Yonina C. Eldar, Alfred O. Hero III, Li Deng, Jeff Fessler, Jelena Kovačević, H. Vincent Poor, and Steve Young

Challenges and Open Problems in Signal Processing: Panel Discussion Summary from ICASSP 2017

This column summarizes the panel on open problems in signal processing, which took place on 5 March 2017 at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in New Orleans, Louisiana. The goal of the panel was to draw attention to some of the challenges and open problems in various areas of signal processing and generate discussion on future research areas that can be of major significance and impact in signal processing. Five leading experts representing diverse areas within signal processing made up the panel:

- Li Deng, Microsoft: machine learning
- Jeff Fessler, the University of Michigan: medical imaging
- Jelena Kovačević, Carnegie Mellon University: graph signal processing
- H. Vincent Poor, Princeton University: wireless communication
- Steve Young, the University of Cambridge: speech and language processing.

It was organized and moderated by Yonina Eldar from the Technion and Alfred O. Hero III from the University of Michigan.

The panel drew a very large crowd and stimulated a vibrant discussion on directions, trends, and challenges of signal processing in the 21st century and in the era of big data. In this column, we summarize the main points raised by the panelists and the audience in each of the aforementioned topics. Our goal and hope is to further the discussion on some of the main challenges and opportunities for signal processing in the coming years and to highlight areas where, as a community, working and collaborating together, we may have impact on theory, applications, and education.

Next, we summarize open problems in the previously mentioned areas, highlighted by the participants: open problems in machine learning, medical imaging, graph signal processing, physical layer wireless communications, and speech and language processing. A common cross-cutting theme that emerged was the opportunity to improve performance by the better integration of accurate physical models into state-of-the-art algorithms.

Open problems in machine learning

Machine learning aims to give computers the ability to learn by exploiting data instead of being explicitly programmed. There are many approaches in machine learning, including support vector machines, decision-tree learning, artificial neural networks, Bayesian networks, genetic algorithms, rule-based learning, and inductive logical programming, among others [3]. In recent years, the fastest growing area of machine learning comes from neural networks and related generative models, where carefully designed hierarchies are built into the overall machine-learning models to form multiple layers of latent representations that disentangle the confounding factors and complexity in the raw data. This type of hierarchical model and the associated machine-learning algorithms are called *deep learning* [1], [2], which represents the most recent and influential advance in machine learning over the past decade. The first successful application of deep learning in real-world tasks came from speech recognition in our signal processing community and was published in this magazine [13], followed quickly with computer vision, natural language processing, robotics, speech synthesis, and image rendering [2].

Despite impressive empirical successes of deep learning and other machine-learning approaches, many open problems remain to be solved. Current deep-learning methods typically lack interpretability, in contrast to traditional machine-learning techniques based on linear models. In a number of applications, deep-learning methods achieve recognition accuracy close to or exceeding that of humans, but they require considerably more training data, power consumption, and computing resources than humans. In addition, although accuracy results are often statistically impressive, they are often unreliable on an individual basis. Finally, most of the current deep-learning models have no reasoning and explaining capabilities, making them vulnerable to disastrous failures or attacks without the ability to foresee and thus to prevent them.

To overcome these challenges, both fundamental and applied research is needed. One potential breakthrough in machine learning is in developing

Digital Object Identifier 10.1109/MSP.2017.2743842
Date of publication: 13 November 2017

interpretable deep-learning models with the aim of creating new algorithms and methods that can overcome current limitations of machine-learning systems in their lack of ability to explain the actions, decision, and prediction outcomes to human users while promising to perceive, learn, decide, and act on their own. This new class of machine-learning systems will allow users to understand and thus trust the system's outputs and to foresee and predict future system behaviors. To this end, neural networks and symbolic systems need to be integrated, enabling the machine-learning systems themselves to construct models that will explain how the world works. That is, they will discover by themselves the underlying causes or logical rules that shape their prediction and decision-making processes interpretable to human users in symbolic and natural language forms. An initial work in this direction makes use of an integrated neural-symbolic representation called *tensor-product neural memory cells,* which can be decoded back to symbolic form without loss of information after extensive learning in the neural-tensor domain.

