AI-Empowered Maritime Internet of Things: A Parallel-Network-Driven Approach

Tingting Yang, Jiacheng Chen, and Ning Zhang

Abstract

As one of the key technologies for realizing a fully digitalized world, the Internet of Things (IoT) requires ubiquitous connections across both land and sea. However, due to lack of infrastructure such as optical fibers and base stations, maritime communications inevitably face a highly complex and heterogeneous environment, which greatly challenges the connection reliability and traffic steering efficiency for future service-oriented maritime IoT. With the recent burgeoning application of artificial intelligence (AI) in many fields, an AI-empowered autonomous network for maritime IoT is envisioned as a promising solution. However, AI typically involves training/learning processes, which require realistic data/environment in order to obtain valuable outcomes. To this end, this article proposes the parallel network, which can be regarded as the "digital twin" of the real network and is responsible for realizing four key functionalities: self-learning and optimizing, state inference and network cognition, event prediction and anomaly detection, and knowledge database and snapshots. We then explain how various AI methods can facilitate the operation of the parallel-network-driven maritime network. A case study is provided to demonstrate the idea. Research directions are also outlined for further studies.

Introduction

The prominent wireless communications technologies represented by 4G/5G and Wi-Fi have successfully served the burgeoning data demand coming from both humans and things for the past decade. In order to realize a full-fledged information society in the future, communications for vertical domains need to be further developed. One such domain is concerned with the ocean, since a maritime network can support a wide spectrum of maritime Internet of Things (IoT) services such as smart ports [1] and autonomous ship control [2]. Therefore, to facilitate both human and non-human activities at sea, a reliable and efficient network for maritime IoT is necessary.

Compared to its terrestrial counterpart, the maritime network faces the following two severe situations.

Environment complexity. On one hand, the communication range is scale-free, meaning that both long-distance (e.g., spanning an ocean) and near-shore communications should be supported. Solutions from terrestrial networks cannot work, since on land, short-range communication is realized via access networks (e.g., 4G/Wi-Fi), and long-range communication is realized through optical fibers. On the other hand, the marine radio environment is complicated and volatile, and is susceptible to sudden weather conditions such as storms.

Network heterogeneity. Due to environment complexity, it can be very difficult to implement maritime communications with a single type of network or communications technology. Therefore, the maritime network should be a heterogeneous space-air-ground-sea integrated network [3]. However, without proper network resource management, the gain of heterogeneity cannot be fully exploited.

Besides, the maritime network should also be service-oriented so as to meet different quality of service (QoS) requirements. Consequently, it is critical for the maritime network to connect things reliably and deliver services efficiently in the highly complex and heterogeneous communication environment.

Obviously, it is difficult to make appropriate decisions on how to utilize the resources under such conditions. We advocate that network intelligence and autonomy need to be enhanced so as to effectively address the above challenges. In recent years, artificial intelligence (AI) has gained unprecedented attention from both academia and industry. Fields such as computer vision, speech recognition, and natural language processing have made tremendous progress with the help of AI approaches. Due to the generality of AI, researchers have started applying these methods to communications and networking [4]. Most of these pioneering research works focus on solving specific problems with certain AI methods, and have proved the effectiveness of AI. However, still lacking are studies on how to apply AI methods to the maritime network in a systematic manner. Furthermore, AI methods typically require training/learning processes with realistic data/environment, which has become the bottleneck for applying AI.

This article studies how to adopt AI methods in the service-oriented maritime network. In the remainder of this article, we first introduce the real network, namely the physical substrate network, which is a heterogeneous space-air-ground-sea integrated network. It also adopts state-of-the-art technologies such as software defined networking (SDN), network slicing, and edge computing. Then we delineate the parallel network, which is the core component of the AI-empowered

This work was supported in part by Natural Science Foundation of China under Grant 61771086, Dalian Outstanding Young Science and Technology Talents Foundation under Grant 2017RJ06, Liaoning Province Xingliao Talents Program under Grant XLYC1807149, Guangdong Province Basic and Applied Basic Research Foundation under Grant 2019B1515120084, and the Verification Platform of Multi-tier Coverage Communication Network for Oceans (PCL2018KP002).

