THE CONVERGENCE OF MACHINE LEARNING AND COMMUNICATIONS

Wojciech Samek[1], Slawomir Stanczak[1,2], Thomas Wiegand[1,2]
[1]Fraunhofer Heinrich Hertz Institute, 10587 Berlin, Germany
[2]Dept. of Electrical Engineering & Computer Science, Technische Universität Berlin, 10587 Berlin, Germany

Abstract - The areas of machine learning and communication technology are converging. Today's communication systems generate a large amount of traffic data, which can help to significantly enhance the design and management of networks and communication components when combined with advanced machine learning methods. Furthermore, recently developed end-to-end training procedures offer new ways to jointly optimize the components of a communication system. Also, in many emerging application fields of communication technology, e.g. smart cities or Internet of things, machine learning methods are of central importance. This paper gives an overview of the use of machine learning in different areas of communications and discusses two exemplar applications in wireless networking. Furthermore, it identifies promising future research topics and discusses their potential impact.

Keywords – Artificial intelligence, communications, machine learning, wireless networks

1. INTRODUCTION

We are currently observing a paradigm shift towards "smart" communication networks that take advantage of network data. In fact, modern communication networks, and in particular mobile networks, generate a large amount of data at the network infrastructure level and at the user/customer level. The data in the network contains a wealth of useful information such as location information, mobility and call patterns. The vision of network operators is to either enable new businesses through the provisioning of this data (or the information contained within it) to external service providers and customers or to exploit the network data for in-house services such as network optimization and management.

In order to make the vision reality, there is a strong need for the development and implementation of new machine learning methods for big data analytics in communication networks. The objective of these methods is to extract useful information from the network data while taking into account limited communication resources, and then to leverage this information for external or in-house services.

Moreover, machine learning methods are a core part in many emerging applications of communication technology, e.g. smart cities [47] or the Internet of things [46]. Here, topics such as monitoring, fault prediction and scheduling are addressed with modern learning algorithms. The use of machine learning methods in communications may provide information about individuals that affect their privacy. Therefore, various privacy-preserving approaches to data analysis have been recently proposed (e.g. [1]). Machine learning methods are also widely applied to tackle security-related problems in communications, e.g. as part of defense mechanisms against spam attacks and viruses [21].

The increasing convergence can be also observed in specific domains of communications such as image and video communication. While the direct approach to designing compression algorithms using autoencoders has provided very limited results compared to the state-of-the-art approach, the use of machine learning as an enhancing component for aspects like video encoding, bit allocation or other parts became a promising research direction [55]. As most video signals are stored as compressed data, the topic of object recognition and tracking in the compressed domain is also of high relevance [41]. Video streaming is another application which benefits from the use of learning algorithms [28].

Despite the successful use of machine learning methods in various communication applications, there are still many challenges and questions that need to be addressed. For instance, the large size and high computational demands of modern machine learning algorithms prevent the large-scale use of these models in embedded devices. Also 5G networks call for novel machine learning-based approaches to radio resource management and

network management approaches that can cope with uncertainties and incomplete channel and network state information. Other problems concern reliability, privacy and security aspects of machine learning models.

The following clauses review the literature (clause 2), present two applications (clause 3) and discuss future research topics in machine learning and communications (clause 4). The paper concludes with a summary (clause 5).

2. MACHINE LEARNING IN COMMUNICATIONS

This clause reviews the use of machine learning algorithms in different application fields of communications (see Figure 1).

2.1. Communication networks

Routing has a significant impact on the network's performance, and it is a well-studied topic in communications. Machine learning methods have been used to tackle different types of routing problems in the past, including shortest path routing, adaptive routing and multicasting routing. The authors of [7] proposed an algorithm for package routing in dynamically changing networks based on reinforcement learning. This algorithm learns a routing policy which balances between the route length and the possibility of congestion along the popular routes. Extensions on this idea have been proposed in [27]. Other researchers approached the routing problem with genetic algorithms [34]. Here alternative routes are created by crossover and mutation of the existing routes. Genetic algorithms have been also used for tackling the multicasting routing problem which emerges when data is sent to multiple receivers through a communication network [54]. Also, in mobile ad-hoc networks the construction of multicast trees has been addressed using genetic algorithms. Here additional objectives such as bounded end-to-end delay and energy efficiency are added to the optimization [30].

Several works (e.g. [17]) have also used machine learning techniques for throughput or traffic prediction in communication networks. This is an important topic as with a dynamic throughput control and allocation one can fulfill the quality of service (QoS) requirements while efficiently utilizing the network resources. For instance, the authors of [28] applied neural networks for variable-bit-rate video traffic prediction in order to dynamically allocate throughput for real-time video applications. Traffic identification is another important topic for network operators as it helps them to manage their networks, to assure the QoS and to deploy security measures.

