European Conference on Optical Communication (ECOC) 2022 © Optica Publishing Group 2022

Reinforcement-Learning-based Network Design and Control with Stepwise Reward Variation and Link-Adjacency Embedding

Kenji Cruzado(1), Ryuta Shiraki(1), Yojiro Mori(1), Takafumi Tanaka(2), Katsuaki Higashimori(2), Fumikazu Inuzuka(2), Takuya Ohara(2), Hiroshi Hasegawa(1)

(1) Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi, 464-8603 Japan, cruzado.kenji.p5@s.mail.nagoya-u.ac.jp
(2) NTT Corporation, 1-1 Hikari-no-oka, Yokosuka, Kanagawa, 239-0847 Japan

Abstract *We propose a reinforcement-learning-based network design and control algorithm that introduces reward variation dependent on maximum link utilization and link-adjacency embedding as input parameters. Up to 65%/20% capacity enhancement relative to first-fit and link-congestion-aware methods is verified. ©2022 The Authors*

Introduction

Motivated by the continual steep growth in IP traffic volume around the world, a lot of effort is being devoted to enhancing network capacity with the introduction of higher link parallelism by spatial-division-multiplexing (SDM) [1-3] and the use of unused frequency bands in fibres [4,5]. Another trend is to pursue better network utilization by introducing flexible bandwidth assignment to paths by the ITU-T flexgrid [6-8] and dynamic path operations enabled by software-defined-networking (SDN) [9,10]. In order to combine these studies, we have to find a way to dynamically operate complex networks that have high spatial parallelism, fine granular frequency assignment, non-uniform transmission loss over multiple frequency bands, and impairments such as inter-core crosstalk at multi-core fibres. Emerging machine-learning (ML) techniques are expected to play a key role in developing efficient operation schemes of such future networks.

The essential task of optical network operation is routing (including fibre and core selection), and frequency / spectrum assignment (RWA/RSA). The application of reinforcement learning (RL), a variant of ML, to RWA/RSA has recently commenced [11-15]. Several challenges have been elucidated to account for network-specific constraints such as connection latency [16], multiband allocation [17,18], and robustness guarantees [19]. We have substantially improved learning efficiency by splitting the original RWA/RSA into two compact sub-problems; RL-based routing for all wavelengths / frequencies and wavelength / frequency assignment considering route optimality. However, the network operation task is still complex necessitating the adequate numbers of agents and selection of the best one to sufficiently outperform conventional non-ML algorithms.

In this paper, we propose a novel RL-based network design and control method inspired by the use of link congestion information in an efficient heuristic RWA/RSA algorithm. We start with our previous RL-based routing with fibre and core selection for each wavelength / frequency due to its efficiency given by the use of a single common neural network (NN) for all wavelengths / frequencies. Link-adjacency, which connects pairs of links by nodes, is newly embedded into the input vectors of the NN. Reward for each path setup is varied according the maximum utilization ratio of all links in the network. Our stepwise reward control corresponds to a link weighting technique, a heuristics based on congestion levels. Our advances enable stable and efficient learning with a small number of agents. Numerical simulations confirm that a given network can accommodate up to 65% and 20% more paths than the simple first-fit and the congestion-aware heuristics, respectively.

RL-based Network Design and Control Method that Considers State-Value Saturation and Link-Adjacency

This paper considers transparent optical path networks whose optical paths are located on a uniformly spaced frequency grid. Although this assumption corresponds to ITU-T fixed grid networks, the discussion in this paper can be generalized to ITU-T flexible grid networks with aligned path accommodation by defining several regular grids with different spacing; each path is located on a regular frequency grid whose grid spacing equals the frequency bandwidth of the path. Aligned path accommodation, named semi-flexible grid, has been proven to matching the routing performance of the conventional flexible grid [20]. Two configurations have been evaluated. The first one is referred as episodic network design; the traffic demand is just a random sequence of path setup requests and the objective is to maximize the number of

