ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

A survey of Machine Learning applied to Computer Networks

Alexander Gepperth[1] and Sebastian Rieger[1]

1- Fulda University of Applied Sciences - Applied Computer Science
Leipziger Straße 123, 36037 Fulda - Germany

Abstract. We review the current state of the art in the domain of machine learning applied to computer networks. First of all, we describe recent developments in computer networking and outline the potential fields for machine learning that arise from these developments. We discuss challenges for machine learning in this particular field, namely the inherent big data aspect of computer networks, and the fact that learning very often needs to be conducted in a streaming setting with non-stationary data distributions. We discuss practical issues like privacy protection and computing resources before finally outlining potential technological benefits of this emerging scientific field.

1 Introduction

The application of machine learning techniques in the area of computer networks is a very promising area of study. Steadily growing network traffic, especially in the Internet or in huge data centers, as well as the continuously increasing number of endpoints due to mobile devices and Internet of Things (IoT) applications, implies an enormous management and monitoring effort when designing and operating these networks. Network automation mechanisms based on, e.g., Software-Defined Networking (SDN), can leverage advanced telemetry solutions to allow fine-grained traffic management. Large scale data transfer and ever-growing bandwidths raise the demand for open-loop network management assistance or sustained closed-loop auto-remediation, self-healing or -optimization techniques.

On the machine learning side, challenges arise due to the inherent "big data" aspect of network traffic. Because of this, learning often has to be conducted "on the fly" on streaming data, without storing any data at all. Learning is further complicated by the non-stationarity of the data which can produce the catastrophic forgetting effect to which Deep Neural Networks (DNN) are particularly vulnerable[1]. Last but not least, the acquisition of sufficient amounts of good-quality training data is often difficult due to privacy protection issues, and the results of machine learning are often not generalizable because all networks and their users have strongly individual characteristics.

Methods or applications of machine learning for computer network management and monitoring:

- machine learning for network automation and programmability in the data, control, management or knowledge plane cognitive/autonomic network management and monitoring

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

- machine learning in network fault, configuration, accounting, performance or security management
- big data and deep learning approaches for network traffic engineering and routing
- in-network computing using machine learning
- streaming (network) data processing by deep neural networks
- continual learning on network traffic data
- acquisition of training data from computer networks and generalization of results

2 Survey of Recent Works

Over the last years, several surveys and conference tracks regarding the application of machine learning for computer networks have been established. For example, [2] gives an overview of different applications for machine learning in computer networks and also shows up current advancements and opportunities. The applications identified in the article include information cognition, traffic prediction, traffic classification, resource management, network adaption, performance prediction and configuration extrapolation. Another survey presented in [3] categorizes possible applications in traffic prediction, traffic classification, traffic routing, congestion control, resource management, fault management, QoS and QoE management and network security. Both surveys mention traffic classification and traffic prediction among the first applications for machine learning in networking. These subfields of network management also leveraged machine learning to enhance traffic engineering and control in earlier publications [4, 5]. Another example for using machine learning to classify network traffic flows and their throughput in simulated data center networks is presented in [6]. In contrast to simulated traffic used in these publications however, flow characteristics in real-world networks are often fluctuating and complex [7]. Abrupt and gradual changes in computer networks and traffic characteristics can pose a significant challenge for the application of machine learning models. Machine learning based flow size prediction used for improved routing can be found in [8].

As proposed for knowledge-defined networking [9] or cognitive network management [10], machine learning techniques in networking can especially be combined with network monitoring and network softwarization and virtualization established in SDN [11, 12] and Network Functions Virtualization [13]. The combination of machine learning based classification with SDN, forming an adaptive traffic engineering framework is presented in [14]. An approach that proposes the use of deep reinforcement learning on synthetic network traffic for routing can be found in [15]. [16] introduces the combination of learning from existing Dijkstra-based routing algorithms and imitating them with higher performance using a dynamic routing framework for SDN to optimize network throughput

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

based on simulated data. A solution with supervised deep learning for routing decisions based on real traffic demands is presented in [17]. The model is using aggregated known traffic demands as an input to optimize the overall path utilization. [18] analyzes real-world network flows captured from a university campus network. Thereby, the prevalence of small flows and the classification of features is discussed and the importance of data collection in real networks is emphasized. The relevance of deep learning based traffic classification and prediction in SDN is explained in [19]. Traffic analysis and routing optimization with deep learning are explicitly named as major future research problems. This is also supported by the necessary shift from rule-based network traffic control to mechanisms using artificial intelligence (AI), e.g., due to steadily increasing traffic volumes [20]. [21] argues that a network AI can be used to predict future network traffic from past data to evolve network management and automation. Using a network AI, [22], [23] and [24] focus on intelligent traffic routing for aggregated traffic characteristics and improved network analytics.