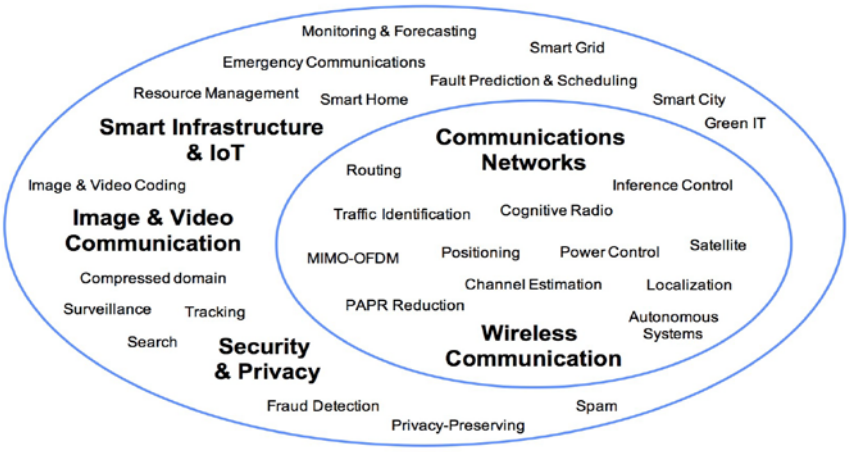

Fig. 1. Applications on machine learning in different areas of communications.

Here, machine learning methods recognize statistical patterns in the traffic data by analyzing packet-header and flow-level information. An excellent review of traffic classification with machine learning methods is [35].

2.2. Wireless communications

To achieve a high efficiency at the desired QoS, it is essential in wireless systems to continuously adapt different parameters of MIMO-OFDM systems, in particular the link parameters, to the variations in the communication environment. Various work (e.g. [52]) tackle this parameter selection problem using machine learning. Due to the dynamic nature of the

wireless communication environment, there is also a strong need for adapting hardware parameters, e.g. to select a suitable set of transmit and receive antennas [25].

The problem of reducing the peak-to-average power ratio (PAPR) is one of the key aspects in the design of OFDM-based wireless systems. This problem has attracted much attention for many years. Application examples of machine learning to the PAPR reduction problem include neural networks [23] and set-theoretic approaches [10] that are particularly suitable for online learning. Methods of machine learning and compressive sensing can also provide a key ingredient in enhancing the efficiency of OFDM channel estimation. For instance, the authors of [13] address the problem by considering a neural network with known pilot signals at its input and the corresponding channel response at its output. Other works (e.g. [39]) turn their attention towards the problem of channel estimation in MIMO systems in the presence of nonlinearities. Learning-based approaches have been also applied for the estimation of mmWave channels [32].

In order to enable an efficient and reliable opportunistic spectrum access, several approaches based on supervised, unsupervised, or reinforcement learning have been proposed in the literature. For instance, the study [44] considers a cognitive radio system with cooperative spectrum sensing where multiple secondary users cooperate to obtain robust spectrum-sensing results. Other approaches [8] apply distributed adaptive learning to tackle this problem.

Power control is a key mechanism for resource allocation in wireless systems. Machine learning has attracted some attention in the context of MIMO power control (e.g. [31]). Various learning-based approaches (e.g. [19]) have also been proposed to tackle the inter-cell interference problem, which may have a detrimental impact on the performance of wireless users in mobile networks. Furthermore, human supervision is still an indispensable element of current network management tools that are used to operate and manage mobile networks. Much research effort has been spent in the last decade to fully automate the network management process and with it to realize the vision of self-organizing networks that operate without human intervention (see [2]).

Information on the position of wireless devices is a key prerequisite for many applications. Machine learning methods have been used for localization [53] as well as navigation and positioning in car-to-car communication systems [40].

2.3. Security, privacy and communications

Machine learning methods play a pivotal role in tackling privacy and security-related problems in communications. For instance, they monitor various network activities and detect anomalies, i.e. events that deviate from the normal network behavior. Various machine learning methods have been applied for network anomaly detection in the past (see [45]). Other security applications are automatic spam filtering [21] and phishing attack detection [4].

Preserving data privacy is an important security aspect in communications, especially when sensitive data is involved. The design of machine learning algorithms that respect data privacy has recently gained increased attention. The authors of [1] demonstrated that it is possible to build a decision-tree classifier from corrupted data without significant loss in accuracy compared to the classifiers built with the original data, while at the same time it is not possible to accurately estimate the original values in the corrupted data records. This way one can hide private information from the algorithm, but still obtain accurate classification results.