Digital Object Identifier: 10.1109/MNET.011.2000020

Tingting Yang is with Dongguan University of Technology and Peng Cheng Laboratory; Jiacheng Chen (corresponding author) is with Peng Cheng Laboratory; Ning Zhang is with the University of Windsor.

maritime network. The parallel network can be regarded as the "digital twin" of the real network. It is generated through a data-driven approach with realistic data from the real network. Hence, it can be used as an ideal platform for AI techniques to optimize resource management of the real network. We further explain how to apply AI methods in the parallel-network-driven maritime network. A case study is presented as an intuitive illustration on the benefits of the proposed method. Finally, open research issues are discussed, and the article is concluded.

Network Architecture and Enabling Technologies

The Space-Air-Ground-Sea Integrated Network

The space-air-ground-sea integrated maritime network is shown in Fig. 1. It is a heterogeneous network supporting satellite-based, airborne-based, and shore-based communication links. It also allows data exchange among peer users in an ad hoc manner. Thus, it can realize short to extremely long-distance communications. Also, the diversity of communications is helpful for handling the harsh and dynamic radio environment. Furthermore, the network operation can be fully decoupled, for example, control/data or uplink/downlink can use separate resources/links for transmission. Moreover, the design and application of network protocols should be flat and decentralized, such that the network operation can be fast and agile, which is critical for the heterogeneous network architecture. More descriptions are given below.

Shore-based communications: A ship can directly communicate with the base station (BS) located alongside the coastline. The BS can be a 5G BS, a dedicated maritime radio BS operating on special channels, or a cognitive-radio-based BS operating on TV white space (TVWS) channels. Basically, shore-based communications is a natural extension of terrestrial broadband communications, so it is easy to deploy and can provide high-bandwidth data services to users near the shore.

Relay-based communications: When a ship goes farther from the coast, it may use a relay to exchange data with the users on land. An airborne relay such as an unmanned aerial vehicle (UAV) [5] or a high-altitude platform (HAP) can be a preferable choice since it can be deployed in a flexible way. A buoy relay can also be used in certain scenarios, for example, as a maritime IoT gateway for collecting data from sensors floating around it.

Satellite-based communications: Wherever it is, a ship can always rely on satellite-based communications [6], which can provide global Internet access through low Earth orbit (LEO)/medium Earth orbit (MEO) constellations or geostationary Earth orbit (GEO) backbone networks. However, accessing via satellites usually has a lower priority than other access methods due to its relatively high latency and limited data rate.

Direct communications: Ships within a certain geographical area can form an ad hoc network and directly communicate with each other if Internet access is not required. One use case is a ship platoon, which is similar to a vehicle platoon on land. Direct communications can also be integrated with relay-based communications if one ship in the group shares its Internet access capability with others.

