Research Article

Optimization Algorithm for Multipath Transmission of Distance Education Resources Using Reinforcement Learning

Wang Yajun\(^1\) and Zhao Liang\(^2\)

\(^1\)Chengdu College of Arts and Sciences, Sichuan, Chengdu 610000, China  
\(^2\)School of International Education, Chengdu College of Arts and Sciences, Sichuan, Chengdu 610000, China

Correspondence should be addressed to Zhao Liang; ies@cdcas.edu.cn

Received 14 March 2022; Accepted 19 April 2022; Published 10 May 2022

Academic Editor: Mian Ahmad Jan

Copyright © 2022 Wang Yajun and Zhao Liang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Distance education resources are the essential components of modern distance learning systems, and the development of high-quality resources is a potent assurance of modern distance education quality. The standard transmission efficiency assessment approach ignores the optimal data transmission stop rule and is unable to calculate the ideal stop time, resulting in excessive data transmission energy consumption and an evaluation result that does not match the real value. As a result, this study provides a reinforcement learning-based optimization technique for multipath transmission of distance education materials, because to calculate the optimal transmission rate thresholds in different detection time periods and realize the optimization of resource data transmission efficiency. According to the simulation result, the suggested efficiency optimization approach uses less energy on average and transmits data at a faster pace, improving resource data transmission efficiency. The result also shows that the proposed method performs well in terms of packet loss as compared to other existing methods.

1. Introduction

Education by distance, often known as network education or distance learning, is a type of education in which the tutor and learner both are separated by distance and occasionally by time. Distance education materials are critical in achieving the teaching objectives of distance education and in conveying the teaching contents across long distances [1]. Since December 1998, the Chinese ministry of education has planned and implemented “Modern Distance Education Project” as part of the 21st century education revitalization action plan, with a total of 68 colleges and institutions participating in trial projects. Unfortunately, the development of online education in China has been sluggish so far, and the outlook for the future, as well as the creation of distance education materials, is bleak at best. The resource development of distance education is greatly affected by some issues such as a single construction theme, an arbitrary distribution of resources, an unorganized classification of resources, and lack of evaluation methods. In fact, resource development is the concept and cornerstone of distance education and teaching in a networked setting [2–4]. First and foremost, high-quality distant education resources must be developed for Yuanzheng Education in order to continue to grow and prosper.

According to the framework of the national medium-and long-term education reform and development plan, our country is in an atmosphere where the entire society produces a learning environment for the entire population, which is a positive development. Society is in desperate need of inventive multifaceted skills, and the advancement of mobile technology has the potential to meet the learning requirements of a wide range of adult learners. When it comes to distance and open education in industrialized nations such as China and US, online education has nearly become synonymous with the terms [5–9]. Best Colleges, an American education research consulting agency, questioned online students and university officials. The results revealed that 79 percent of online learners and 76 percent of graduates believe that online education is better than or equal to campus education, and that online learning is booming. In addition, new developments have motivated
the researchers to devote more attention to mobile learning and the development of digital abilities. For eight consecutive years, under the banner of Mobile Learning Week, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) has brought together experts in related fields from around the world to discuss how to accelerate national learning through affordable and powerful mobile technology. Specifically, according to the report released by the 2018 Mobile Learning Week, digital skill is the ability to use digital devices, communication tools, and networks to obtain and manage information, create and share digital content, and communicate and collaborate with others. These abilities enable people to gain significant convenience in the course of their daily lives and social practices, while also gently improving their ability level and achieving self-shaping. The traditional learning methods have constraints in terms of time and space; however, mobile learning overcomes these limitations. The design of mobile learning resources needs to be capable of promoting learners’ ability to adapt, shape, and choose their learning environment, while also meeting the requirements of educational research activities in the modern day. At the moment, Open University’s mobile learning instructional tools are not adequate and do not properly match the learning demands of students. As a result, the development of relatively comprehensive mobile learning materials has emerged as a pressing issue that the Open University must address as soon as possible [10–18].

Although resource and data transmission over a network has a wide range of applications and has been used in different sectors [6], further, it also faces several technological challenges. For instance, insufficient node energy and a lack of continuous supply, especially in the event of excessive resource data, are likely to cause data loss or damage during the transmission process, resulting in data transmission inefficiency that is incongruous with reality in an ideal environment [19–21]. How to effectively increase the effectiveness and availability of total resources and data transmission of a network, as well as how to accurately assess its actual efficiency, has emerged as a critical topic in the field of network resource management that must be resolved immediately at this point [22, 23].