2.4. Smart services, smart infrastructure and IoT

With the recent advances in communication technology the new field of “smart” applications has gained increased attention (e.g. smart homes, smart cities, smart grids, Internet of things). Machine learning algorithms are often the core part of such applications. For instance, the authors of [14] used a neural network based prediction algorithm to forecast and manage the power production of a photovoltaic plant. Other researchers have applied similar techniques to traffic light control [48] in smart cities or context aware computing in IoT [37]. Machine learning can also help detecting malicious events before they occur, e.g. in smart-grid networks [18]. Tasks such as the prediction of a resource usage, estimation of task response times, data traffic monitoring and optimal scheduling have also been tackled with learning algorithms [49].

2.5. Image and video communications

Machine learning methods have been used for various tasks in multimedia communication and processing (e.g. more than 200 applications of neural networks for images are summarized in [16]). Signal compression is one important field of application of these methods as it is part of almost every

multimedia communication system. A survey on image compression methods with neural networks can be found in [24]. Tracking is another well-studied topic in machine learning which is also relevant in multimedia communication. A new generation of object tracking methods based on deep neural networks has been described in [12]. Tracking algorithms which make use of the compressed video representation have also gained attention recently [41]. In multimedia applications such as video streaming the quality of the displayed video is of crucial importance. Different machine learning methods have been proposed to estimate the subjective quality of images perceived by a human [5, 6].

3. EXEMPLAR APPLICATIONS IN WIRELESS NETWORKING

The design and operation of wireless networks is a highly challenging task. On the road to the fifth generation of mobile networks (5G), researches and practitioners are, in particular, challenged by a multitude of conflicting requirements and promises as ever higher data-rates, lower-latency and lower energy consumption.

The main cause of the problems and limitations in the context of 5G is the radio propagation channel. This so-called wireless channel can strongly distort transmission signals in a manner that varies with frequency, time, space and other system parameters. The channel distortions are therefore of random nature and are notoriously difficult to estimate and predict. In addition, the wireless channel is a shared communication medium so that different wireless (communication) links must share the available communication resources. In modern mobile networks, this leads to interference between different mobile users, which in turn may have a detrimental impact on network operation. As a result, the capacity of wireless links is of an ephemeral and highly dynamic nature, and it depends on global channel parameters such as path loss, path delay and carrier phase shifts, all of which vary with time, frequency and space.

Against this background, it is not surprising that the problem of reconstructing, tracking and predicting channel parameters play a prominent role in the design and operation of modern wireless networks such as 5G. Traditional approaches to this problem are usually based on the assumptions that 1) the wireless channel can be modeled with a sufficient accuracy and 2) a sufficient number of pilot-based channel measurements can be performed in real time. However, the continuously increasing need for high-spectral efficiency and the utilization of extremely high frequencies (above 6 GHz) makes these assumptions untenable in future networks. A potential solution will not be an adaptation or extension within an existing framework, but rather a paradigm shift is necessary to meet the requirements of 5G networks. This in turn requires large strides both with respect to theoretical foundations and practical implementations.

Modern wireless networks collect and process a large amount of data and this data (including measurement data) can be used for tackling the previously mentioned problem of channel reconstruction, tracking and prediction. Therefore, in this context, special attention has been given to the development of new machine learning algorithms that are able to process spatially distributed data in real time while efficiently using scarce wireless communication resources. This calls for the development of distributed algorithms that in addition must provide robust results, have good tracking (online) capabilities, and exhibit a relatively low complexity. Finally, they need to exploit context and side information such as spatial and temporal sparsity in the wireless channel.

In the following subclause, we present one promising machine learning approach to the problem of reconstructing and tracking path loss maps in cellular networks. Clause 3.2 exemplifies the possibility of designing deep neural networks that exploit sparsity in the input data and which are amenable to real-time implementation.

3.1. Reconstruction of radio maps

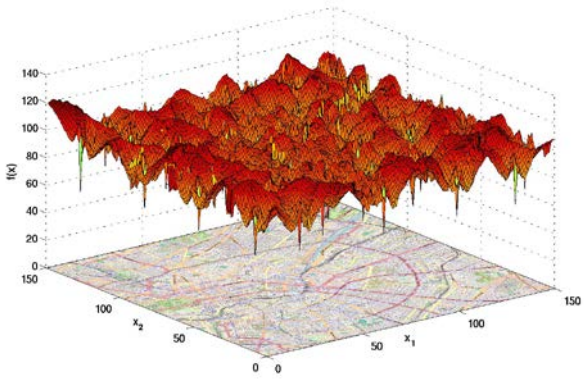

Fig. 2. An example of a path loss map for the downlink scenario with multiple base stations. The path loss map is a 2-dimensional function that assigns to a geographical position its path loss to the strongest base station.