Another potential breakthrough in machine-learning research is in new algorithms for reinforcement and unsupervised deep learning, which make use of weak or even no training signals paired to inputs to guide the learning. Effective reinforcement-learning algorithms will allow machine-learning systems to learn via interactions with possibly adversarial environments and with themselves.

The most challenging problem, however, is unsupervised learning, for which no satisfactory machine-learning algorithms have been devised so far in practical applications. The development of unsupervised learning techniques is significantly behind that of supervised and reinforcement deep learning. The most recent development in unsupervised learning exploits sequential output structure and advanced optimization methods to alleviate the need for using labels in training prediction systems [12].

Future advances in unsupervised learning include taking into account new sources of learning signals such as the structure of input data and building conditional generative models. In this context, the recent popular topic of generative adversarial networks [2] is a highly promising direction exploiting the long-standing concept of analysis by synthesis. A closely related open problem is multimodal deep learning with cross-domain information as low-cost supervision. Standard speech recognition, image recognition, and text classification methods make use of supervision labels within each of the speech, image, and text modalities separately. This is far from how children learn to recognize speech and classify text. For example, children often get a distant "supervision" signal for speech sounds by an adult pointing to an image scene or text.

A final future direction for tackling open problems in machine learning is the paradigm of learning-to-learn or metalearning; i.e., how to design a machine-learning system that improves or automatically discovers a learning algorithm. Learning-to-learn is a powerful emerging paradigm and a fertile research direction expected to impact a wide range of real-world applications.

Open problems in medical imaging

Medical image reconstruction is the process of forming interpretable images from the data recorded by an imaging system. Until recently, there have been two primary methods for image reconstruction: analytical and iterative. Analytical methods use idealized mathematical models for the imaging system. Typically, these techniques consider only the geometry and sampling properties of the imaging system and ignore details of the system physics and measurement noise. These reconstruction approaches have been used extensively because they require modest computation.

Over the past two decades, image reconstruction has evolved from the exclusive use of analytical methods to a wider use of model-based approaches that account for the physics and statistics. Usually the problems are ill posed, so that maximum-likelihood (ML) methods would propagate excessive noise from the measurements into the reconstructed image. Using priors or regularizers can overcome this limitation. A popular approach is to base iterative methods on maximum a posteriori (MAP) estimation. MAP estimation encompasses 1) modeling the system, 2) developing signal models to serve as priors, 3) developing faster optimization algorithms, and 4) assessing the quality of the reconstructed image.

The transition from analytical to iterative algorithms took place at widely different dates in different modalities. In positron emission tomography (PET) and single-photon emission computed tomography (SPECT), a seminal paper on an expectation maximization (EM) algorithm in the early 1980s led to more than a decade of research before a key acceleration method called *ordered subsets* (*OS*) (related to incremental gradients in the optimization field) helped lead to the commercial adoption of OS-EM for clinical PET and SPECT in about 1997, using an (unregularized) ML approach. This transition provided a dramatic improvement in image quality. Human PET scanners only recently began to provide MAP methods clinically using a modification of a Gaussian Markov random field prior and a convergent OS algorithm.

In X-ray computed tomography (CT), iterative image reconstruction first became available commercially for the CT part of SPECT-CT scanners in about 2010, using a different OS algorithm published a decade earlier. In 2012, the first U.S. Food and Drug Administration (FDA)-approved iterative MAP method targeted at reduced X-ray dose became available for clinical CT, building on an *IEEE Transactions on Signal Processing* paper from two decades earlier. This approach also uses a modified Gaussian MRF to make it edge preserving.