The approach of transmitting packets according to the streaming mechanism, and subsequently transmitting packets for educational content, has been advocated by several academicians, among others [24, 25]. In order to design a service framework that is compliant with the streaming transmission of educational data, the streaming transmission method needs to be employed that handles the data in nodes independently. The data packetization technique is used to safeguard educational data transmission from outside sources of interference and to ensure that the transmission of educational data is stable. It has the potential to significantly increase the efficiency with which instructional data is transmitted. Network-based educational resources have improved in terms of quality, but there are still some issues such as educational data corruption that need to be addressed [26–29]. In addition, researchers have presented a method for evaluating the effectiveness of network resource distribution that is based on quantitative recursive scheduling. The method is ready to use after creating a data flow model of educational network resources, mapping the bit error rate of network education resource data classification into a set of probability density functions, and using the quantitative recursive feature points of educational resource data and use them to generate the evaluation outcomes of the feature points in question [30–32]. Although this evaluation approach has a high degree of accuracy, the computation process is complicated and time-consuming. Others have proposed the problem of minimizing the average energy consumption of data transmission based on the rate at which data is received by the transmission network.

The problem of minimizing the average energy consumption based on the data arrival rate is developed using the features of random variation in wireless channel quality, and then, it is translated into an optimal stopping problem, which demonstrates that the optimal stopping rule exists. Additionally, the energy consumption optimization strategy of data transmission based on the data arrival rate is realized, but this method suffers from a high packet loss rate [33–37].

This work provides an optimization technique for multi-path transmission of distance education resources that is based on reinforcement learning in order to address the issues that have arisen with the previously mentioned methods. This work analyzes the data rate’s transmission energy consumption in accordance with the existing transmission delay requirements. Apart from this, it also establishes the state and reward functions, introduces the minimum expected cost, and then determines the optimal stopping rule according to the framework, and then calculates the optimal time stop time according to the rule.

The remaining structure of the paper is laid down as follows: Section 2 discusses the existing challenges and resource construction of distance education. Section 3 is about the proposed method of the paper. The results and discussion of the paper are elaborated in Section 4 of the paper. Finally, Section 5 concludes the basic theme and ideas of the paper.

2. Distance Education Resource Construction and Challenges

2.1. Challenges. Resource optimization is concerned with ways to more efficiently manage and utilize resources in order to boost overall profitability and to reduce the expenses. The topic of resource optimization may be found in a variety of contexts, ranging from international trade to small-scale companies to everyday issues such as clothes, food, housing, and transportation. Natural resources underpin the entire human economy, as well as its production and daily living activities. The traditional operational research methods, including combinatorial optimization, linear programming, non-convex optimization, and other technologies, have been widely used in resource optimization scenarios for a long time, including shipping optimization, taxi dispatch, supply chain management, and cargo packing, with impressive results. Despite the fact that operation
research methods have proven to be extremely effective in solving the resource optimization problems discussed above, but there are still numerous challenges that need to be overcome for operations research methods to have a significant impact on the solution. These difficulties are primarily caused by the following three factors.

2.1.1. A Tremendous Amount of Available Solution Space. In many practical scenarios, there are many resource nodes, complex dependencies, and a long time for the problem to be solved. As a result, building an operation research model with hundreds or thousands of variables and constraints is very necessary. But it will slow down the system response and will increase the computing cost for finding a solution, making it very tough to apply an operation research model. It is employed in various situations where there is a strong necessity for timeliness, such as taxi dispatch.

2.1.2. A Great Deal of Uncertainty. Concerns about resource optimization are typically based on events that will occur in the future. For example, shipping companies must balance the scheduling of containers based on forecasted supply and demand conditions in order to remain competitive. Taxi dispatching orders must be matched with future orders to be successful. When planning the supply chain, it is important to consider how much capacity each link will have in the future. In addition, warehouse capacity and final consumer demand are taken into consideration while determining the supply strategy. It is necessary to make explicit forecasts about future conditions and then create models based on those explicit predictions to solve this problem by utilizing an operation research-based strategy. In any case, precision of the prediction is always restricted, and when a long-term prediction is necessary, the accuracy is even more difficult to ensure. As a result, this circumstance leads to poor solution quality, low optimization efficiency, and even the impossibility to implement the acquired solution.