We consider the downlink of a cellular network in which a number of base stations (transmitters) send the data to mobile users. While the users move over a geographical area covered by the network and send periodically their path loss measurements to the base stations, the problem is to reconstruct and track a (complete) path loss map in an online fashion from these measurements. The path loss determines the reduction in the power density of a transmission signal as it propagates from a base station to a geographical position. Note that for every geographical position, its path loss is defined with respect to the strongest base station, which is the base station with the smallest path loss. A radio map is then a function $f: \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ that assigns to every geographical position in a given area its path loss associated with the strongest base station. Figure 2 shows an example of a path loss map (the 2-dimensional function over the geographical area) for the downlink of a cellular network.

The general setting is as follows: Each base station collects path loss measurements sent by a subset of mobile users and it updates its estimate of the path loss map in an online manner whenever a new measurement arrives. Measurements may contain errors since geographical location cannot be determined with arbitrary precision and measured path loss values can be erroneous. Finally, measurements are not uniformly distributed over a given geographical area so that more measurements may be available for some sub-areas than for others. The challenge is to reliably reconstruct the path loss map, including the path loss values for geographical positions for which no path loss measurements are available.

The problem was considered in [26] where the authors propose using a multi-kernel approach based on adaptive projection methods. To be more precise, consider an arbitrary base station and let $(x_n, y_n) \in \mathbb{R}^2 \times \mathbb{R}$ be its measurement at time $n \in \mathbb{N}$, where $x_n \in \mathbb{R}^2$ is a sample (position measurement) at time n and $y_n \in \mathbb{R}_{\geq 0}$ is the corresponding response (a noisy path loss measurement). An estimate $\hat{f} : \mathbb{R}^2 \mapsto \mathbb{R}_{\geq 0}$ of the path loss map must be consistent with the available measurements. To this end, we require that $\forall_{n\in\mathbb{N}} |y_n - \hat{f}(x_n)| \leq \epsilon$ for some suitably chosen small $\epsilon > 0$. In [26], this requirement is met by projecting the estimate $\hat{f}$ on the hyperslabs given by $S_n = \{f \in \mathcal{H} : |y_n - \langle f, \kappa(x_n, \cdot)\rangle| \leq \epsilon\}, n \in \mathbb{N}$ where $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS) and $\kappa : \mathbb{R}^2 \times \mathbb{R}^2 \mapsto \mathbb{R}$ is the reproducing kernel for $\mathcal{H}$ so that $\langle f, \kappa(x_n, \cdot)\rangle = f(x_n)$ (the reproducing property). Due to a lack of space, we refer the reader to [26] for a rigorous definition of the concept of RKHS.

Since S_n is a closed convex set, the method of projection on convex sets (POCS) [20] provides the basis for the development of an iterative algorithm. However, the POCS framework cannot be directly applied to our problem at hand because the number of measurements grows without bound as time evolves. Therefore, the authors of [26] considered a different algorithmic approach that is a special case of the adaptive projected sub-gradient methods (APSM) developed in [50, 11, 9]. These methods open the door to distributed implementation and real-time online processing via adaptive parallel projections on closed convex sets such as the hyperslabs. Moreover, they allow for incorporating context information in a systematic manner, while exhibiting relatively low-complexity and robustness against errors. For more details the reader is referred to [50, 11, 9, 26].

The main disadvantage of the APSM-based approach is the need for choosing appropriate kernel functions. In fact, in practical scenarios, different geographical positions require different kernel functions that in addition need to be adapted over time due to the dynamic nature of the wireless environment. Since a real-time optimization of the kernel functions is an intricate task, inspired by the work [51], the authors of [26] developed a multi-kernel approach that adapts kernel functions over time and space by choosing them from a large set of predefined kernel functions, while maintaining low-complexity and real-time capabilities. In the following, we briefly explain this approach.

To this end, let $\{\kappa_m\}_{m=1}^{M}$ with $\kappa_m : \mathbb{R}^2 \times \mathbb{R}^2 \mapsto \mathbb{R}$ be a given set of some predefined kernel functions, where $M \gg 1$ is sufficiently large to include all relevant kernel functions.

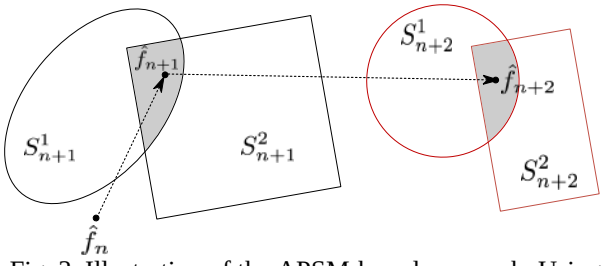

Fig. 3. Illustration of the APSM-based approach: Using parallel projection methods, the estimate $\hat{f}$n follows the intersections of the hyperslabs at times n+1 and n+2 to ensure consistency with new measurements and good tracking capabilities with online processing. At each time, there are two hyperslabs corresponding to two measurements.