For verification, prediction models can be cross-checked, e.g., with existing evaluations of the interpretability of deep learning models used in the area of computer networks [25]. An interesting option is a generative replay approach, whereby generated characteristic traffic is combined with prior data to ensure the adaptability of the prediction model [26].

3 Challenges in the Area of Computer Networks

High bisection and link bandwidths, offer large amounts of data to be used for machine learning. First, the mere data volume continuously transferred over currently prevalent link speeds around 10 Gbit/s accounts for ≈4,5 terabyte per hour. While this volume is created on a single link, even small local area networks of mid-sized organizations can contain thousands of individual links. Furthermore, presently available Ethernet link speeds allow rates of up to 400 Gbit/s. For example, in 2015 Google presented the Jupiter datacenter network generation, that is used in a single Google datacenter, to have a bisection bandwidth of 1.3 Pbit/s (≈1,000,000 Gbit/s) [27]. The bisection bandwidth is defined by the sum of the bandwidths used by the links that need to be cut to section the network to two roughly equally sized partitions. As mentioned in [27], the bisection bandwidth of Google's datacenter networks is constantly increasing (from 2 Tbit/s in 2004, 10 Tbit/s in 2006, 82 Tbit/s in 2008, 207 Tbit/s in 2009 to the aforementioned 1.3 Pbit/s in 2012) and can be assumed to be much higher in the recent past.

Consequently, collecting and processing data being transferred over large computer networks holds significant challenges. When collecting the traffic transferred over multiple links in the same network, a large amount of data is redundant, as multiple subsequent links (along a path in the network) transfer the same packet sequence. However, to analyze the traffic characteristics in a network, each link or at least path has to be considered, e.g., to allow for efficient network utilization and traffic flow prediction. Furthermore, collecting

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

the data on multiple links is necessary if the overall state of the network (e.g., link or component outage, latency, packet loss etc.) is used in the machine learning model. This also includes changes in the topology and configuration of the network. Therefore, a common approach is not to simply use the entire payload (as, e.g., offered by pcap, RSPAN etc.), but rather only (aggregated) metadata of the transferred packets. A typical solution is to extract only parts of the payload and network state (e.g., INT) or focus on the packet headers, e.g., containing addresses, type and length information etc. (e.g., NetFlow, sFlow, IPFIX). Aggregation can be based, e.g., on a network flow, for example a single metadata record per download (esp., a 5-tuple containing source IP address, destination IP address, source transport port, destination transport port, transport protocol) or further aggregated forms of management data (e.g., SNMP, NETCONF, RESTCONF, YANG). Nonetheless, also these aggregated forms of network traffic data collection can lead to multiple gigabytes to terabytes of data per day, as, e.g., observed in university campus core networks [18, 28, 29].

For predictive analysis or network management, the data needs to be collected over longer time frames, as traffic in a network is typically fluctuating and sudden spikes are common, e.g., due to popular downloads or external events. Collecting training data to be used for machine learning only on short intervals (e.g., seconds or minutes) might not contain these spikes, consequently negatively impacting the accuracy of the machine learning model, e.g., used for prediction. Besides these abrupt events in the state, also gradual changes in the state of the computer network need to be considered for predictive network management and monitoring. For example, the overall consumed bandwidths and hence utilization of links is constantly rising. This effect can be observed, e.g., by looking at the 5-year traffic statistics of DE-CIX as one of the largest internet exchange points worldwide [30] (currently transferring a peak traffic of ≈8,000 Gbit/s). Besides the constant increase in traffic, also the aforementioned fluctuation can be derived from the peaks while looking at these statistics. Another fluctuation can be observed in monthly or weekly statistics of these internet exchange points. As expected, traffic peaks are higher in the evening compared to other times of day and comparatively low in the early morning. These gradual, periodic, seasonal and abrupt changes need to be considered as concept drift for predictive analytics and machine learning.