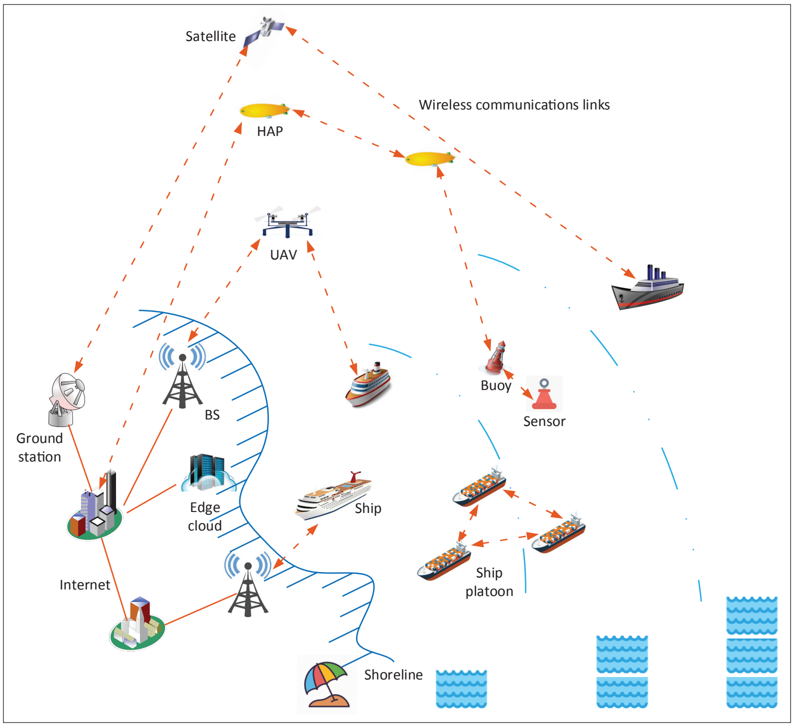

FIGURE 1. The space-air-ground-sea integrated maritime network.

SDN, Network Slicing, and Edge Computing

For such a heterogeneous network, network management involving interworking between different access technologies and cooperative resource allocation can be a great challenge [7]. To this end, SDN can be utilized for the space-air-ground-sea integrated network [8]. Recently, SDN has been widely adopted in various wireless networks [9], showing its generality as a network management and resource control tool. As shown in Fig. 2, SDN is able to virtualize the heterogeneous network substrate and provide a unified interface for network traffic orchestration. This is achieved through the separation of control and forwarding of network traffic. An SDN controller implements the network traffic control functionalities by setting the forwarding rules of each physical equipment in the network. With the help of an SDN controller, network slices can be generated for specific purposes. Each network slice is allocated dedicated network resources in order to guarantee its functionality. Applications can assign network slices for their users, considering users' access capabilities and QoS requirements. Edge computing is also useful for IoT [10] since it can provide computing resources close to end devices. Hence, an edge cloud is deployed at the seashore to provide accessible computation resources for the SDN controller and AI methods.

Parallel Network

Due to the complexity and heterogeneity of the maritime network, it is difficult to fully optimize the network in advance. To this end, we resort to AI methods and propose a parallel-network-driv-

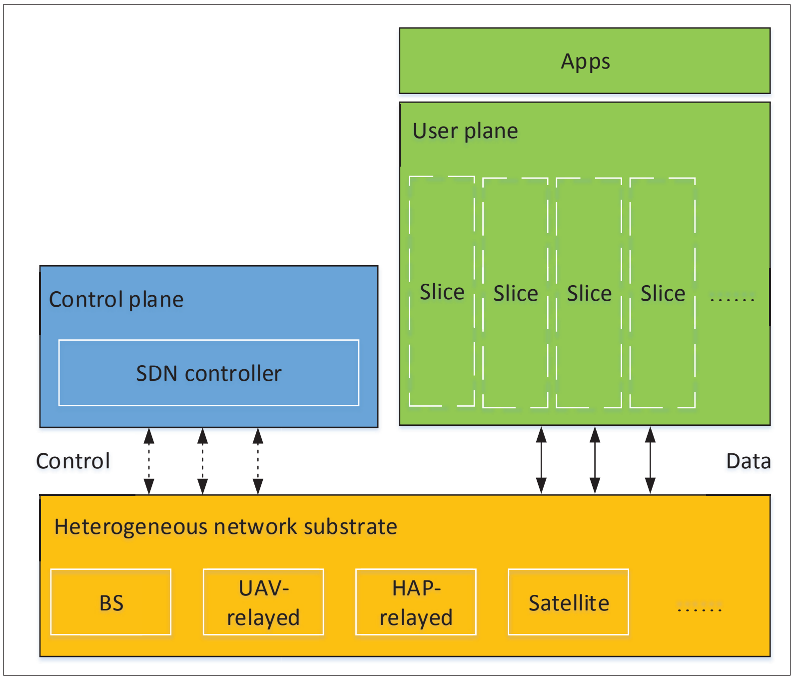

FIGURE 2. SDN and network slicing for the maritime network.