Since the number of measurements grows linearly with n, we take into account only the most relevant data which is contained in the dictionary $\{(x_i,y_i)\}_{n\in In}$ where $I_n \subseteq \{n,n-1,...,1\}$ is the dictionary index set at time n. The cardinality $I_n = |I_n|$ of the dictionary must be sufficiently small to satisfy the hardware limitations on memory size and processor speed. With these definitions, for an arbitrary time n, the estimate $\hat{f}_n(x)$ of the path loss at position x can be written as a weighted sum of kernel functions: $\hat{f}_n(x) = \langle A_n, K_n \rangle = \mathrm{trace}(A^{\mathrm{T}}{}_n, K_n)$. Here $K_n = K_n(x) \in \mathbb{R}^{M\times In}$ is a given kernel matrix (evaluated at x) with $[K_n]_{i,m} = \kappa_m(x,x_i)$, and $A_n \in \mathbb{R}^{M\times In}$ is a parameter matrix that needs to be optimized. We point out that since the kernel matrix depends on the position $x \in \mathbb{R}^2$, the parameter matrix should be optimized for different geographical positions.

The most obvious requirement on the parameter matrix A is that it must be chosen to fit the estimate to the measurements. This can be achieved by minimizing the distance (with some suitably chosen metric) of A from the set $S_n = \{A \in \mathbb{R}^{M\times In} : |\langle A_n, K_n \rangle - y_n| \leq \epsilon\}$ for some sufficiently small $\epsilon > 0$. Since M is large, the problem is however computationally prohibitive for many applications in wireless networks. Therefore, the authors of [26] extended the objective function by adding to the distance metric two regularization terms that impose some sparsity in A when the new regularized objective function is minimized. As a result, the approach not only fits the estimate function to the measurements but also discards irrelevant data in the dictionary and reduces the impact of unsuitable kernels.

The regularized objective function provides a basis for the development of new iterative algorithms in [26] based on the forward-backward splitting methods and sparsity-based iterative weighting methods. The algorithms provide good tracking capabilities for the problem of reconstructing and tracking time-varying path loss maps. For more details, we refer the reader to [26].

3.2. Deep neural networks for sparse recovery

Recently, compressed sensing and deep learning have emerged as theoretical and practical toolsets to unleash full potential and approach fundamental theoretical bounds, whether it be for pilot decontamination in channel estimation, user identification, activity detection or PAPR reduction. While in many cases researchers are well aware of optimal solutions, e.g. in terms of optimization problems for channel estimation using a minimal number of pilots, implementing these solutions in embedded devices is considered unfeasible due to unpredictable termination times and incalculable loss of early stopping. In this regard, a provisional solution aimed at large-scale measurement campaigns and utilizing black-box data-driven machine learning techniques. While this approach fits well with many imaging problems, it was soon stripped of its enchantment for communication systems due to the necessity of measuring and preprocessing RF signals under diverse sets of environmental conditions resulting in extremely large training times and disappointing performance gains. In addition, there is still no commonly accepted neural network de-facto standard or baseline architecture for particular communication problems akin to AlexNet or GoogleNet in the imaging domains. One step to close this important gap was made in [29] by using multidimensional Laplace transform techniques to design optimal neural networks for particular sparse recovery problems revealing a very intriguing connection between commonly employed neural networks comprising weights, threshold functions, rectified linear (ReLU) and rectified polynomial (ReP) activation functions and volume and centroid computation problems over sparsity inducing sets. We refer the reader to Figure 4 for a geometric illustration of a small sparse recovery problem. Here, x is to be recovered from dimensionality reduced measurement y = Ax given that x belongs to a particular sparsity inducing set (blue). Then, the neural network of [29] outputs the estimate $\hat{x}$ that minimizes the expected error over the uncertainty set (intersection between red and blue). Using such geometric ideas in the design of neural networks allows for bypassing costly search over exponential candidate networks that consume large portions of available computing resources. Indeed, practitioners can still apply fine-tuning to reduce a possible model mismatch and reduce reconstruction errors even further.

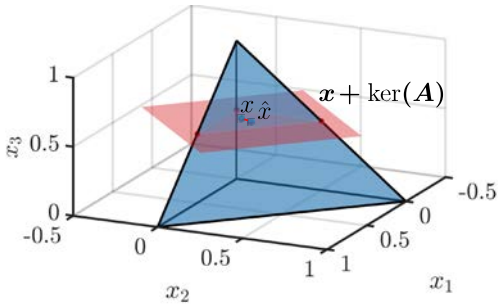

Fig. 4. Geometry of sparse recovery.

