Advances in Machine Learning and Deep Neural Networks

By RAMA CHELLAPPA, *Life Fellow IEEE*
Guest Editor

SERGIOS THEODORIDIS, *Life Fellow IEEE*
Guest Editor

ANDRE VAN SCHAIK, *Fellow IEEE*
Guest Editor

We are currently experiencing the dawn of what is known as the fourth industrial revolution. At the center of this historical happening, as one of the key enabling technologies, lies a discipline that deals with data and whose goal is to extract information and related knowledge that is hidden in it, in order to make predictions and, subsequently, take decisions. Machine learning (ML) is the name that is used as an umbrella to cover a wide range of theories, methods, algorithms, and architectures that are used to this end.

This special issue covers promising developments in the related areas of machine learning and deep neural networks and offers possible paths for the future.

ML as a subdiscipline of what is known as artificial intelligence (AI) is the field that has been consolidated as an entity of its own sometime around 1986, following the publication of the seminal paper [1]. However, it is around 2010, with the "rediscovery" of deep neural networks combined with the availability of large data sets and affordable computer power, that it is really flourishing and it has now been established as a field that dominates the area of AI and it is diffused to a wide range of practical applications. It is difficult to think of an area where ML has not been applied. Deep neural networks (NNs) have offered performances much beyond of the incremental nature compared to previous techniques.

However, after ten years of success and intense research, it seems that some saturation has been reached. A number of people raise the question, what comes next? What is the next leap forward?

In view of the previous discussion, a number of remaining drawbacks, yet unsolved after ten years of intense research, come into the main scene of discussion. Deep networks usually need huge amounts of data for their training. Also, the energy consumption to execute the algorithms in standard digital computers is huge. Training such networks for real-life applications is a formidable task that needs the use of cloud computing and powerful GPUs. Hence, if the next demand will be to have such algorithms running in, e.g., our phones, in autonomous cars, and in our everyday appliances and sensors, which comprise the communications backbone of what is known as Industry 4, then the field needs to overcome a number of major challenges.

On the algorithmic front, over the recent years, a line of research focusses on the so-called federated learning and edge computing that explores collaborative/distributed processing where data are distributed among a number of agents, as opposed to the more classical centralized processing.

Digital Object Identifier 10.1109/JPROC.2021.3072172

Such processing facilitates real-time computing that is needed in a number of applications, such as in communication networks. On the hardware front, a major research effort has been dedicated in developing power-efficient architectures based on neuromorphic computing concepts and brain-inspired algorithms. Besides computational issues, current models have revealed several new concerns that are related to the security and interpretability of the obtained results, which are of paramount importance in certain disciplines, such as in financial and medical applications. The existence of adversarial examples poses a potential threat since one can fool any system. Furthermore, current ML and deep network algorithms can only make predictions. This is far away from the original goals set by the pioneers of AI, which was to build systems that "reason." One of the major related aspects is that of causality, which comprises a significant characteristic of what we call human intelligence. One should be able to "understand" and interpret the obtained results. Can we build machines that deal with the notion of causality besides simple prediction? Furthermore, several theoretical issues have emerged, while training such networks. According to classical ML theory, as we know from any textbook, overparametrizing a model will lead to overfitting to the training data specificities. Yet, while training deep NNs, often with many more parameters compared to the available size of the training data, experiments have revealed that very good generalization properties can still be obtained.

It is true to say that most of the research effort over the last ten years or so has been invested in deep neural networks, which, no doubt, changed the way that we view ML. Yet, is it the end? Is there something beyond deep neural networks, that could provide better solutions, or may be used as a complementary method? Is it about the time for the ML community and funding bodies to look for new inspiring ideas?

I. OVERVIEW OF THE SPECIAL ISSUE

The scope of this issue is to put together a number of papers, written by world experts in the field, that try to provide some answers, to the possible extent, to the previously discussed problems, and at the same time to summarize research that has been carried in the related areas. The topics covered range from theory and algorithms to applications and hardware implementations. We made an effort to select key areas of high current research interest in all three directions. The special issue comprises 14 invited papers. On the "theory" front, the selected topics refer to causal inference in the context of deep networks, adversarial learning, graph deep networks, a spline approximation view of deep networks, analysis of overparameterized networks, and the adoption of tropical geometry as a tool for ML. All these papers deal with some of the open questions and major challenges associated with ML and deep networks and try to offer possible paths for the future. On the applications' front, some "typical" topics have been selected and the related papers provide a review of more recent related results. Specifically, anomaly detection, medical imaging, computer vision, generative adversarial networks for image and video synthesis, computational media intelligence/multimodal ML, and wireless communications are considered. Finally, on the hardware front, two papers are dedicated to brain-inspired algorithms and neuromorphic computing.