2.1.3. The Logic of the Scenario Is Complicated and Variable. Due to the complexity of business logic, many logics cannot be effectively described by the constraints of operation research in practical problems. For instance, some policy and regulatory requirements in cross-border trade, reputation, and loss of customers caused by unmet customer need in the supply chain, etc. Consequently, manual design is required for the establishment of the operation research model in order to approach these limitations, resulting in the model being subject to subjective judgment. Meanwhile, the business logic of the scenario (such as the business model and regulatory compliance needs) will evolve with time. Once it has changed, a large amount of personnel are necessary to readjust the model to accommodate the new modifications, resulting in high labor costs and stable models in the process.

The issues listed above go beyond the scope of typical operation research methods and necessitate the development of whole new approaches. With the continuous progress of information technology and the continuous decline of storage equipment prices, various industries have accumulated a large amount of historical data, including route data, ship departure data, supply and demand relationship data, taxi vehicle trajectory data, and order demand data, as well as package size and destination distribution data in the field of express delivery. It is important to note that this valuable data contain complicated changes in the business as well as the uncertainty of numerous events, which indirectly represent the problem’s operating logic. How to make full use of the data, uncover patterns, and learn strategies from them is a significantly difficult task, but it also represents a tremendous potential to solve the problem of resource optimization by leveraging machine learning.

Given that reinforcement learning (RL) excels in the field of serialized decision-making and achieves superior performance in the field of multiagent collaboration, some of its outstanding characteristics have also attracted the attention of researchers working in the field of resource optimization. For starters, RL-based solution and decisions are extremely efficient in their execution, despite the fact that teaching RL techniques takes a long time and these training assignments can be completed on a computer’s desktop. In practice, only the trained model is required for inference, allowing for approximate real-time decision-making to be done in the vast majority of situations. Second, the way of employing RL does not necessitate making explicit predictions; instead, the model can learn rules and strategies from interaction experience and large amounts of data, thereby assisting in the formulation of appropriate decisions. Finally, with real-time logic, the model is not necessary to represent the business logic, and the business logic can be considered totally a black box, removing both the issues and subjectivity involved with the characterization of intricate business logic. The agent can react promptly when the business environment changes by making use of the change signals included in the data to refine the most suited optimization scheme by engaging with the business environment more rapidly and acutely. As a result of these qualities, the solution of the RL algorithm combined with industry big data has been increasingly utilized in the field of resource optimization in recent years, with a series of great outcomes being obtained.

2.2. Construction Principles. In order to construct the teaching content of a course in order, we complete the implementation of distance education teaching activities through various resources according to the specific structure of the course and develop distance education resources according to the teaching objectives and teaching plans of distance education. The educational and pedagogical features of teaching resources must be clearly defined. As a result, when developing mobile learning resources, the following rules should be followed in terms of structure, design, knowledge content, and resource acquisition.

2.2.1. The Refining of Structure as a Rule of Thumb. Within a course’s learning resources, the knowledge modules that are included in each section of the learning resources are distinct from the knowledge modules that are adjacent to them, and they exhibit the qualities of refinement at the structural level. The independent knowledge modules are linked together in
a sequence to build a learning unit that is administered by a specific organization in order to meet the learning objectives that have been specified.

2.2.2. Design with the Modular Approach in Mind. The knowledge modules, learning units, and learning subjects that make up a course are at the heart of the learning process. In turn, they are developed into reusable and reproducible learning tools, which contain both individual knowledge points and comprehensive knowledge points. In this coconstruction and sharing course model, learners can choose the corresponding learning content that corresponds to their requirements and interests. This coconstruction and sharing course model can shatter the traditional course structure that is excessively rigid in its content.

2.2.3. The Notion of Content Miniaturization Is Used Here as Well. Miniaturized learning resources allow for the reuse of knowledge modules and the creation of new courses for learners by reducing their size. Students who are in a state of temporal fragmentation exhibit traits such as distraction and short duration at the same time. Miniaturized learning tools are appreciated by students for their convenience. We should effectively acquire knowledge points in a piecemeal period of time.