4. FUTURE RESEARCH TOPICS

This clause discusses four future research topics in machine learning and communications.

4.1. Low complexity models

State-of-the-art machine learning models such as deep neural networks are known to work excellently in practice. However, since the training and execution of these models require extensive computational resources, they may not be applicable in communication systems with limited storage capabilities, computational power and energy resources, e.g. smartphones, embedded systems or IoT devices. Recent work addressed this problem and proposed techniques for reducing the complexity of deep neural networks. For instance, the authors of [22] demonstrated that the size of VGG-16, a popular deep neural network for image classification, can be reduced by over 95% with no loss of accuracy. In communication applications such compression techniques can be used to store and transmit models efficiently. Other authors (e.g. [15]) targeted the problem of weight binarization in deep neural networks. This type of discretization can be useful, e.g. when adapting models to processor architectures which do not allow floating point operations.
Further research on these topics is of high importance as it can be expected that a large number of new applications would emerge, if the complexity of state-of-the-art models can be reduced to a level, which allows their use in computationally limited environments at minimal performance loss.

4.2. Standardized formats for machine learning

The standardization of algorithms and data formats is of high importance in communications, because it increases the reliability, interoperability and modularity of a system and its respective components. With the increasing use of learning algorithms in communication applications, the need for standardized formats for machine learning is also rising.
For instance, standardized formats could be used to specify how to train, adapt, compress and exchange machine learning models in communication applications. Furthermore, there could be standardized formats for the data and standards which determine how multiple machine learning models interact with each other. Other formats could be specifically designed for ensuring that a model fulfills certain security or privacy requirements.

4.3. Security and privacy mechanisms

Machine learning models are often used in a black-box manner in today's applications. This prevents the human expert from comprehending the reasoning of the algorithm and from validating its predictions. Although recent works [3, 33] proposed techniques for explaining the predictions of a machine learning model, further research on this topic is of high importance as the lack of transparency can be a large disadvantage in communication applications.
Moreover, it is well-known that deep neural networks can be easily fooled or may behave in an unexpected way when being confronted with data with different properties to the data used for training the model [43]. Thus, the establishment of mechanisms which increase the reliability of the model is a prerequisite for large-scale use in communication applications. Such mechanisms can be implemented on different levels, e.g. be an integral part of the model, be integrated into the communication protocol or be part of a separate inspection process.
Besides interpretability and security aspects, future research also needs to investigate how to effectively encrypt machine learning models and how to ensure data privacy during and after learning.

4.4. Radio resource and network management

The end-to-end performance of mobile networks is strongly influenced by the choice of radio resource (e.g. beamforming and medium access control parameters) and network management (e.g. handover parameters, neighborhood lists, loads and power budgets) parameters. Moreover, some of the parameters must be continuously adapted on a relatively short time scale to time-varying radio propagation conditions and changing network topologies [42].
Current approaches are inadequate to cope with the growth of autonomous network elements in 5G small cell deployments based on mobile cloud RAN architectures. Therefore, 5G networks call for new model and data-driven radio resource management and network management methods that are augmented by machine learning techniques for extracting knowledge from the system and gradual learning in the presence of inherent uncertainties and the lack of complete channel and network state information [38]. The realization of these ideas in the context of 5G will require modifications of existing protocols and the development of new ones.

5. CONCLUSION

This paper discussed the increasing mutual influence of machine learning and communication technology. Learning algorithms were not only shown to excel in traditional network management tasks such as routing, channel estimation or PAPR reduction, but also to be a core part of many emerging application fields of communication technology, e.g. smart cities or Internet of things. The availability of large amounts of data and recent improvements in deep learning methodology will further foster the convergence of these two fields and will offer new ways to optimize the whole communication pipeline in an end-to-end manner [36].

However, before resource-intensive models such as deep neural networks can be applied on a large scale in communication applications, several practical challenges (e.g. complexity, security, privacy) need to be solved. Furthermore, more research is required on theoretical topics at the intersection of communications and machine learning, e.g. incremental learning, learning in non-stationary environments or learning with side information.

REFERENCES

[1] R. Agrawal and R. Srikant. Privacy-preserving data mining. In ACM SIGMOD Manag. Data, pages 439– 450, 2000.

[2] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans. A survey of self organisation in future cellular networks. IEEE Commun. Surveys Tuts., 15(1):336–361, 2013.

[3] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLOS ONE, 10(7):e0130140, 2015.

[4] R. B. Basnet, S. Mukkamala, and A. H. Sung. Detection of phishing attacks: A machine learning approach. Soft Comp. Appl. Indust., 226:373–383, 2008.