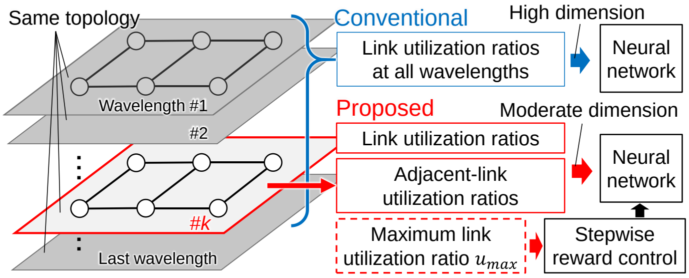

Fig. 1: Conventional and proposed composition of input of NN.

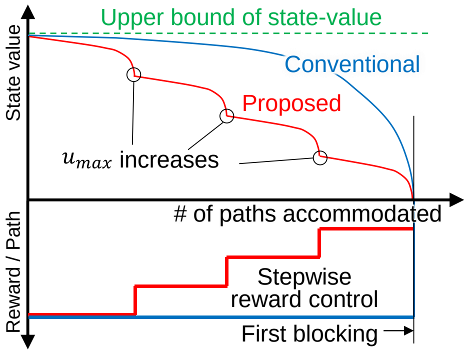

Fig. 2: Comparison of conventional and proposed reward and state-value.

successfully accommodated paths until the first path blocking. The other is dynamic network control; the traffic demand is represented by the continuous arrival of setup/teardown requests and the objective is to maximize the expected number of accommodated paths subject to a given path blocking ratio.

The status of a network is represented by a vector whose components are the utilization ratios of all frequency indexes on all links. As the high-dimensionality of the vector degrades the efficiency of ML-based design and control, we proposed an efficient RL-based design and control algorithm that adopts a more compact state expression (See Fig. 1). A vector of the link utilization ratio of a frequency index is fed to a neural network approximating state-values for the frequency and the neural network is commonly used for all frequency indexes to accelerate the learning process [21]. This formulation has validated for semi-flexible grid networks [22]. However, these methods suffer from the state-value saturation issues explained below and need to be improved. In this paper, we introduce the following novel techniques to RL-based network design and control.

Stepwise reward variation subject to maximum link utilization ratio

Suppose that we have a system with state s and the state will be changed to s' by action $a(s,s')$. An episode is a finite sequence of states and the goal is to maximize the total reward up to failure by selecting appropriate actions for all states. A TD(0)-based algorithm updates state-value V at state s as

$$V(s) \leftarrow V(s) + \alpha[r(s,s') + \gamma V(s') - V(s)] \qquad (1)$$

where $r(s,s')$ is the reward associated with the action, discount rate $\gamma \in [0,1]$, and a sufficiently small $\alpha \in [0,1]$ [23]. Value $V(s)$ will gradually converge to $\gamma V(s') + r(s,s')$ after sufficiently many updates for all possible states.

In order to apply similar algorithms to optical network design and control, the objective will be the maximization of the number of paths successfully accommodated by the given network. Thus we could use $r(s,s') = 1$ for all s, s' so that the total reward equals the number of accommodated paths. Then, in the inference stage, the TD(0)-based algorithm selects next actions so as to maximize the state-values of next states $V(s')$. The discount rate $0 < \gamma < 1$ contributes to stabilize training and enhances learning efficiency. On the other hand, the state-values will saturate with the bound of $1/(1-\gamma)$ (See Fig. 2). As there exist numerous network states and the number of episodes is relatively limited, the estimation error of state-values will not be negligible. The non-negligible error and saturation in state-value estimation trigger the selection of non-optimal actions except at the end of the episode.

Considering the essential limitation in RL, we propose a novel stepwise reward variation associated with maximum link utilization. Let the utilization ratio of a link be the ratio of the number of paths traversing the link to the maximum number of paths to be accommodated. The reward is written as $1 - u_{max}$ where u_{max} is the largest utilization ratio of all links in the network. This makes the state-value a pseudo linear function of the expected number of paths to be accommodated to the network.