The challenges posed by velocity and volume (as well as redundancy) of data can be addressed using existing solutions, e.g., leveraging related big data techniques. Abrupt fluctuation and gradual changes in the network state and topology can be addressed by applying machine learning models and predictive analytics to different time and learning intervals. Also, considering context information from the network (e.g., paths, topology etc.), e.g., originating from existing network management and monitoring or automation mechanisms can help to consider these effects in the prediction models. All the challenges need to be taken into account to allow a generalization of computer network models used for machine learning, leading to better accuracy and hence applicability of resultant mechanisms like predictive analysis and management.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

4 Challenges in the Area of Machine Learning

The challenges for machine learning when applied to computer networks are manifold:

Physical data acquisition In computer networks, common hardware components that have direct access to network traffic (like, e.g., routers or switches) have limited computational capacity, which makes it necessary to export the data to more powerful devices. This can lead to problems with export protocols (like NetFlow) that are vendor-specific and contain hidden parameters that cannot be changed.

Privacy protection Analyzing network traffic, even if just connection meta-information (protocol, sender, receiver) is involved, requires access to potentially sensitive information that is, to various degrees, protected by national laws. Acquiring and even publishing such data therefore requires a complete, documented processing chain that ensures that sensitive information like IP addresses are anonymized while preserving essential content, and that the data acquisition process does not lead to additional security risks.

Big data and streaming data Data throughput in large computer networks is significant. Apart from issues related to the collection of data, training machine learning models in real time poses an even bigger challenge: first of all, it is out of the question to store all of the data for later training. Rather, individual samples have to be processed as they arrive, using small mini-batches at most with few training iterations. This imposes strong constraints on the complexity of machine learning models, but also on the types of models that can be used (tale, e.g., SVMs which are ruled out by these constraints).

Non-stationarity and catastrophic forgetting A very important issue related to the ”streaming data” aspect is the problem of non-stationary data. If data are collected over a certain time period, one can always randomly shuffle collected data and ensure an approximately uniform data distribution. However, computer network traffic can exhibit highly non-stationary characteristics (to see this, consider WiFi utilization in an university campus on weekdays or week-ends) on time scales that are not always obvious: if training is conducted in a streaming fashion, the machine learning model will, over time, be exposed to strongly varying data distributions. Such a temporal variability is known to lead to an effect termed catastrophic forgetting [1] in most current machine learning models (DNNs, SVMs linear classifiers) and must be actively avoided. This can be done in various fashions, either by maintaining statistically significant holdout datasets (termed ”replay buffers” in reinforcement learning literature, see, e.g., [31]) or by using models that are (more) robust against catastrophic forgetting (see, e.g., [1, 32]).

Time-variable class imbalances A potential consequence of non-stationary data distributions is, at least for classification settings, a variable proportion of individual classes over time. This is a challenge for training strategies that rely on static per-class re-weighting of loss gradients, or on oversampling based on fixed assumptions about class frequencies. Such strategies would have to be

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

adapted over time, introducing additional hyper-parameters like time constants, thus leading to additional complexities.

Validity and evaluation of trained models A central assumption in machine learning (see, e.g., [33]) is that samples for training and testing/applying the model are drawn from the same underlying probability distribution. For data whose distribution is non-stationary over time (see previous paragraphs), this obviously no longer holds, which means that a model trained on data from time interval $T = [t_1, t_2]$ will not necessarily be valid outside T, in particular for $t \gg t_2$. This poses problems especially for evaluating model performance and the interpretation of such performance measures (see [34] for a related discussion).