[5] S. Bosse, D. Maniry, T. Wiegand, and W. Samek. A deep neural network for image quality assessment. In IEEE ICIP, pages 3773–3777, 2016.

[6] S. Bosse, K.-R. Müller, T. Wiegand, and W. Samek. Brain-computer interfacing for multimedia quality assessment. In IEEE Syst. Man Cyber., pages 002834– 002839, 2016.

[7] J. A. Boyan and M. L. Littman. Packet routing in dynamically changing networks: A reinforcement learning approach. NIPS, pages 671–671, 1994.

[8] R. L. Cavalcante and S. Stanczak. A distributed subgradient method for dynamic convex optimization problems under noisy information exchange. IEEE J. Sel. Topics Signal Process., 7(2):243–256, 2013.

[9] R. L. G. Cavalcante and S. Stanczak. Robust settheoretic distributed detection in diffusion networks. In IEEE ICASSP, 2012.

[10] R. L. G. Cavalcante and I. Yamada. A flexible peak-toaverage power ratio reduction scheme for ofdm systems by the adaptive projected subgradient method. IEEE Trans. Signal Process., 57(4):1456–1468, 2009.

[11] R. L. G. Cavalcante, I. Yamada, and B. Mulgrew. An adaptive projected subgradient approach to learning in diffusion networks. IEEE Trans. Signal Process., 57(7):2762–2774, 2009.

[12] Y. Chen, X. Yang, B. Zhong, S. Pan, D. Chen, and H. Zhang. Cnntracker: online discriminative object tracking via deep convolutional neural network. Appl. Soft Comput., 38:1088–1098, 2016.

[13] C.-H. Cheng, Y.-H. Huang, and H.-C. Chen. Channel estimation in ofdm systems using neural network technology combined with a genetic algorithm. Soft Comput., 20(10):4139–4148, 2016.

[14] L. Ciabattoni, G. Ippoliti, A. Benini, S. Longhi, and M. Pirro. Design of a home energy management system by online neural networks. IFAC Proc. Volumes, 46(11):677–682, 2013.

[15] M. Courbariaux, I. Hubara, D. Soudry, R. El-Yaniv, and Y. Bengio. Binarized neural networks: Training deep neural networks with weights and activations constrained to +1 or -1. arXiv:1602.02830, 2016.

[16] M. Egmont-Petersen, D. de Ridder, and H. Handels. Image processing with neural networksa review. Pattern Recognit., 35(10):2279–2301, 2002.

[17] A. Eswaradass, X.-H. Sun, and M. Wu. A neural network based predictive mechanism for available bandwidth. In IPDPS, pages 33a–33a, 2005.

[18] Z. M. Fadlullah, M. M. Fouda, N. Kato, X. Shen, and Y. Nozaki. An early warning system against malicious activities for smart grid communications. IEEE Netw., 25(5), 2011.

[19] A. Galindo-Serrano and L. Giupponi. Distributed qlearning for aggregated interference control in cognitive radio networks. IEEE Trans. Veh. Technol., 59(4):1823–1834, 2010.

[20] L. Gubin, B. Polyak, and E. Raik. The method of projections for finding the common point of convex sets. USSR Comput. Math. Math. Phys., 7(6):1–24, 1967.

[21] T. S. Guzella and W. M. Caminhas. A review of machine learning approaches to spam filtering. Expert Syst. Appl., 36(7):10206–10222, 2009.

[22] S. Han, H. Mao, and W. J. Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. arXiv:1510.00149, 2015.

[23] Y. Jabrane, V. P. G. Jimenez, A. G. Armada, B. A. E.´ Said, and A. A. Ouahman. Reduction of power envelope fluctuations in ofdm signals by using neural networks. IEEE Commun. Lett., 14(7):599–601, 2010.

[24] J. Jiang. Image compression with neural networks–a survey. Signal Process. Image, 14(9):737–760, 1999.

[25] J. Joung. Machine learning-based antenna selection in wireless communications. IEEE Commun. Lett., 20(11):2241–2244, 2016.

[26] M. Kasparick, R. L. G. Cavalcante, S. Valentin, S. Stanczak, and M. Yukawa. Kernel-based adaptive online reconstruction of coverage maps with side information. IEEE Trans. Veh. Technol., 65(7):5461–5473, 2016.

[27] S. Kumar and R. Miikkulainen. Dual reinforcement q-routing: An on-line adaptive routing algorithm. In ICANNE, 1997.

[28] Y. Liang. Real-time vbr video traffic prediction for dynamic bandwidth allocation. IEEE Trans. Syst., Man, Cybern. C, 34(1):32–47, 2004.