en approach to facilitate the application of AI in the maritime network. The concept of parallel network is illustrated in Fig. 3. Generally speaking, a parallel network is the "digital twin" of a real network and runs at the edge cloud. More specifically, it is a duplicate of the real network reproduced from realistic network state data, which may include node positions, channel qualities, link capacities, traffic flows, and so on, and even node energy level if it is an energy harvesting device [11]. A network sensing slice can be used to acquire the network state data. The sensing slice ensures the resources for bringing back information to parallel network with light probing packets that have negligible influences on the real network. The SDN controller can also control and manage the traffic flows in the parallel network. A parallel network has the following functionalities:

State inference and network cognition: In order to reproduce the real network, a parallel network is fed with realistic network state data. Hence, the timeliness and density of data will directly affect the accuracy of the parallel network. However, it is inevitable that data may be delayed and even lost during transmission. Also, the frequency of sending probing packets cannot be too high since network resources are limited. To this end, the parallel network should be able to infer the network state based on historical data and pre-built models, which can be continuously refined and rectified by the newly arriving data. Furthermore, by observing the state inference results, human engineers may have novel understandings on how the network performance is affected by external factors.

Self-learning and optimizing: It is difficult to find optimal decisions on how to utilize the network resources due to the complexity and heterogeneity. To this end, we can adopt AI-based learning methods for selecting appropriate access networks, data paths, and so on. These methods learn through interacting with the real network. In order to accelerate the learning process, especially the initial period when the learning model is not efficient enough, the parallel network can be used as a substitute for the real network since it can provide a realistic environment for the network controller to continuously self-improve without interfering with the real network.

Event prediction and anomaly detection: Except for reflecting the real network, the parallel network should also be able to "move one step forward in advance." Specifically, it should predict future incidental events such as node malfunction and channel quality degradation. In this way, precautions can be made such that service delivery may not be affected. Also, it should detect anomalous behaviors of network nodes, and thereafter trigger certain actions accordingly.

Knowledge database and snapshots: Data from both real and parallel networks can be permanently stored in a database as knowledge, which can be utilized to gain more insights on the network, such as pinpointing the network bottlenecks or cutting off redundant network resources, via data mining technologies. States of the parallel network at a certain time point can also be saved as network snapshots (i.e., historical states of network), which are useful for recovery from large-scale network breakdown.

AI Functions in the Maritime Network

In this part, we explain the important roles of AI algorithms in the proposed parallel-network-driven maritime network. We only focus on the general functions of different types of AI methods and identify their possible uses. Due to limited space, we do not discuss which specific algorithms should be selected and how algorithms should be specifically implemented.

Reinforcement Learning

Reinforcement learning has proven its superiority in solving decision making problems in deterministic environments such as games, from simple Atari arcade games to complicated Dota 2. Basically, reinforcement learning can be described with a Markov decision process (MDP), in which a state and an action leads to the next state regardless of the previous states (i.e., the Markov property). Each state is associated with a reward, and the objective is to maximize the total reward within certain steps of actions. Sometimes positive rewards are only defined in a few states; hence, reinforcement learning needs to be capable of finding the "true reason" for obtaining the reward. Reinforcement learning models are usually trained with a feedback loop process so as to determine the expected value of each action at each state. The recently developed deep reinforcement learning ingeniously integrates deep neural networks with traditional models.

Decision making problems are common in the maritime network (e.g., video transmission scheduling [12, 13]), and reinforcement learning is a promising solution due to the complexity and heterogeneity. For example, it can be used to select proper network access methods for end users based on their access capabilities and service requirements, as well as the current network state. Similarly, it is helpful to determine how to route traffic on appropriate data paths to end users. As mentioned earlier, the training process of reinforcement learning can be conducted in the parallel network. Furthermore, training techniques such as experience replay can be exploited with the data stored in the knowledge database.