A. Theory and Algorithms

Toward Causal Representation Learning

by B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio

This article reviews fundamental concepts of causal inference and relates them to crucial open problems of machine learning, including transfer learning and generalization, thereby assaying how causality can contribute to modern machine learning research. This also applies in the opposite direction: most work in causality starts from the premise that the causal variables are given. A central problem for AI and causality is, thus, causal representation learning, the discovery of high-level causal variables from low-level observations. Some implications of causality for machine learning are discussed and some key research areas at the intersection of the two disciplines of ML and graphical causality are proposed.

Optimism in the Face of Adversity: Understanding and Improving Deep Learning Through Adversarial Robustness

by G. Ortiz-Jiménez, A. Modas, S.-M. Moosavi-Dezfooli, and P. Frossard

This article presents an in-depth review of the field of adversarial robustness in deep learning and provides a self-contained introduction to its main notions. In contrast to the mainstream pessimistic perspective of adversarial robustness, the paper reveals some of the main positive aspects that it entails. The authors highlight the intuitive connection between adversarial examples and the geometry of deep neural networks and eventually explore how the geometric study of adversarial examples can serve as a powerful tool to better understand deep learning. Furthermore, the authors demonstrate the broad applicability of adversarial robustness, by providing an overview of the main emerging applications of adversarial robustness beyond security. A major strength of the article lies in that it presents a new perspective to understand deep learning and to supply readers with intuitive tools and insights on how to use adversarial robustness to improve it.

Graph Neural Networks: Architectures, Stability, and Transferability

by L. Ruiz, F. Gama, and A. Ribeiro

This article deals with graph neural networks (GNNs) that operate on data supported by graphs. In the current context, such networks are treated as generalizations of convolutional

neural networks (CNNs), in which individual layers contain banks of graph convolutional filters, as opposed to the more classical notion of convolutional filters. Filters employ pointwise nonlinearities and are stacked in layers. It is shown that GNN architectures exhibit equivariance to permutation and stability to graph deformations. These properties help explain the good performance of GNNs that can be observed empirically. The underlying concepts are illustrated by the application of GNNs to recommendation systems, decentralized collaborative control, and wireless communication networks.

Mathematical Models of Overparameterized Neural Networks
by C. Fang, H. Dong, and T. Zhang

The focus of this article is on theoretical developments concerning the analysis of overparameterized neural networks. In a number of fairly recent articles, it has been shown that such systems behave like convex systems under various restricted settings, as, for example, in the case for two-layer NNs, and when learning is restricted locally in the so-called neural tangent kernel space around specialized initializations. The article discusses some of these recent developments that lead to a significant better understanding of neural networks. The emphasis of the reported results lies on the analysis of two-layer neural networks and explains the key mathematical models, with their algorithmic implications. Finally, challenges in understanding deep neural networks and some current research directions are discussed.

Mad Max: Affine Spline Insights Into Deep Learning
by R. Balestriero and R. G. Baraniuk

In this article, the bridge between deep networks (DNs) and approximation theory via spline functions and operators is rigorously established. It is shown that a DN constructs a set of signal-dependent, class-specific templates against which the signal is compared via a simple inner product; the links to the classical theory of optimal classification via matched filters and the effects of data memorization are established. Based on this, the authors propose the use of a simple penalty term that can be added to the cost function of any DN learning algorithm to force the templates to be orthogonal with each other; this leads to significantly improved classification performance and reduced overfitting with no change to the DN architecture. This point of view establishes links of DNs to the theory of vector quantization (VQ) and K-means clustering, which opens up a new geometric avenue to study how DNs organize signals in a hierarchical fashion. To validate the utility of the VQ interpretation, a new distance metric for signals and images is developed that quantifies the difference between their VQ encodings.

Tropical Geometry and Machine Learning
by P. Maragos, V. Charisopoulos, and E. Theodosis

This article is an example of how tools that have been developed in different disciplines can be combined and used in the context of machine learning tasks. Tropical geometry is a relatively recent field in mathematics and computer science combining elements of algebraic and polyhedral geometry. Tropical geometry recently emerged in the analysis and extension of several classes of problems and systems in both classical machine learning and deep learning. The article first summarizes some introductory ideas and objects of tropical geometry, providing a theoretical framework for both the max-plus algebra that underlies tropical geometry as well as its extensions to general max algebras. Then, the authors survey the state-of-the-art and recent progress in three areas. First, a purely geometric approach for studying the representation power of neural networks with piecewise-linear activations is established. Then, the tropical geometric analysis of parametric statistical models, such as HMMs is considered. Finally, the authors derive optimal solutions and an efficient algorithm for the convex regression problem, using concepts and tools from tropical geometry and max-plus algebra.