Link-adjacency embedding to input vectors

In our previous work, a vector of utilization ratio of each frequency on all links is used to calculate the state-value and the set of vectors for all frequencies represents the exact status of the network (See Fig. 2). All the components in each vector are just arrayed in parallel; however, which pairs of links are bridged by nodes is strongly correlated with successful path accommodation and, as a result, with state-values. We call such pairs of links adjacent link pairs. In this paper, we propose to use a vector where each component corresponds to an adjacent link pair and the value is the maximum of utilization ratios of the links in the pair (See Fig. 3).

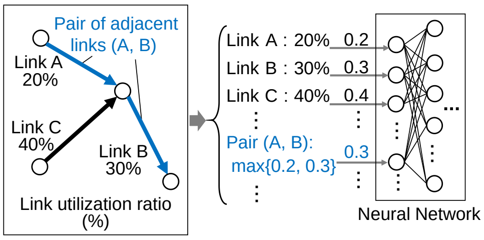

Fig. 3: Calculation of link-adjacency and composition of input to NN.

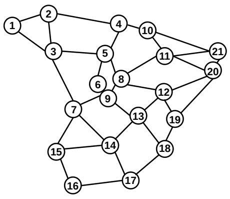

Fig. 4: Spanish Telefónica Network.

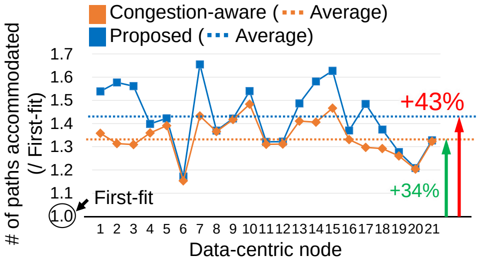

Fig. 5: Number of accommodated paths to a network for different data-centric node locations.

Due to space limitations, we briefly summarize our network design and control algorithm below. The algorithm calculates the state-values subject to given information on traffic distribution matrix T whose component (i, j) is the expected number of paths from node i to node j, and then finds an appropriate route for each path setup request.

Proposed Network Design and Control Algorithm

Step 1. Initial setup stage

Find all adjacent link pairs for the given network topology and fix the dimension of input vectors. Randomly generate episodes, sequences of path setup requests that follow given traffic distribution matrix T. Conduct episodic network design for these episodes and train the common neural network for all frequency indexes by processing path setup requests and updating state-values with Eq.(1).

Step 2. Design/control stage

If a path teardown request arrives, remove the path immediately. If a path setup request arrives, in ascending order of frequency index, search for available routes for the selected frequency. If found, select a route that maximizes the state-value and set up the path. Otherwise, increase the wavelength index and search for available routes again. If no route is available, terminate (for episodic design) or block the request (for dynamic control).

Numerical Simulations

We verified the performance of the proposed algorithm on Spanish Telefónica network with 21 nodes and 35 links (Fig. 4). In order to highlight the difference in routing between the proposal and conventional scheme, we assume that the number of wavelengths is one; i.e. no wavelength assignment. The maximum number of paths on each link is set to 20. A data-centralized traffic distribution is assumed; traffic volume from/to a selected node is four times that between the other nodes. The selected node is called the data-centric node. Path setup requests are generated according to the distribution and episodic network designs are conducted; the metric for the comparison is the number of paths accommodated up to blocking.

The dimension of the input vector representing the link utilization and adjacency is 244. We adopt a three-layer NN to approximate the state-values in the broad parameter space. The numbers of neurons of input, hidden, and output layers are, respectively, 244, 32, and 1. The NN is trained with 400 episodes, where each episode starts from the empty state and ends with the first blocking. The number of agents / NNs is set to 16, much lower than is typical for typical ML-based methods. The step-size of link utilization ratio is set to 10%.