2.2.4. The Resource Acquisition Association’s Basic Concept of Operation. Despite the fact that video-based learning tools are available on their own, the substantive content of each video must be linked to the content of the others. It is vital to build a relationship between separate resources for learners to associate the knowledge points provided in each video in a short amount of time. The learner subconsciously links the independent films together in the continuous learning process as a result of this implicit connection.

3. Methods

In order to maximize the data transmission efficiency of educational resources, it is first essential to organize and merge the transmission efficiency of educational resources across a network of computers. Before any data can be calculated in either unit, the network must be separated into units with the same continuous length and discontinuous nodes. Following the calculation findings, the data cells are clustered, and the data transmission efficiency is integrated in accordance with the data transmission efficiency criteria. The following measures will be kept in consideration.

Based on the assumption that the two data units are represented by the letters $c_1$ and $c_2$, then the distance between the data in the two cells will be computed by applying as follows:

$$\text{dist}(c_1, c_2) = \frac{\{c_1, c_2\} \cap \phi_{c_1} \cap \phi_{c_2}}{\text{Density}_{c_1} \times \text{Density}_{c_2}},$$

where $\phi_{c_1}$ is the simple distribution coefficient of the data cell $c_2$ to $c_1$, near($c$) represents the overall distance value, Density$_{c_1}$ represents the cell radius of $c_1$, and Density$_{c_2}$ represents the cell radius of $c_2$.

Suppose $r(s)_{xij}$ is used to represent the initialization configuration message of the sink node $S$ of the data, $T_{S_{app}}$ represents the current timestamp, $p'_{f/h}$ represents the data transmission delay calculated according to the timestamp, $g_{d_{fg}}'$ represents the remaining network node energy, and $E'_{app}$ represents the node grouping identification, then we have

$$p'_{app} = \frac{E'_{app} \times g_{d_{fg}}'}{p'_{f/h} \times \frac{T_{S_{app}} \times r(s)_{xij}}{u'_{fgj}}}.$$

where $u'_{fgj}$ is the distance factor.

Assuming that $\phi_{f_{fg}}'$ is used to represent the subnode reading of the data as a whole and $k'_{cn}$ represents the data delay constraint of the sink node, then the integration of network data transmission efficiency is as follows:

$$e'_{fgj} = \frac{k'_{cn} \times \phi_{f_{fg}}'}{p'_{app} \times W_{rate}} \times R_{app} \pm \text{dist}(c_1, c_2).$$

The wireless channel of the wireless sensor terminal quality is randomly transformed according to the network operation, and $\tau$ is described as the maintenance time of the channel. Then, the detection channel needs to be separated by a period to keep the detection time less than $\tau$, and the detection energy consumption is $E_D$. The transmitting terminal transmits data after detecting channels, the transmission time is $t_n$ and the amount of data to be transmitted is $C_n$. The amount of transmitted data will not exceed the channel holding time. Then, $t_n = \min \{C_n/t_n, \tau\}$, the amount of transmitted data is $Q_n = \min \{C_n, t_n \tau\}$. The transmission power sent to the terminal is $p$, and the transmission energy consumption is $t_n \cdot p$.

The data value is greater than $E_D$. When the channel completes the $n$-th detection, the energy consumption of a round of data transmission can be represented by $Neotp +$, and according to the energy consumption, the average energy consumption $Z_n = E_u/Q_u$ of the transmitted unit data can be obtained. Among them, the minimum average energy consumption is $Z_n$. Then, $N$ can be represented as the best stop time, and $N \geq 1$ at the same time. Then, the maximum transmission delay rate of the target data can be represented by $D_{max}$. Define $M = [D_{max}/\tau]$ and we know that $1 \leq N \leq M$.

Define the average energy consumption efficiency of transmitting unit data as $\xi$, and the calculation of $\xi$ is

$$\xi = \frac{\sum_{j=1}^{M} E_{N_j} \sum_{j=1}^{M} Q_{N_j}}{\sum_{j=1}^{M} F_{N_j}},$$

where $\phi_{c_2}$ is the simple distribution coefficient of the data cell $c_2$ to $c_1$, near($c$) represents the overall distance value, Density$_{c_1}$ represents the cell radius of $c_1$, and Density$_{c_2}$ represents the cell radius of $c_2$.