B. Applications

A Unifying Review of Deep and Shallow Anomaly Detection
by L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, and K.-R. Müller

Anomaly detection is a major problem in machine learning for decades and a number of approaches have been proposed over the years. More recently, deep learning techniques for anomaly detection have improved the state of the art in detection performance on complex datasets such as large collections of images or text. In this article, the authors aim to identify the common underlying principles as well as the assumptions that are often made implicitly by various methods. In particular, connections between classic "shallow" and novel deep approaches are established and it is shown how this relationship might cross-fertilize or extend both directions. An empirical assessment of major existing methods that is enriched by the use of recent explainability techniques is presented together with specific worked-through examples with practical advice. Finally, critical open challenges are discussed that identify paths for future research in anomaly detection.

Communication-Efficient and Distributed Learning Over Wireless Networks: Principles and Applications
by J. Park, S. Samarakoon, A. Elgabli, J. Kim, M. Bennis, S.-L. Kim, and M. Debbah

The goal of this article is to provide a holistic overview of relevant communication and ML principles, and thereby present communication-efficient and distributed ML frameworks with selected use cases. Machine learning is a promising enabler for the fifth-generation (5G) communication systems and beyond.

By embedding intelligence into the network edge, edge nodes can proactively carry out decision-making, and thereby react to local environmental changes and disturbances while experiencing zero communication latency. To this end, it is essential to cater for high ML inference accuracy at scale under the time-varying channel and network dynamics, by continuously exchanging fresh data and ML model updates in a distributed way. Taming this new kind of data traffic boils down to improving the communication efficiency of distributed ML by optimizing communication payload types, transmission techniques, scheduling, as well as ML architectures, algorithms, and data processing methods.

A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises

by S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers

Medical imaging has been one of the most prominent applications of machine learning. Deep learning has been widely used in various medical imaging tasks and has achieved remarkable success. Yet, it is known that the success of ML relies largely on the availability of big data with annotations for a single task. However, medical imaging presents unique challenges that often confront such requirements. In this article, the authors first highlight both clinical needs and technical challenges in medical imaging and describe how emerging trends in deep learning are addressing these issues. Topics such as network architecture, sparse and noisy labels, federating learning, interpretability, and uncertainty quantification are discussed. A number of case studies are presented including digital pathology and chest, brain, cardiovascular, and abdominal imaging. The article concludes with a discussion and presentation of promising future directions.

Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications

by M.-Y. Liu, X. Huang, J. Yu, T.-C. Wang, and A. Mallya

Generative adversarial networks (GANs) are one of the major breakthroughs in deep learning and in machine learning in general and have widely been used in various contexts and applications. This article provides an overview of GANs with a special focus on algorithms and applications for visual synthesis. Several important techniques are presented for stabilizing GAN training, which has a reputation for being notoriously difficult. A number of applications are discussed such as image translation, image processing, video synthesis, and neural rendering.

Tensor Methods in Computer Vision and Deep Learning

by Y. Panagakis, J. Kossaifi, G. G. Chrysos, J. Oldfield, M. A. Nicolaou, A. Anandkumar, and S. Zafeiriou

Tensors, or multidimensional arrays, are data structures that can naturally represent visual data of multiple dimensions. This article provides an overview of tensors and tensor methods in the context of representation learning and deep learning, with a particular focus on visual data analysis and computer vision applications. The article starts with presenting some basic definitions of tensors and it proceeds with an emphasis on tensor-based visual data analysis methods. Recent developments on the use of tensors in deep learning architectures are reported, and their implications in computer vision applications are highlighted. To further enable the newcomer to grasp such concepts quickly, the authors provide thorough tutorial-style companion notebooks in Python, covering every section of the article.

Computational Media Intelligence: Human-Centered Machine Analysis of Media

by K. Somandepalli, T. Guha, V. R. Martinez, N. Kumar, H. Adam, and S. Narayanan

The topic treated in this article is the application of deep learning algorithms, combined with audio-visual signal processing, to analyze entertainment media such as film/TV. Text mining and natural language processing allow a nuanced understanding of language use and spoken interactions in media such as News to track patterns and trends across different contexts. Moreover, advances in human sensing offer ways to directly measure the influence of media on an individual's physiology (and brain), while social media analysis enables tracking the societal impact of media content on different cross-sections of the society. The emphasis of the article is on the representative methodologies and algorithms, tools, and systems advancing the area of human-centered media understanding through machine learning in the pursuit of developing computational media intelligence.