A congestion-aware resource allocation method is adopted as a conventional alternative. It adaptively controls the weight of links according to their utilization ratios, and finds the shortest route in terms of link weight. The results are normalized with the result of the basic first-fit route finding method under the same conditions. Figure 5 shows the average number of paths accommodated in 50 trials for different data-centric node portions. The proposed method improves the number of paths accommodated by 6.5% (average) / 20% (maximum) and 43% / 65% from the conventional congestion-aware and first-fit methods irrespective of data-centric node locations. The introduction of a secondary metric such as average latency could differentiate the proposal from the efficient conventional alternative.

Conclusion

In this paper, we have proposed a novel RL-based network design and control method that adopts link-adjacency embedding and stepwise reward variation inspired by an efficient congestion-aware heuristic algorithm. Numerical simulations verified that the proposed method successfully outperformed simple first-fit and congestion-aware heuristic algorithms by up to 65% and 20%, respectively.

References

[1] G. M. Saridis, D. Alexandropoulos, G. Zervas and D. Simeonidou, “Survey and evaluation of space division multiplexing: from technologies to optical networks,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2136-2156, 2015. DOI: 10.1109/COMST.2015.2466458

[2] M. Jinno, “Spatial channel network (SCN): opportunities and challenges of introducing spatial bypass toward the massive SDM era [invited],” *Journal of Optical Communications and Networking*, vol. 11, no. 3, pp. 1-14, 2019. DOI: 10.1364/JOCN.11.000001

[3] J. Comellas, J. Perelló, J. Solé-Pareta and G. Junyent, “Using spatial division multiplexing to avoid fragmentation in flex-grid optical networks,” presented at *2020 22nd International Conference on Transparent Optical Networks (ICTON)*, Bari, Italy, 2020. DOI: 10.1109/ICTON51198.2020.9203291

[4] M. Nakagawa, H. Kawahara, K. Masumoto, T. Matsuda and K. Matsumura, “Performance evaluation of multi-band optical networks employing distance-adaptive resource allocation,” presented at *2020 Opto-Electronics and Communications Conference (OECC)*, Taipei, Taiwan, 2020. DOI: 10.1109/OECC48412.2020.9273660

[5] B. Correia, R. Sadeghi, E. Virgillito, A. Napoli, N. Costa, J. Pedro and V. Curri, “Power control strategies and network performance assessment for C+L+S multiband optical transport,” *Journal of Optical Communications and Networking*, vol. 13, no. 7, pp. 147-157, 2021. DOI: 10.1364/JOCN.419293

[6] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone and S. Matsuoka, “Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies,” *IEEE Communications Magazine*, vol. 47, no. 11, pp. 66-73, 2009. DOI: 10.1109/MCOM.2009.5307468

[7] I. Tomkos, S. Azodolmolky, J. Solé-Pareta, D. Careglio and E. Palkopoulou, “A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges,” *Proceedings of the IEEE*, vol. 102, no. 9, pp. 1317-1337, 2014. DOI: 10.1109/JPROC.2014.2324652

[8] ITU-T, “G694.1,” https://www.itu.int/rec/T-REC-G.694.1, accessed on 10 May 2022.

[9] M. Channegowda, R. Nejabati and D. Simeonidou, “Software-defined optical networks technology and infrastructure: enabling software-defined optical network operations [invited],” *Journal of Optical Communications and Networking*, vol. 5, no. 10, pp. A274-A282, 2013. DOI: 10.1364/JOCN.5.00A274

[10] D. Kreutz, F. M. V. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky and S. Uhlig, “Software-defined networking: a comprehensive survey,” *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14-76, 2015. DOI: 10.1109/JPROC.2014.2371999

[11] X. Chen, B. Li, R. Proietti, H. Lu, Z. Zhu and S. J. B. Yoo, “DeepRMSA: A deep reinforcement learning framework for routing, modulation and spectrum assignment in elastic optical networks,” *Journal of Lightwave Technology*, vol. 37, no. 16, pp. 4155-4163, 2019. DOI: 10.1109/JLT.2019.2923615