In MRI, researchers studied iterative techniques to quantify relaxation parameters, reconstruct data from multiple receive coils, and correct for magnetic field inhomogeneities. A turning point was the introduction of compressed sensing in about 2005, spawning an explosion of research that finally led to FDA approval of compressed sensing MRI products in 2017 using combinations of total variation regularization and wavelet sparsifying transforms. In all of the aforementioned examples, more than a decade passed between the key publication and commercial availability of the method!

Commercial MAP techniques use relative simple priors defined mathematically. The emerging research trend is to explore signal models that are learned from data. In X-ray CT, there are numerous images acquired at "normal" X-ray doses from which one can learn signal models to use later for reconstructing images from low-dose data. Another data-driven option is to learn a sparse signal model during image reconstruction, rather than relying on training data, called *blind* or *adaptive dictionary* (or *transform*) *learning*. This data-driven evolution provides opportunities for signal processing researchers to explore signal models that better solve inverse problems, particularly from limited or noisy data.

One can "unroll the loop" of an iterative reconstruction algorithm and treat it as a sequence of processing steps akin to a deep neural network and then use data to train more aspects of the processing chain. Recent conferences have seen an explosion of such methods. There are many significant challenges because such algorithms are arguably even more nonlinear (and opaque) than the edge-preserving regularization techniques used clinically today. Can one characterize the "resolution" and "noise" properties of such algorithms? What is the best training metric: MSE or diagnostic image quality? What if a patient has significantly different image features than those found in the training data? How well will a method trained for one system configuration (e.g., a certain set of coils in MRI or a certain set of angular views and pitch in CT) generalize to other configurations? Some experts have conjectured that "machine learning will transform radiology significantly within the next five years" but others point out there are significant technical and legal challenges. These questions and more should provide numerous research opportunities for signal processors interested in inverse problems like medical imaging [11].

Open problems in graph signal processing

Today's data is being generated at an unprecedented rate from a diversity of sources. Examples include profile information in social networks, stimuli in brain connectivity networks, and traffic flow in city street networks, among others. A decade ago, a typical data set was supported on a regular lattice; today, the story is quite different. Data is supported on complex and irregular structures. Often, these structures are modeled by graphs, as they are able to describe both the structure and the data associated with that structure. For example, in an online social network, a user's profile may contain the user's date of birth, school attended, professional organizations, and more. Each of these attributes can form a subnetwork with different properties. Using graphs, we want to analyze data supported on such complex structures, allowing us to mine information from online social networks, transportation networks, the power grid, and more, in the same context. While these are representatives of physical-world graphs, other graphs may include abstract concept networks such as knowledge graphs and correlation graphs.

Data science on graphs has been considered from several angles by graph theory, network science, and graph mining, all dealing with graph structure. More recently, the area of graph signal processing has emerged, formalizing the addition of metadata as signals on a graph [4]–[6]. Graph signal processing aims to extend classical signal processing tasks and tools to data on irregular structures modeled by graph signals (see Figure 1). The goal is to gain an understanding of the intrinsic structure of the data by using tools well understood on regular structures, such as filtering and Fourier transforms, and to perform tasks such as sampling, restoration, compression, and topology learning.

Signal processing on graphs is an active area of research; many challenges and opportunities still remain. For example, a number of basic concepts in statistical signal processing and sampling theory have not yet been entirely extended to graphs in a unified way. More advanced challenges include the scale of the data, its heterogeneity, distributed analysis and processing, fusing data from different scales and resolutions, and processing tensor values defined on nodes. Disparate communities such as network science, machine learning, and signal processing are all currently working on these challenges with the tendency to attack such problems either via learning methods or by building models; an important path for advancing this field and dealing effectively with the deluge of data is to combine the tools and integrate these different approaches.

Open problems in physical layer wireless communications

Wireless communications have been a major driver of signal processing research for at least the past three decades, spurred by the development of widespread consumer mobile communications and other applications, which today impact the lives of billions of people—indeed, most people alive today. Here we focus on research in the physical layer of mobile communication networks where signal processing has perhaps had the greatest impact.