5 Resulting Opportunities for Machine Learning

It is often the case that new application domains motivate new theoretical developments. Especially considering the challenges stated in Sec. 4, the investigation of the following aspects would facilitate the deployment of machine learning in computer networks: in the first place, we see the the development of efficient models that are more robust to changing data statistics than DNNs are (see, e.g., [1, 35, 32, 36] for some potential approaches), or else the modification of DNN models in that sense. In computer networking, it is often not important (or even possible) to train perfectly accurate models, whereas execution speed is often a more significant criterion. Furthermore, what would be very helpful are theoretical developments concerning the guarantees that can be given for non-stationary data distribution, in analogy to the guarantees provided by conventional statistical learning theory [33].

6 Technological Perspectives

Machine learning applied to the field of computer networks allows significant enhancements especially in the area of distributed network management and monitoring. These can be classified in open-loop and closed-loop management systems. Open-loop systems can be used to analyze the state of the network and address the challenges mentioned in Section 3. Due to the aforementioned volume, velocity and fluctuation in current network deployments, planing and maintaining network capacity and traffic engineering poses a significant challenge for network administrators. Using open-loop systems, decisions of these administrators can be assisted using, e.g., suggested or semi-automated tasks derived or predicted from machine learning models of the networks. However, velocity and abrupt changes in the network and transported traffic characteristics also lead to requirement for closed-loop systems. Related work has been carried out as adaptive management, self-healing or auto-remediating network management in the past decades. Certainly, the generalizability and applicability of these approaches is cumbersome. Yet, recent advancements in the area of network programmability and automation allow (e.g., originating from SDN and Network Functions Virtualization - NFV) for an application of predictive analytics and maintenance for specific tasks in network management and monitoring. These can be applied to the common management categories formed by the classic FCAPS (fault, configuration, accounting, performance and security management) paradigm of computer network management [37].

For example, fault management can be supported by proactive or predictive analytics and maintenance as well as correlation techniques leveraging machine learning. Configuration management can benefit from the automation of configuration and deployment tasks using machine learning. Accounting management can use predictions, e.g., to evaluate customer profiles. In the area of performance management, machine learning models can be used for capacity planning and predictive traffic engineering. Security management can use machine learning to detect network attack patterns and anomalies as well as applying resilient remediation and mitigation.

7 Conclusion and Future Work

The related work dealing with the application of machine learning in computer networks discussed in this contribution together with the identified challenges and opportunities support the relevance of this emerging scientific field. This is further strengthened by presented recent surveys and scientific conference titles and tracks in this area. Overall, machine learning leverages analysis and processing of network management information residing in the control and management plane. Often, machine learning is therefore described as applied in an intermediate plane between control and management, also referred to as an open- or closed-loop knowledge plane [9]. However, by driving network automation and programmability down to the data plane, i.e., by using concepts like programming protocol-independent packet processors (P4) [38], machine learning does not have to reside only between control and management plane. P4 offers programmability inside the data plane, also referred to as in-network computing. This allows for faster and more fine-grained extraction of network metadata (e.g., in-network or streaming telemetry compared to classic push- or pull-based monitoring), as data being sent from data plane to upper layers, as well as faster and more fine-grained handling and manipulation of data, i.e., packets, within the data plane. This way, the intended state of the network (e.g., fine-grained load balance and fault tolerance) as also alluded to in the upcoming network management paradigm intent-based networking [39], can be ensured in a timely and automated manor on the control as well as the data plane, leading to efficient operation and utilization of the network. Timely in this case especially means, that reactions and proactive measures can be applied in time frames of minutes and seconds or even fractions of seconds, while changes on the management plane, operated by humans, typically happen in much greater intervals. Also, parts of the machine learning application can be placed within the data layer, e.g., in the sense of data preprocessing or collaborative and distributed data analysis. This way, computational intensive machine learning approaches can be placed or centralized in planes on a higher level (e.g., control, management)

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

and data with high velocity can be pre-processed decentralized directly in lower levels, i.e. the data plane. Nonetheless, also human network management on the management plane can benefit from decision support, e.g., based on predictive analytics as mentioned for open-loop network management systems leveraging machine learning.

References

[1] B Pfülb and A Gepperth. A comprehensive, application-oriented study of catastrophic forgetting in DNNs. In *International Conference on Learning Representations (ICLR)*, 2019.

[2] Mowei Wang, Yong Cui, Xin Wang, Shihan Xiao, and Junchen Jiang. Machine Learning for Networking: Workflow, Advances and Opportunities. *IEEE Network*, 2018.