Classification

Classification is one of the most widely used machine learning methods. Generally speaking, a classification model or hypothesis can figure out the class (or label) given the input data (i.e., features). The feature space is problem-specific and can be very large, such as all the pixels of an image. The label can be a binary indicator (e.g., 0 or 1) or have concrete meanings (e.g., a digit from 0 to 9). When the outcome of the model is continuous instead of discrete labels, the corresponding problem is often named regression. Classification belongs to a specific type of machine learning called supervised learning, which requires a large set of labeled data in advance to train the model.

In our scenario, at the very beginning, sufficient data need to be collected from the real network. Specifically, each data sample contains features including the current network state and decisions, and the corresponding label is the next state after decisions have been carried out. When the classification model is fully trained, the following state of the parallel network can be inferred given the current state and decisions. Also, classification can be used for anomaly detection. For example, suppose that we capture the state data of a network node. We also record the status of the node (i.e., whether it is working well or not). With such labeled datasets, a classification model can be built to detect anomalous working status of the node based on its real-time state data.

Clustering

Different from classification, clustering requires no labels on data samples; thus, it is categorized as unsupervised learning. As its name indicates, clustering is used to group data samples that are "close" in feature space into different clusters. Sometimes the target number of clusters is even unknown. The definition of distance between data samples is critical when using clustering algorithms in practice.

Clustering can also be used for anomaly detection in our scenario. Specifically, data samples are grouped based on certain features and distance metrics. The groups that are farther from other groups and smaller in size can be regarded as anomalies. Compared to classification, clustering can identify anomalies that have not appeared before since it does not require labeled anomalous data samples. Another way to use clustering is to group nodes in large-scale maritime ad hoc networks, such as a network of buoys. In [14], the authors show how to use big data to find potential connections among user devices and form user clusters accordingly. In this case, the spatial distance is used for the clustering algorithm. Since the positions of nodes are dynamic, a static grouping method is not applicable. Nodes in the same groups share the same communications resources and can communicate with each other with low latency and power consumption.

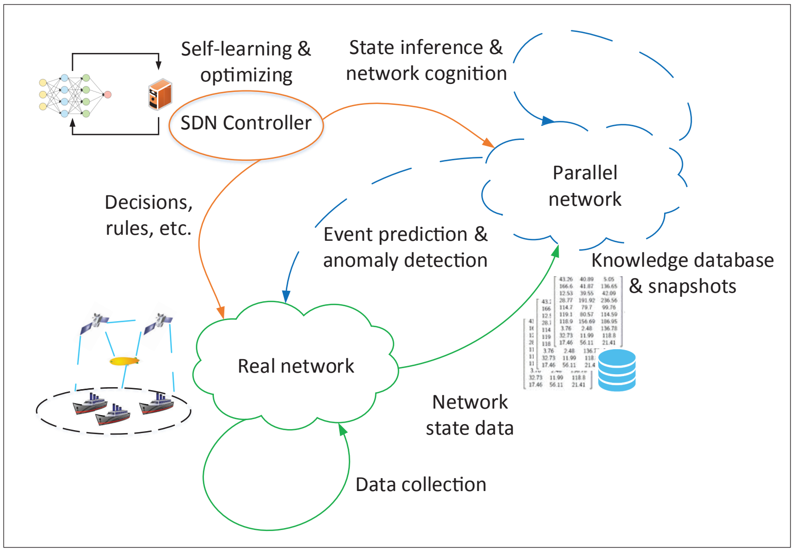

FIGURE 3. Relationship between the real network and the parallel network.

Miscellaneous

This subsection discusses some AI methods that are not covered previously. First, we briefly introduce time series forecasting. The problem works on data that are sequenced based on time, and aims to predict future values, sometimes with a confidence interval. AI methods concerned with time series such as long short-term memory (LSTM) are appropriate to deal with the problem. Time series prediction can be used for event prediction in a parallel network. At a larger scale, traffic burst can be predicted to avoid network congestion. Channel condition of an end user may also be predicted so as to prevent sudden service interruption.