[12] P. D. Choudhury and T. De, “Recent developments in elastic optical networks using machine learning,” presented at *2019 21st International Conference on Transparent Optical Networks (ICTON)*, Angers, France, 2019. DOI: 10.1109/ICTON.2019.884046

[13] J. Suarez-Varela, A. Mestres, J. Yu, L. Kuang, H. Feng, A. Cabellos-Aparicio and P. Barlet-Ros, “Routing in optical transport networks with deep reinforcement learning,” *Journal of Optical Communications and Networking*, vol. 11, no. 11, pp. 547-558, 2019. DOI: 10.1364/JOCN.11.000547

[14] X. Chen, R. Proietti, C. -Y. Liu and S. J. Ben Yoo, “Towards self-driving optical networking with reinforcement learning and knowledge transferring,” presented at *2020 International Conference on Optical Network Design and Modeling (ONDM)*, Barcelona, Spain, 2020. DOI: 10.23919/ONDM48393.2020.9133022

[15] M. Shimoda and T. Tanaka, “Mask RSA: end-to-end reinforcement learning-based routing and spectrum assignment in elastic optical networks,” presented at *47th European Conference on Optical Communication (ECOC 2021)*, Bordeaux, France, 2021. DOI: 10.1109/ECOC52684.2021.9606169

[16] C. Hernandez-Chulde, R. Casellas, R. Martínez, R. Vilalta, and R. Munoz, “Assessment of a latency-aware routing and spectrum assignment mechanism based on deep reinforcement learning” presented at *47th European Conference on Optical Communication (ECOC 2021)*, Bordeaux, France, 2021. DOI: 10.1109/ECOC52684.2021.9605919

[17] P. Morales, P. Franco, A. Lozada, N. Jara, F. Calderón, J. Pinto-Ríos and A. Leiva, “Multi-band environments for optical reinforcement learning gym for resource allocation in elastic optical networks,” presented at *2021 International Conference on Optical Network Design and Modeling (ONDM)*, Gothenburg, Sweden, 2021. DOI: 10.23919/ONDM51796.2021.9492435

[18] N. E. D. E. Sheikh, E. Paz, J. Pinto and A. Beghelli, “Multi-band provisioning in dynamic elastic optical networks: a comparative study of a heuristic and a deep reinforcement learning approach,” presented at *2021 International Conference on Optical Network Design and Modeling (ONDM)*, Gothenburg, Sweden, 2021. DOI: 10.23919/ONDM51796.2021.9492334

[19] H. Ma, J. Zhang and Y. Ji, “Graph sequence attention network-enabled reinforcement learning for time-aware robust routing in OSU-based OTN,” presented at *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, USA, 2022. DOI: 10.1364/OFC.2022.Th2A.18

[20] Z. Shen, H. Hasegawa, K. Sato, T. Tanaka and A. Hirano, “A novel semi-flexible grid optical path network that utilizes aligned frequency slot arrangement,” presented at *39th European Conference and Exhibition on Optical Communication (ECOC 2013)*, London, UK, 2013. DOI: 10.1049/cp.2013.1426

[21] R. Shiraki, Y. Mori, H. Hasegawa and K. Sato, “Dynamic control of transparent optical networks with adaptive state-value assessment enabled by reinforcement learning,” presented at *2019 21st International Conference on Transparent Optical Networks (ICTON)*, Angers, France, 2019. DOI: 10.1109/ICTON.2019.8840405

[22] R. Shiraki, Y. Mori, H. Hasegawa and K. Sato, “Dynamically controlled flexible-grid networks based on semi-flexible spectrum assignment and network-state-value evaluation,” presented at *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, USA, 2020. DOI: 10.1364/OFC.2020.M1B.4

[23] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA: The MIT Press, 2018.