[3] Raouf Boutaba, Mohammad A Salahuddin, Noura Limam, Sara Ayoubi, Nashid Shahriar, Felipe Estrada-Solano, and Oscar M Caicedo. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *Journal of Internet Services and Applications*, 9(1):16, 2018.

[4] Zubair Md Fadlullah, Fengxiao Tang, Bomin Mao, Nei Kato, Osamu Akashi, Takeru Inoue, and Kimihiro Mizutani. State-of-the-Art Deep Learning: Evolving Machine Intelligence Toward Tomorrow's Intelligent Network Traffic Control Systems. *IEEE Comm. Surveys & Tutorials*, 2017.

[5] Thuy Nguyen and Grenville Armitage. A Survey of Techniques for Internet Traffic Classification using Machine Learning. *IEEE Comm. Surveys*, 2008.

[6] Houman Rastegarfar, Madeleine Glick, Nicolaas Viljoen, Mingwei Yang, John Wissinger, Lloyd LaComb, and Nasser Peyghambarian. TCP Flow Classification and Bandwidth Aggregation in Optically Interconnected Data Center Networks. *Journal of Optical Communications and Networking 8*, 2016.

[7] Theophilus Benson, Aditya Akella, and David A Maltz. Network Traffic Characteristics of Data Centers in the Wild. In *10th ACM SIGCOMM Conf. on Internet Measurement*, 2010.

[8] Pascal Poupart, Zhitang Chen, Priyank Jaini, Fred Fung, Hengky Susanto, Yanhui Geng, Li Chen, Kai Chen, and Hao Jin. Online Flow Size Prediction for Improved Network Routing. In *IEEE 24th Internat. Conf. on Network Protocols*, 2016.

[9] Albert Mestres, Alberto Rodriguez-Natal, Josep Carner, Pere Barlet-Ros, Eduard Alarcón, Marc Solé, Victor Muntés-Mulero, David Meyer, Sharon Barkai, Mike J Hibbett, et al. Knowledge-Defined Networking. *ACM SIGCOMM Computer Communication Review*, 2017.

[10] Sara Ayoubi, Noura Limam, Mohammad A Salahuddin, Nashid Shahriar, Raouf Boutaba, Felipe Estrada-Solano, and Oscar M Caicedo. Machine Learning for Cognitive Network Management. *IEEE Communications Magazine*, 2018.

[11] Diego Kreutz, Fernando MV Ramos, Paulo Verissimo, Christian Esteve Rothenberg, Siamak Azodolmolky, and Steve Uhlig. Software-Defined Networking: A Comprehensive Survey. *Proc. of the IEEE*, 2015.

[12] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. OpenFlow: Enabling Innovation in Campus Networks. *ACM SIGCOMM Computer Communication Review*, 2008.

[13] Rashid Mijumbi, Joan Serrat, Juan-Luis Gorricho, Niels Bouten, Filip De Turck, and Raouf Boutaba. Network Function Virtualization: State-of-the-Art and Research Challenges. *IEEE Comm. Surveys & Tutorials*, 2016.

[14] Pu Wang, Shih-Chun Lin, and Min Luo. A Framework for QoS-aware Traffic Classification using Semi-supervised Machine Learning in SDNs. In *IEEE Internat. Conf. on Services Computing*, 2016.

[15] Asaf Valadarsky, Michael Schapira, Dafna Shahaf, and Aviv Tamar. Learning to Route. In *16th ACM Workshop on Hot Topics in Networks*, 2017.

[16] Abdelhadi Azzouni, Raouf Boutaba, and Guy Pujolle. NeuRoute: Predictive Dynamic Routing for Software-Defined Networks. *13th International Conference on Network and Service Management*, 2017.

[17] J. Reis, M. Rocha, T. K. Phan, D. Griffin, F. Le, and M. Rio. Deep neural networks for network routing. In *2019 Internat. Joint Conf. on Neural Networks*, 2019.

[18] Piotr Jurkiewicz, Grzegorz Rzym, and Piotr Borylo. Flow length and size distributions in campus internet traffic. *Computing Research Repository*, 2018.