Then we explore another innovative and powerful AI method, namely generative adversarial networks (GANs). A GAN consists of two neural networks called generator and discriminator. Generator produces data from noise, while discriminator determines whether the data is real or fake based on the input real data. After training, a GAN is able to produce real data with the generator. A GAN is commonly used for data production when the underlying distribution of data is unknown or intractable. In the field of communications, a GAN is a potential solution for channel modeling [15], and it is especially valuable for the maritime scenario, which has not been fully studied yet. In a parallel network, a GAN can also be used to generate realistic traffic, which can further be used for self-learning and optimizing purposes.

At last, we briefly investigate decentralized machine learning, such as federated learning. Simply speaking, federated learning trains local models in a decentralized environment, where edge devices only have local data samples. It is especially useful when the training datasets are heterogeneous. Hence, it may become applicable for the heterogeneous maritime network after nodes have more computation power.

A Case Study

This section presents a case study to illustrate the benefits of the proposed parallel-network-driven maritime network. We consider a ship navigating across an ocean. When the ship is far from land, it

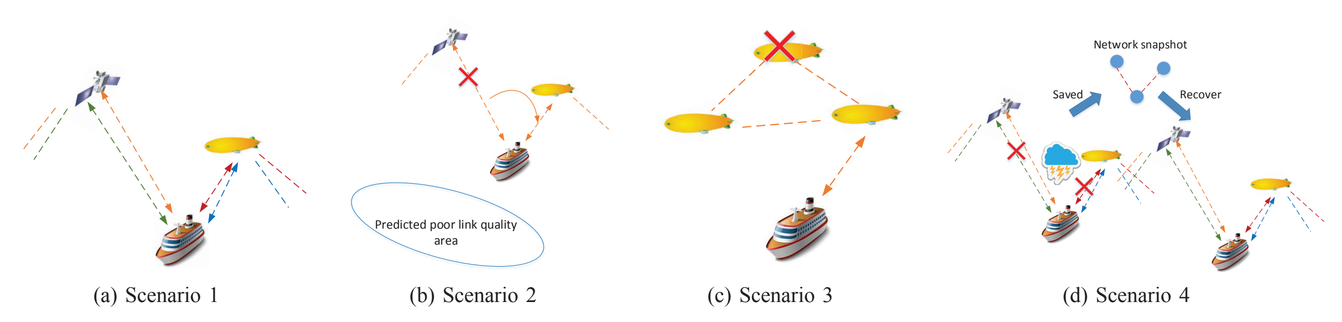

FIGURE 4. A case study: a) traffic steering for different services based on learning models; b) traffic migration based on link state prediction; c) route re-selection based on anomalous node detection; d) recovery from network snapshot.

can only access the Internet through satellites and HAPs. A specific network slice with such access capabilities is assigned for the ship. The following four scenarios are considered.

Scenario 1 (Fig. 4a): The ship is sailing on the sea, with both satellite and HAP access activated simultaneously for transmitting data. Traffic demands from the ship may include different types of services (e.g., bandwidth-intensive surveillance video streaming or delay-sensitive real-time sensing data). State data regarding the ship are continuously collected and reported to the parallel network. An AI model, such as a reinforcement learning model, has been trained for allocating resource, and is run by the SDN controller. Specifically, when new traffic demands arrive, the model will make decisions on assigning traffic flows to links with certain objectives (e.g., QoS guarantee or load balancing). The decisions are then implemented by the SDN controller. With the model, traffic demands of the ship can be guided to different data paths to meet the service requirements. Also, when a traffic burst occurs on one link and services transmitted on this link are affected, the original decisions will be re-evaluated, and a new decision on redirecting the traffic may be made by the model in order to avoid congestion and service quality degradation.

Scenario 2 (Fig. 4b): The ship, as well as many other ships, have traveled over a predefined route many times before, and the historical state data regarding link quality during all of the travel experiences have been recorded. A link quality prediction model is trained with the above time series related AI methods. This time, when the ship is approaching a certain area as it usually does, the model predicts that one link will probably have high packet loss rate based on the historical data. Consequently, the controller will suggest traffic on the link migrate to another link, maybe of another access type, in advance so as to prevent a possible link interruption.