Modern mobile communication networks have been through four generations to date, and the fifth generation (5G)

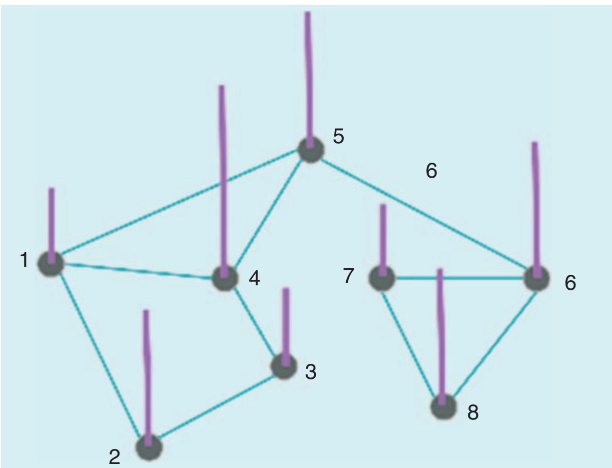

FIGURE 1. A graph signal models data (values on the graph nodes) supported on complex structures (graph nodes).

is rapidly emerging. The key enablers of the most recent deployed generation of mobile networks, the so-called fourth generation (4G), have been the development of methods to exploit the spatial diversity afforded by the wireless medium in the forms of multiple-input, multiple output (MIMO) antenna systems, cooperation, and relaying; the exploitation of frequency diversity through the use of orthogonal frequency-division multiple access signaling; and the development of methods to approach link capacity via the iterative decoding of turbo or low-density parity-check codes. These signal processing advances have allowed 4G networks to meet the challenge of real-time multimedia communications that has been the primary advance of 4G over its predecessors.

The emerging generation of mobile networks, 5G, presents a number of new signal processing challenges. Beyond providing adequate capacity and reliability, 5G networks also add the issue of energy efficiency, required to support several new applications areas. These include the so-called Internet of Things (IoT), which is envisioned to involve orders-of-magnitude more terminals than 4G networks in highly densified configurations of low-complexity terminals; systems requiring autonomy or telecontrol, in which low latency and very high reliability are critical; and immersive experiences, such as virtual reality, which require very high bandwidth streaming [7].

These requirements give rise to a number of open problems and potential solutions. Solutions enabling densification and the consequent interference management include cloud radio access networks, massive MIMO systems, millimeter wave techniques, and transceivers that can harvest radio-frequency energy from their surroundings. Substantial capacity enhancements are also needed, and some techniques for providing greater capacity (in addition to densification of resources) include full duplex transmission and nonorthogonal multiple-access techniques, both of which will be enabled by sophisticated signal processing. Security is another issue in which signal processing has a key role to play; traditionally, security has been a higher-layer issue, with encryption being a primary mechanism. However, with highly dense networks of low-complexity terminals connected via loosely organized networks, new methods are needed. Physical layer security is such a promising method, which relies on signal processing techniques, such as coding, beamforming, and signal design. Finally, many emerging applications, such as autonomous vehicles and factory automation, require low-latency, high-reliability communications via short packets. Since the existing theory of reliable data transmission is largely based on analyses in the asymptote of infinite block-length, new theories are needed to understand the limits of reliable communication in this regime. In addition, in applications such as autonomous driving, worst-case metrics may be more appropriate than the standard average-case analysis.

Open problems in speech and language processing

Spoken language processing encompasses methods and techniques for transforming and manipulating speech, text, and a wide variety of related symbolic representations. Examples are speech recognition (speech→words), natural language understanding (words→meaning), natural language generation (meaning→words), speech synthesis (words→speech), and machine translation (words in L1→words in L2). Modern applications of spoken language processing will typically incorporate many if not all of these component technologies [8]–[10]. For example, intelligent agents such as Siri and Alexa require all of the aforementioned technologies to support conversations over a wide range of topics in many languages.