The energy consumption minimization formula is

$$\xi^* = \inf_{N \in \mathbb{N}} \frac{E[Q_N]}{E[Q_N]} = \inf_{N \in \mathbb{N}} \frac{E[NED + P\epsilon N]}{E[Q_N]}.$$
Rewards are given as follows:

\[ Y_e(\xi) = e'_n nE_D + P_{in} - \xi Q_n. \]  

(6)

Then

\[ V(\xi) = \inf_{N \in N} \{ E[nE_D + P_{in} - \xi Q_n] \}. \]  

(7)

The optimal stop time for the educational computer is

\[ N^* (\xi) = \arg \inf_{N \in N} \{ E[nE_D + P_{in} - \xi Q_n] \}. \]  

(8)

The RL transition probability and cumulative discounted rewards are given as follows:

\[ P[S = s_{t+1} | S = s_t, A = a_t], \]
\[ R_t = \sum_{s = 0}^{T} y^{t-1} r(s_t, a_t). \]  

(9)

RL policy update function is given as follows:

\[ \nabla \phi J(\phi) = E_{r, pA} \left[ \nabla_{\theta} Q(s, a ; \theta) \big|_{a = \pi(s ; \phi)} \right] \nabla_{\phi} \pi(s ; \phi). \]  

(10)

This paper models distance education resource data transfer as a Markov decision process and the difference between the stop time and the optimal stop time as a reward in RL.

\[ R_t = \sum_{s = 0}^{T} y^{t-1} \inf_{N \in N} \{ nE_D + P_{in} - \xi Q_n \}. \]  

(11)

The loss function is

\[ L(\theta) = E_{s,a,i'} \left[ (y - Q(s, a ; \theta))^2 \right], \]
\[ y = r + \gamma Q(s', a' ; \theta'). \]  

(12)

To improve network transmission efficiency, we employ the principle of optimal stopping problems to detect and compare various optimization methods separately, as well as to obtain certification of network threats and user behaviors in order to construct their detection matrix and evaluate transmission resource optimization.

For example, we use the equation \( p = \{ p_A, p_D \} \) to describe the initial efficiency of network transmission, where \( p_A \) represents the running speed of data during transmission and \( p_D \) represents the collection of operating strategies available. A number of various consequences of efficiency optimization will be identified using the following formula:

\[ \phi(k) = \frac{u'(h^k) \times \{ p_A, p_D \}}{p} f(k)e(k), \]  

(13)

where \( \phi(k) \) denotes the transfer speed.

In short, this paper develops a multipath transmission optimization algorithm for distance education resources based on RL. It models the transmission efficiency and optimal stopping time in traditional transmission optimization as the MDP process and then uses RL to solve the problem of multipath transmission optimization.

### 4. Result and Discussion

Simulated experiments are carried out to test the data transmission efficiency of distance education resources through a specific data detector in order to verify the efficacy of the data transmission efficiency optimization of distance education resources presented in this paper and to evaluate the accuracy of the mathematical model developed.

By comparing the resource data transmission efficiency optimization with the other data transmission strategies based on simulation results and assessing the proposed optimization scheme, it is concluded that it can increase the network resource data transmission efficiency and reliability. The average transmission efficiency is calculated by comparing the total data volume received by the receiving terminal to the total data volume transmitted by the sending terminal. As the value of the variable changes, the data volume of the data transmission will grow in proportion, resulting in a delay in the transmission of the data. In addition, the amount of invalid data will be minimized. Figure 1 shows...
the result of the Data Transmission Service (DTS) algorithm in terms of transmission efficiency, and the transmission efficiency of our proposed model is shown in Figure 2. The data transmission efficiency of the DTS algorithm is simulated and compared to the results of the proposed algorithm. It can be concluded from the following figures that our proposed model performs better than the DTS algorithm in terms of data transmission efficiency.

The model proposed in this study has the maximum transmission efficiency when compared to DTS, and it can obtain the threshold of the highest transmission data in a variety of periods, as shown in Figure 2. The anticipated
resource data packet loss rate is compared with the actual resource data packet loss rate in order to demonstrate that the resource transmission efficiency has been improved. When it comes to the Internet mistakes, if the rate of data packet loss is low, it can be concluded that the network’s stability is good and the transmission efficiency is high. Figure 3 depicts a comparison of the packet loss rate between different networks.

In Figure 3, the red line shows the chances of data packet loss in our proposed method while the other color represents the packet loss of the other existing methods, including DTS, DTS+, and DTPS. It is quite clear from Figure 3 that the chances of packet loss in our proposed method are minimal as compared to the other existing methods.