Scenario 3 (Fig.4c): Several HAPs are providing Internet services to the ships within a certain area. The real-time throughput is reported to the parallel network. An anomaly detection model is built to supervise the working state of these nodes. Suddenly, the throughput of one HAP drops and diverges from the others. This unfamiliar pattern is detected by the model, which indicates that the HAP may encounter a failure. New routes should be arranged to circumvent the node.

Scenario 4 (Fig. 4d): When the ship is sailing in the middle of the ocean, a sudden storm appears and strikes the ship. All the radio links are severely affected, and communications can hardly be maintained. However, before the storm arrives, a network snapshot in the parallel network is saved. When the ship leaves the storm, the network can be rebuilt quickly based on the previous snapshot.

Conclusion and Future Works

In this article, we have introduced a systematic approach for applying AI methods in a service-oriented maritime network. We have presented the heterogeneous maritime network architecture, which is further enhanced by SDN, network slicing, and edge computing technologies. Multiple access methods are necessary for the maritime network to face the complex environment and meet diverse service requirements. Operation of the real network is driven by its duplicate, namely parallel, network. We have explained the functionalities of the parallel network and showed its significant role in optimizing resource allocation and maintaining network states. AI methods that can be used in the parallel-network-driven maritime network have been discussed, based on their functions. We have provided a case study to demonstrate the key utilities of parallel network and AI models. Here, we list several critical issues that need to be further studied.

Feature space definition: Although AI methods seem to be automatic, one important step usually needs human involvement (i.e., feature selection). For a complex system such as the heterogeneous maritime network, the selected features should be able to fully characterize it. This can be very challenging since it is difficult to choose the features straightforwardly like the pixels of an image. Also, not every piece of raw data should be used as a feature, or the feature space will be large and redundant and can slow the AI methods. The network designers should exploit their expertise and experience to select appropriate features.

Algorithm selection and evaluation: The field of AI is becoming ever more popular, and a plethora of new algorithms with various features and superiorities are developed from time to time. Therefore, how to select a suitable algorithm for each AI function in the maritime network requires careful consideration. On one hand, the algorithm itself should be evaluated from different aspects such as efficiency, stability, and complexity. On the other hand, practical issues when implementing the algorithm should also be considered, for example, whether the model is easy to tune.

Coping with errors: One thing that needs to be stressed is that AI methods may generate errors. Since the accuracy of an AI model is surely not 100 percent (i.e., errors cannot be avoided), measures should be taken to cope with errors. Also, we should design a system that is robust and resilient enough so as to tolerate a certain level of errors and degrade gracefully when errors occur.

References

[1] Y. Yang *et al.*, "Internet of Things for Smart Ports: Technologies and Challenges," *IEEE Instrumentation & Measurement Mag.*, vol. 21, no. 1, 2018, pp. 34–43.

[2] A. A. Allal *et al.*, "Toward a New Maritime Communication System in Detroit of Gibraltar Where Conventional and Autonomous Ships Will Co-Exist," *2017 Int'l. Conf. Wireless Networks and Mobile Commun.*, 2017, pp. 1–8.

[3] J. Liu *et al.*, "Space-Air-Ground Integrated Network: A Survey," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 4, 4th qtr., 2018, pp. 2714–41.

[4] Z. M. Fadlullah *et al.*, "State-of-the-Art Deep Learning: Evolving Machine Intelligence Toward Tomorrow's Intelligent Network Traffic Control Systems," *IEEE Commun. Surveys & Tutorials*, vol. 19, no. 4, 4th qtr. 2017, pp. 2432–55.

[5] X. Li *et al.*, "Maritime Coverage Enhancement Using UAVs Coordinated with Hybrid Satellite-Terrestrial Networks," arXiv:1904.02602, 2019.

[6] T. Wei *et al.*, "Hybrid Satellite-Terrestrial Communication Networks for the Maritime Internet of Things: Key Technologies, Opportunities, and Challenges," arXiv: 1903.11814, 2019.