Since virtually all spoken language processing involves classification and/or prediction, modern approaches depend heavily on statistical models and machine learning. A major breakthrough in recent years has been the widespread deployment of deep learning [9]. The ability of neural networks to automatically learn low-level features, the use of attention mechanisms to learn which features are important, and the flexibility to scale parameter sets both in width and depth has led to significant performance improvements. For example, word error rates for real-time large vocabulary speaker-independent speech recognition are now routinely below 10%, and speech synthesis quality is acceptable for most applications.

The renaissance of neural networks has also been the catalyst for the development of a powerful toolbox of core network components (such as deep neural networks, long short-term memory networks, convolutional neural networks, and more) and development tools (such as TensorFlow, Torch, and others), which allow solutions to complex problems to be assembled, trained, and deployed quickly and at a relatively low cost.

Despite the undoubted progress witnessed over the last decade, there remain many challenges. The recognition of fluent conversations between human speakers and speech in high levels of background noise or in the presence of a competing talker still falls well short of human performance. Our ability to understand the meaning of natural language sentences, especially in the context of past interactions and a changing real-world environment, remains extremely limited.

Two emerging trends aimed at addressing some of the challenges are continuous representations and end-to-end training. In particular, there is currently a shift away from symbolic representations to continuous space representations. An already well-established example of this is the use of word embeddings. By projecting discrete words into a continuous high-dimensional space, many of the problems associated with synonyms, antonyms, and rare words are mitigated by the use of simple well-behaved distance metrics. The extension of embeddings to represent whole sentences and conversations enables variable-length sequences to be mapped into fixed-length vectors that can then be manipulated using conventional classification and prediction models. There is also increasing emphasis on end-to-end training. Conventional systems are typically built as a pipeline of processes for which each component interface needs to be explicitly defined, and training data needs to be appropriately labeled at every component interface. This is expensive and inevitably results in information loss as the signal propagates through the pipeline. By treating component interfaces as hidden variables, and training end to end, costs are reduced and performance increases.

In summary, the extensive use of machine learning coupled with the availability of large-scale computing and very large data sets have led to a significant improvement across all areas of speech and language processing. Ultimately, however, the real challenge will concern our ability to extract and manipulate the underlying meaning of word sequences and this is an area that has so far remained rather elusive.

Authors

Yonina C. Eldar (yonina@ee.technion.ac.il) is a professor in the Department of Electrical Engineering at the Technion–Israel Institute of Technology, Haifa, where she holds the Edwards Chair in Engineering. She is also an adjunct professor at Duke University, Durham, North Carolina, a research affiliate with the Research Laboratory of Electronics at the Massachusetts Institute of Technology, and was a visiting professor at Stanford University, California. She is a member of the Israel Academy of Sciences and Humanities and is a Fellow of the IEEE and EURASIP. She has received many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award, the IEEE/AESS Fred Nathanson Memorial Radar Award, the IEEE Kiyo Tomiyasu Award, the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, and the Wolf Foundation Krill Prize for Excellence in Scientific Research. She is the editor-in-Chief of *Foundations and Trends in Signal Processing* and serves the IEEE on several technical and award committees.

Alfred O. Hero III (hero@eecs.umich.edu) received the B.S. degree (summa cum laude) from Boston University in 1980 and the Ph.D. degree from Princeton University in 1984, both in electrical engineering. He is the John H. Holland Distinguished University Professor of Electrical Engineering and Computer Science and the R. Jamison and Betty Williams Professor of Engineering at the University of Michigan, Ann Arbor. He is also the codirector of the university's Michigan Institute for Data Science. He has served as president of the IEEE Signal Processing Society and as a member of the IEEE Board of Directors. He has received numerous awards for his scientific research and service to the profession, including several best paper awards, the IEEE Signal Processing Society Technical Achievement Award in 2013, and the 2015 Society Award. He is a Fellow of the IEEE.