C. Implementations

Advancing Neuromorphic Computing With Loihi: A Survey of Results and Outlook

by M. Davies, A. Wild, G. Orchard, Y. Sandamirskaya, G. A. Fonseca Guerra, P. Joshi, P. Plank, and S. R. Risbud

Neuromorphic computing aims to implement learning algorithms on chips that are more directly inspired by the form and function of biological neural circuits so they can process new knowledge, adapt, behave, and learn in real time at low power levels. At the heart of this article lies Intel's Loihi, a neuromorphic research processor designed to support a broad range of spiking neural networks with sufficient scale, performance, and features to deliver competitive results compared to state-of-the-art contemporary computing architectures. The article provides a survey of results obtained to date with Loihi across the major algorithmic domains under study, including deep learning approaches as well as novel approaches that aim

to more directly harness the key features of spike-based neuromorphic hardware. It is demonstrated that brain-inspired networks using recurrence, precise spike-timing relationships, synaptic plasticity, stochasticity, and sparsity perform certain computations with orders of magnitude lower latency and energy compared to state-of-the-art conventional approaches.

Brain-Inspired Learning on Neuromorphic Substrates
by F. Zenke and E. O. Neftci

This article provides a mathematical framework for the design of practical online learning algorithms for neuromorphic substrates. A direct connection between real-time recurrent learning (RTRL), an online algorithm for computing gradients in conventional recurrent neural networks (RNNs), and biologically plausible learning rules for training spiking neural networks (SNNs) is established. This framework bridges the gap between synaptic plasticity and gradient-based approaches from deep learning and lays the foundations for powerful information processing on future neuromorphic hardware systems. ■

REFERENCE

[1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by backpropagation errors," *Nature*, vol. 323, pp. 533–536, 1986.

ABOUT THE GUEST EDITORS

Rama Chellappa (Life Fellow, IEEE) is a Bloomberg Distinguished Professor in the Departments of Electrical and Computer Engineering and Biomedical Engineering (School of Medicine) at Johns Hopkins University. His current research interests span many areas in image processing, computer vision, machine learning and pattern recognition.

Prof. Chellappa has received several awards from IEEE, the International Association of Pattern Recognition, the University of Southern California and the University of Maryland. He has been recognized as an Outstanding ECE by Purdue University and received the Distinguished Alumni Award from the Indian Institute of Science. He served as the Editor-in-Chief of PAMI, as a Distinguished Lecturer of the IEEE Signal Processing Society and as the President of IEEE Biometrics Council. He is a Golden Core Member of the IEEE Computer Society. He is a Fellow of IEEE, IAPR, OSA, AAAS, ACM, AAAI, and NAI and holds eight patents.

Sergios Theodoridis (Life Fellow, IEEE) is currently a Professor Emeritus with the National and Kapodistrian University of Athens, Athens, Greece, and a Distinguished Professor with Aalborg University, Aalborg, Denmark. He is the author of the book *Machine Learning: A Bayesian and Optimization Perspective* (Academic Press, 2nd Edition, 2020), the coauthor of the best-selling book *Pattern Recognition* (Academic Press, 4th Edition, 2009), the coauthor of the book *Introduction to Pattern Recognition: A MATLAB Approach* (Academic Press, 2010), and the coeditor of the book *Efficient Algorithms for Signal Processing and System Identification* (Prentice Hall, 1993). His research interests span a wide range of areas in the intersection of signal processing and machine learning.

Prof. Theodoridis is a Fellow of the IET and EURASIP and a Corresponding Fellow of the Royal Society of Edinburgh (RSE). He has received a number of best paper awards, including the 2014 *IEEE Signal Processing Magazine* Best Paper Award and the 2009 IEEE Computational Intelligence Society IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award. He was a recipient of the 2017 EURASIP Athanasios Papoulis Award, the 2014 IEEE Signal Processing Society Education Award, and the 2014 EURASIP Meritorious Service Award. He has served as the Vice President for the IEEE Signal Processing Society, as the President for the European Association for Signal Processing (EURASIP), and as the Editor-in-Chief for IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Andre van Schaik (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the University of Twente, Enschede, The Netherlands, in 1990, and the Ph.D. degree in neuromorphic engineering from the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland, in 1998.

He is a pioneer in neuromorphic engineering and currently the Director of the International Centre for Neuromorphic Systems, Western Sydney University, Penrith, NSW, Australia. From 1991 to 1994, he was a Researcher with the Swiss Centre for Electronics and Microtechnology (CSEM), where he developed the first commercial neuromorphic chip—the optical motion detector used in Logitech trackballs since 1994. In 1998, he was a Postdoctoral Research Fellow with the Department of Physiology, The University of Sydney, Sydney, NSW, Australia, where he became a Senior Lecturer and a Reader at the School of Electrical and Information Engineering in 1999 and 2004, respectively. In 2011, he became a Full Professor at Western Sydney University. He has authored more than 200 articles and is an inventor of more than 35 patents. He has also founded three technology startups. His research focuses on all aspects of neuromorphic engineering, encompassing neurophysiology, computational neuroscience, software and algorithm development, and electronic hardware design.

Dr. van Schaik is a Fellow of IEEE for his contributions to neuromorphic circuits and systems.