[7] D. Kidston and T. Kunz, "Challenges and Opportunities in Managing Maritime Networks," *IEEE Commun. Mag.*, vol. 46, no. 10, Oct. 2008, pp. 162–68.

[8] N. Zhang *et al.*, "Software Defined Space-Air-Ground Integrated Vehicular Networks: Challenges and Solutions," *IEEE Commun. Mag.*, vol. 55, no. 7, July 2017, pp. 101–09.

[9] T. Chen *et al.*, "Software Defined Mobile Networks: Concept, Survey, and Research Directions," *IEEE Commun. Mag.*, vol. 53, no. 11, Nov. 2015, pp. 126–33.

[10] S. Li *et al.*, "Joint Admission Control and Resource Allocation in Edge Computing for Internet of Things," *IEEE Network*, vol. 32, no. 1, Jan./Feb. 2018, pp. 72–79.

[11] D. Zhang *et al.*, "Utility-Optimal Resource Management and Allocation Algorithm for Energy Harvesting Cognitive Radio Sensor Networks," *IEEE JSAC*, vol. 34, no. 12, 2016, pp. 3552–65.

[12] T. Yang and X. S. Shen, *Maritime Wideband Communication Networks: Video Transmission Scheduling*, Springer, 2014.

[13] T. Yang *et al.*, "Efficient Scheduling for Video Transmissions in Maritime Wireless Communication Networks," *IEEE Trans. Vehic. Tech.*, vol. 64, no. 9, 2014, pp. 4215–29.

[14] S. Han *et al.*, "Dense-Device-Enabled Cooperative Networks for Efficient and Secure Transmission," *IEEE Network*, vol. 32, no. 2, Mar./Apr. 2018, pp. 100–06.

[15] Y. Yang *et al.*, "Generative Adversarial-Network-Based Wireless Channel Modeling: Challenges and Opportunities," *IEEE Commun. Mag.*, vol. 57, no. 3, Mar. 2019, pp. 22–27.

Biographies

Tingting Yang [M'13] (yangtingting820523@163.com) received her Ph.D. degree from Dalian Maritime University, China, in 2010. She is currently a professor at Dongguan University of Technology, China. Since September 2012, she has been a visiting scholar at the Broadband Communications Research (BBCR) Lab at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. Her research interests are in the areas of maritime wideband communication networks, DTN networks, and green wireless communication. She serves as the Associate Editor-in-Chief of *IET Communications*, as well as an Advisory Editor for SpringerPlus. She also serves as an Associate Chair for IEEE ICC '20 and ICC '21.

Jiacheng Chen (chenjch02@pcl.ac.cn) received his Ph.D. degree in information and communications engineering from Shanghai Jiao Tong University, China, in 2018. From December 2015 to December 2016, he was a visiting scholar at BBCR Lab, University of Waterloo. Currently, he is an assistant researcher at Peng Cheng Laboratory, Shenzhen, China. His research interests include future network design, 5G/6G networks, and resource management.

Ning Zhang [M'15, SM'18] (ning.zhang@ieee.org) received his Ph.D. degree from the University of Waterloo in 2015. After that, he was a postdoctoral research fellow at the University of Waterloo and the University of Toronto, Canada, respectively. He serves as an Associate Editor of the *IEEE Internet of Things Journal*, *IEEE Transactions on Cognitive Communications and Networking*, *IEEE Access*, *IET Communications*, and *Vehicular Communications* (Elsevier); and a Guest Editor of several international journals, such as *IEEE Wireless Communications*, *IEEE Transactions on Industrial Informatics*, and *IEEE Transactions on Cognitive Communications and Networking*. He also serves/has served as a Track Chair for several international conferences and a Co-Chair for several international workshops. He received the Best Paper Awards from IEEE GLOBECOM in 2014, IEEE WCSP in 2015, the *Journal of Communications and Information Networks* in 2018, IEEE ICC in 2019, IEEE Technical Committee on Transmission Access and Optical Systems in 2019, and IEEE ICCC in 2019, respectively.