A comparison of the failure rate of the multipath educational resource transmission over the network is shown in Figure 4. Here, the green color bars show the failure rate of the traditional methods whereas the red bars show the transmission failure of the proposed method. From Figure 4, it is quite clear that the failure rate of our proposed method is less than the traditional method.

5. Conclusion

Distance education resources are essential for fulfilling the distance education teaching goals and delivering educational information across long distances. The development of high-quality resources is a potent assurance of the overall quality of current distance education programs. At any moment of time, the quality of the network channel might change. As a consequence, the higher the channel quality, the larger the transmission efficiency, and the higher the quality efficiency of distance education resource data transmissions are accomplished when the energy consumption of network node transmission is managed. The amount of data that is exchanged will grow in the future. As a result, when the channel quality is high, the amount of data that can be communicated as well as the efficacy of remote education data may both be raised to a degree. This paper proposes a multipath transmission optimization algorithm for distance education resources based on reinforcement learning. It also implements the optimization method of resource data reaching a certain rate, which is based on the optimal stopping rule of data transmission. In the experiments, it is demonstrated that the approach is capable of producing the optimization effect to a certain extent, and the evaluation findings are nearly identical to those obtained after efficiency optimization, demonstrating that the model’s evaluation results are correct.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. N. Gunawardena and M. S. McIsaac, “Distance Education,” in Handbook of Research on Educational Communications and Technology, pp. 361–401, Routledge, 2013.

[2] S. Wang, S. Bi, and Y. J. A. Zhang, “Deep reinforcement learning with communication transformer for adaptive live streaming in wireless edge networks,” IEEE Journal on Selected Areas in Communications, vol. 40, no. 1, pp. 308–322, 2022.

[3] T. Ha, J. Oh, D. Lee, J. Lee, Y. Jeon, and S. Cho, “Reinforcement learning-based resource allocation for streaming in a multi-modal deep space network,” in 2021 International Conference on Information and Communication Technology Convergence (ICTC), pp. 201–206, Ho Chi Minh City, 2021.

[4] J. Luo, F. R. Yu, Q. Chen, and L. Tang, “Adaptive video streaming with edge caching and video transcoding over software-defined mobile networks: a deep reinforcement learning approach,” IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 1577–1592, 2020.

[5] R. C. de Castro Junior, T. C. Medeiros, H. M. Honório, E. Sant’Ana, and P. S. da Silva Santos, “Moodle: teaching strategies in distance education in oral medicine,” Education Research International, vol. 2017, Article ID 4279141, 4 pages, 2017.

[6] S. Elifdroussi, M. Lachgar, H. Kabaili, A. Rochdi, D. Goujdami, and L. El Firdoussi, “Assessing distance learning in higher education during the COVID-19 pandemic,” Education Research International, vol. 2020, Article ID 8890633, 13 pages, 2020.

[7] L. Rajasingh, “Will mobile learning bring a paradigm shift in higher education?,” Education Research International, vol. 2011, Article ID 528495, 10 pages, 2011.

[8] X. Zhang, “Intelligent recommendation algorithm of multimedia English distance education resources based on user model,” Journal of Mathematics, vol. 2022, Article ID 2012700, 8 pages, 2022.

[9] X. Ding, J. Niu, and Y. Han, “Research on distance education development in China,” British Journal of Educational Technology, vol. 41, no. 4, pp. 582–592, 2010.

[10] Y. Su, M. Liwang, Z. Gao, L. Huang, X. Du, and M. Guizani, “Optimal cooperative relaying and power control for IoT networks with reinforcement learning,” IEEE Internet of Things Journal, vol. 8, no. 2, pp. 791–801, 2021.

[11] T. Mahmood, J. Li, Y. Pei et al., “An intelligent fault detection approach based on reinforcement learning system in wireless sensor network,” The Journal of Supercomputing, vol. 78, no. 3, pp. 3646–3675, 2022.

[12] J. Xu and B. Ai, “Experience-driven power allocation using multi-agent deep reinforcement learning for millimeter-wave high-speed railway systems,” IEEE Transactions on Intelligent Transportation Systems, vol. 81, pp. 1–11, 2021.

