, stochastic gradient descent, and value function approximation, effectively solving personalized motion target decision-making problems based on real-time interaction. Finally, a practical case was used to verify the practical significance of the research problem and the effectiveness of the algorithm.

REFERENCES

[1] Huang, R. , He, H. , Zhao, X. , Wang, Y. , & Li, M. . (2022). Battery health-aware and naturalistic data-driven energy management for hybrid electric bus based on td3 deep reinforcement learning algorithm. Applied Energy, 321(9089),13.

[2] WeiLI, ChunhuaZHENG, & DezhouXU. (2022). Research on energy management strategy of fuel cell hybrid vehicles based on deep reinforcement learning. Journal of Integration Technology, 10(03), 47-60.

[3] Li, Y. , Qin, X. , Chen, H. , Han, K. , & Zhang, P. . (2022). Energy-aware edge association for cluster-based personalized federated learning.2(545),13

[4] Fang, Y. , Pu, J. , Yuan, C. , Cao, Y. , & Liu, S. . (2022). A control strategy of normal motion and active self-rescue for autonomous underwater vehicle based on deep reinforcement learning. AIP Advances, 12(1), 546-.

[5] Liu, X. , Wang, Y. , & Zhang, K. . (2023). Energy management strategy based on deep reinforcement learning and speed prediction for power-split hybrid electric vehicle with multidimensional continuous control. Energy Technology: Generation,Conversion,Storage,Distribution.67(57),235

[6] Oh, S. , Lee, S. J. , & Park, J. . (2022). Effective data-driven precision medicine by cluster-applied deep reinforcement learning. Knowl. Based Syst., 256(36), 109877.
[7] Zade, A. E. , Haghighi, S. S. , & Soltani, M. . (2022). Deep neural networks for neuro-oncology: towards patient individualized design of chemo-radiation therapy for glioblastoma patients. Journal of biomedical informatics, 127(46), 104006.
[8] Wu, D. , Yang, X. , Shen, Z. , Wang, Y. , & Dong, B. . (2022). Learning to scan: a deep reinforcement learning approach for personalized scanning in ct imaging. Inverse Problems and Imaging, 16(1), 179-195.
[9] Ahmadian, M. , Ahmadian, S. , & Ahmadi, M. . (2023). Rderl: reliable deep ensemble reinforcement learning-based recommender system. Knowledge-Based Systems, 263(797), 110289-.
[10] Zhao, J. , Li, H. , Qu, L. , Zhang, Q. , Sun, Q. , & Huo, H. , et al. (2022). Dcfgan: an adversarial deep reinforcement learning framework with improved negative sampling for session-based recommender systems. Information Sciences, 596(2354), 222-235.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Jan 31, 2024
Accepted: Mar 18, 2024