[19] A. R. Mohammed, S. A. Mohammed, and S. Shirmohammadi. Machine learning and deep learning based traffic classification and prediction in software defined networking. In *2019 IEEE Internat. Symp. on Measurements Networking*, 2019.

[20] J. Xu and K. Wu. Living with artificial intelligence: A paradigm shift toward future network traffic control. *IEEE Network*, 2018.

[21] Jiao Zhang, Tao Huang, Shuo Wang, and Yun-jie Liu. Future internet: trends and challenges. *Frontiers of Information Technology & Electronic Engineering*, 2019.

[22] H. Yao, T. Mai, C. Jiang, L. Kuang, and S. Guo. Ai routers & network mind: A hybrid machine learning paradigm for packet routing. *IEEE Comp. Intelligence Magazine*, 2019.

[23] Krzysztof Rusek, José Suárez-Varela, Albert Mestres, Pere Barlet-Ros, and Albert Cabellos-Aparicio. Unveiling the potential of graph neural networks for network modeling and optimization in sdn. In *Proc. of the 2019 ACM Symp. on SDN Research*. Association for Computing Machinery, 2019.

[24] Z. Zhuang, J. Wang, Q. Qi, H. Sun, and J. Liao. Toward greater intelligence in route planning: A graph-aware deep learning approach. *IEEE Systems Journal*, 2019.

[25] Ying Zheng, Ziyu Liu, Xinyu You, Yuedong Xu, and Junchen Jiang. Demystifying deep learning in networking. In *Proc. of the 2nd Asia-Pacific Workshop on Networking*. ACM, 2018.

[26] Sebastian Lettner and Andreas Blenk. Adversarial network algorithm benchmarking. In *Proc. of the 15th Internat. Conf. on Emerging Networking EXperiments and Technologies*. Association for Computing Machinery, 2019.

[27] Arjun Singh, Joon Ong, Amit Agarwal, Glen Anderson, Ashby Armistead, Roy Bannon, Seb Boving, Gaurav Desai, Bob Felderman, Paulie Germano, et al. Jupiter rising: A decade of Clos topologies and centralized control in google's datacenter network. *ACM SIGCOMM Computer Communication Review*, 45(4):183–197, 2015.

[28] Christoph Hardegen, Benedikt Pfülb, Sebastian Rieger, Alexander Gepperth, and Sven Reißmann. Flow-based throughput prediction using deep learning and real-world network traffic. In *International Conference on Network and Service management*. IEEE, New York, 2019.

[29] Benedikt Pfülb, Christoph Hardegen, Alexander Gepperth, and Sebastian Rieger. A study of deep learning for network traffic data forecasting. In *International Conference on Artificial Neural Networks*, pages 497–512. Springer, 2019.

[30] DE-CIX location Frankfurt 5-year/1-month statistics. `https://www.de-cix.net/en/locations/germany/frankfurt/statistics`, 2020.

[31] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017.

[32] German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 2019.

[33] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

[34] A Gepperth and B Hammer. Incremental learning algorithms and applications. In *European Sympoisum on Artificial Neural Networks (ESANN)*, 2016.

[35] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. Continual learning: A comparative study on how to defy forgetting in classification tasks. *arXiv preprint arXiv:1909.08383*, 2019.

[36] Jonathan Schwarz, Jelena Luketina, Wojciech M Czarnecki, Agnieszka Grabska-Barwinska, Yee Whye Teh, Razvan Pascanu, and Raia Hadsell. Progress & compress: A scalable framework for continual learning. In *International Conference on Machine Learning (ICML)*, 2018.

[37] Alexander Clemm. *Network management fundamentals.* Cisco Press, 2006.

[38] Pat Bosshart, Dan Daly, Glen Gibb, Martin Izzard, Nick McKeown, Jennifer Rexford, Cole Schlesinger, Dan Talayco, Amin Vahdat, George Varghese, et al. P4: Programming Protocol-Independent Packet Processors. *ACM SIGCOMM Computer Comm. Review*, 2014.

[39] Haipeng Yao, Chunxiao Jiang, and Yi Qian. Intention based networking management. In *Developing Networks using Artificial Intelligence*, pages 199–244. Springer, 2019.