[29] S. Limmer and S. Stanczak. Optimal deep neural networks for sparse recovery via laplace techniques. arXiv, 2017.

[30] T. Lu and J. Zhu. Genetic algorithm for energy-efficient qos multicast routing. IEEE Commun. Lett., 17(1):31– 34, 2013.

[31] P. Mertikopoulos and A. L. Moustakas. Learning in an uncertain world: Mimo covariance matrix optimization with imperfect feedback. IEEE Trans. Signal Proc., 64(1):5–18, 2016.

[32] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath. Channel estimation in millimeter wave mimo systems with one-bit quantization. In Asilomar Signals Syst. Comput., pages 957–961, 2014.

[33] G. Montavon, W. Samek, and K.-R. Müller. Methods for interpreting and understanding deep neural networks. arXiv:1706.07979, 2017.

[34] M. Munetomo, Y. Takai, and Y. Sato. A migration scheme for the genetic adaptive routing algorithm. In IEEE Sys. Man Cyber., volume 3, pages 2774–2779, 1998.

[35] T. T. Nguyen and G. Armitage. A survey of techniques for internet traffic classification using machine learning. IEEE Commun. Surveys Tuts., 10(4):56–76, 2008.

[36] T. J. O'Shea, T. Erpek, and T. C. Clancy. Deep learning based mimo communications. arXiv:1707.07980, 2017.

[37] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos. Context aware computing for the internet of things: A survey. IEEE Commun. Surveys Tuts., 16(1):414–454, 2014.

[38] T. W. Rondeau. Application of Artificial Intelligence to Wireless Communications. PhD thesis, Virginia Tech, 2007.

[39] M. Sanchez-Fernandez, M. de Prado-Cumplido, J. Arenas-Garc´ıa, and F. Perez-Cruz. Svm multiregres-´ sion for nonlinear channel estimation in multiple-input multiple-output systems. IEEE Trans. Signal Process., 52(8):2298–2307, 2004.

[40] Skog and P. Handel. In-car positioning and navigation technologiesa survey. IEEE Trans. Intell. Transp. Syst., 10(1):4–21, 2009.

[41] V. Srinivasan, S. Lapuschkin, C. Hellge, K.-R. Müller, and W. Samek. Interpretable human action recognition in compressed domain. In IEEE ICASSP, pages 1692– 1696, 2017.

[42] S. Stanczak, M. Wiczanowski, and H. Boche. Fundamentals of resource allocation in wireless networks: theory and algorithms, volume 3. Springer, 2009.

[43] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. arXiv:1312.6199, 2013.

[44] M. Thilina, K. W. Choi, N. Saquib, and E. Hossain. Machine learning techniques for cooperative spectrum sensing in cognitive radio networks. IEEE J. Sel. Areas Commun., 31(11):2209–2221, 2013.

[45] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin. Intrusion detection by machine learning: A review. Expert Syst. Appl., 36(10):11994–12000, 2009.

[46] C.-W. Tsai, C.-F. Lai, M.-C. Chiang, L. T. Yang, et al. Data mining for internet of things: A survey. IEEE Commun. Surveys Tuts., 16(1):77–97, 2014.

[47] P. J. Werbos. Computational intelligence for the smart grid-history, challenges, and opportunities. IEEE Comput. Intell. Mag., 6(3):14–21, 2011.

[48] M. Wiering. Multi-agent reinforcement learning for traffic light control. In ICML, pages 1151–1158, 2000.

[49] Xu, M. Zhao, J. Fortes, R. Carpenter, and M. Yousif. Autonomic resource management in virtualized data centers using fuzzy logic-based approaches. Cluster Comput., 11(3):213–227, 2008.

[50] Yamada and N. Ogura. Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions. Num. Func. Analys. Optim., 25(7/8):593–617, 2005.

[51] Yukawa. Multikernel adaptive filtering. IEEE Trans. Signal Process., 60(9):4672–4682, 2012.

[52] S. Yun and C. Caramanis. Reinforcement learning for link adaptation in mimo-ofdm wireless systems. In GLOBECOM, pages 1–5, 2010.

[53] Zhang, Y. Zhuo, and Y. Zhao. Mobile location based on svm in mimo communication systems. In ICINA, volume 2, pages V2–360, 2010.

[54] Q. Zhang and Y.-W. Leung. An orthogonal genetic algorithm for multimedia multicast routing. IEEE Trans. Evol. Comput., 3(1):53–62, 1999.

[55] Y. Zhang, S. Kwong, X. Wang, H. Yuan, Z. Pan, and L. Xu. Machine learning-based coding unit depth decisions for flexible complexity allocation in high efficiency video coding. IEEE Trans. Image Process., 24(7):2225–2238, 2015.