[13] J. Wang, Y. Gao, C. Zhou, R. Simon Sherratt, and L. Wang, “Optimal coverage multi-path scheduling scheme with multiple mobile sinks for WSNs,” Computers, Materials & Continua, vol. 62, no. 2, pp. 695–711, 2020.

[14] Y. Liu, Y. Deng, H. Zhou, M. Elkashlan, and A. Nallanathan, “A general deep reinforcement learning framework for grant-free NOMA optimization in mURLLC,” 2021, https://arxiv.org/abs/2101.00515.

[15] X. Meng, L. Wu, and S. Yu, “Research on resource allocation method of space information networks based on deep reinforcement learning,” Remote Sensing, vol. 11, no. 4, p. 448, 2019.
[16] L. Xiao, Q. Li, T. Chen, E. Cheng, and H. Dai, “Jamming games in underwater sensor networks with reinforcement learning,” in 2015 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, Denver, 2015.

[17] D. Zhang, H. Fan, and J. Zhang, “Novel resource allocation algorithm of edge computing based on deep reinforcement learning mechanism,” in 2021 IEEE Int’l Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPAN/BDCloud/SocialCom/SustainCom), pp. 437–444, Tianjin, 2021.

[18] R. Xie, X. Jia, and K. Wu, “Adaptive online decision method for initial congestion window in 5G mobile edge computing using deep reinforcement learning,” IEEE Journal on Selected Areas in Communications, vol. 38, no. 2, pp. 389–403, 2020.

[19] T. Dong, Z. Zhuang, Q. Qi et al., “Intelligent joint network slicing and routing via GCN-powered multi-task deep reinforcement learning,” IEEE Transactions on Cognitive Communications and Networking, vol. 26, 2021.

[20] Z. Zhang, H. Chen, M. Hua, C. Li, Y. Huang, and L. Yang, “Double coded caching in ultra dense networks: caching and multicast scheduling via deep reinforcement learning,” IEEE Transactions on Communications, vol. 68, no. 2, pp. 1071–1086, 2020.

[21] M. Naeem, G. De Pietro, and A. Coronato, “Application of reinforcement learning and deep learning in multiple-input and multiple-output (MIMO) systems,” Sensors, vol. 22, no. 1, p. 309, 2022.

[22] F. Tang, B. Mao, Y. Kawamoto, and N. Kato, “Survey on machine learning for intelligent end-to-end communication toward 6G: from network access, routing to traffic control and streaming adaption,” IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1578–1598, 2021.

[23] X. Liu, W. Yang, and Y. Chen, “Resource allocation method of coastal wireless communication network based on machine learning,” Journal of Coastal Research, vol. 97, no. sp1, pp. 223–228, 2019.

[24] X. Liu, C. Sun, W. Yu, and M. Zhou, “Reinforcement-learning-based dynamic spectrum access for software-defined cognitive internet of things,” IEEE Transactions on Industrial Informatics, vol. 18, no. 6, pp. 4244–4253, 2022.

[25] F. Restuccia and T. Melodia, “DeepWiERL: Bringing deep reinforcement learning to the Internet of Self-Adaptive Things,” in IEEE INFOCOM 2020 - IEEE Conference on Computer Communications, pp. 844–853, Abu Dhabi, 2020.

[26] Q. N. Tran, N. S. Vo, M. P. Bui, T. M. Phan, Q. A. Nguyen, and T. Q. Duong, “Spectrum sharing and power allocation optimised multi-hop multi-path D2D video delivery in beyond 5G networks,” IEEE Transactions on Cognitive Communications and Networking, vol. 43, 2021.

[27] S. A. Butt, K. A. Bakar, N. Javaid et al., “Exploiting layered multi-path routing protocols to avoid hole regions for reliable data delivery and efficient energy management for IoT-enabled underwater WSNs,” Sensors, vol. 19, no. 3, p. 510, 2019.

[28] H. Zhao, M. Zhang, H. Yu, T. Mao, and H. Zhu, “Multi-path TCP path scheduling optimization based on q-learning in vehicular heterogeneous networks,” in 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–5, Perth, 2018.

[29] A. Qadeer, M. J. Lee, and K. Tsukamoto, “Flow-level dynamic bandwidth allocation in SDN-enabled edge cloud using heuristic reinforcement learning,” in 2021 8th International Conference on Future Internet of Things and Cloud (FiCloud), pp. 1–10, Casablanca, 2021.