Li Deng (l.deng@ieee.org) received his Ph.D. degree from the University of Wisconsin–Madison. He was a tenured professor from 1989 to 1999 at the University of Waterloo, Ontario, Canada, and then joined Microsoft Research, Redmond, Washington, where he was the chief scientist of artificial intelligence (AI) and partner research manager. He recently joined Citadel as its chief AI officer. He is a Fellow of the IEEE, the Acoustical Society of America, and the International Speech Communication Association. He was the editor-in-chief of *IEEE Signal Processing Magazine* (2009–2011) and *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2012–2014), for which he received the IEEE Signal Processing Society (SPS) Meritorious Service Award. He received the 2015 IEEE SPS Technical Achievement Award for "outstanding contributions to automatic speech recognition and deep learning" and numerous best paper and scientific awards for the contributions to AI, machine learning, multimedia signal processing, speech and human language technology, and their industrial applications.

Jeff Fessler (fessler@umich.edu) received his B.S.E.E. degree from Purdue University, West Lafayette, Indiana, in 1985 and his Ph.D. degree from Stanford University, California, in 1990. He is the William L. Root Professor of Electrical Engineering and Computer Sciences at the University of Michigan, where he has worked since 1990. He received the 2013 IEEE Edward Hoffman Medical Imaging Scientist Award and serves as an associate editor of *IEEE Transactions on Computational Imaging*. He was the technical program cochair of the 2002 IEEE International Symposium on Biomedical Imaging (ISBI) and the general chair of ISBI 2007. His research group focuses on imaging problems. He is a Fellow of the IEEE.

Jelena Kovačević (jelenak@cmu.edu) received her M.S. and Ph.D. degrees in electrical engineering from Columbia University, New York, in 1988 and 1991, respectively. She is currently the Hamerschlag University Professor, head of the Department of Electrical and Computer Engineering, and a professor of biomedical engineering at Carnegie Mellon University (CMU), Pittsburgh, Pennsylvania. She received the IEEE Signal Processing Society Technical Achievement Award, the Dowd Fellowship at CMU, the Belgrade October Prize, and the E.I. Jury Award at Columbia University. She has coauthored a number of award-winning papers and is a coauthor of the textbooks *Wavelets and Subband Coding* and *Foundations of Signal Processing*. She is a Fellow of the IEEE and was the editor-in-chief of *IEEE Transactions on Image Processing*. Her research interests include applying data science to a number of domains such as biology, medicine, and smart infrastructure; she is an authority on multiresolution techniques such as wavelets and frames.

H. Vincent Poor (poor@princeton.edu) is the Michael Henry Strater University Professor of Electrical Engineering at Princeton University, New Jersey. His interests include information theory and signal processing, with applications in wireless networks and related fields. He is an IEEE Fellow, a member of the U.S. National Academy of Engineering and the U.S. National Academy of Sciences, and a Foreign Member of the Royal Society. He received the Technical Achievement and Society Awards of the IEEE Signal Processing Society in 2007 and 2011, respectively. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal and honorary doctorates from several universities.

(continued on page 23)

ICIP 2016 Competition on Mobile Ocular Biometric Recognition

Rattani, A.; Derakhshani, R.; Saripalle, S.K; Gottemukkula, V.

The aim of this competition is to evaluate and compare the performance of mobile ocular biometric recognition schemes in visible light on a large scale database (VISOB Data Set ICIP 2016 Challenge Version) using standard evaluation methods. Four different teams from universities across the world participated in this competition, submitting five algorithms altogether. The best results were obtained by a team from Norwegian Biometrics Laboratory (NTNU, Norway).

2016

Semantic Context and Depth-Aware Object Proposal Generation

Zhang, H.; He, X.; Porikli, F.; Kneip, L.

This paper presents a context-aware object proposal generation method for stereo images. The authors propose to incorporate additional geometric and high-level semantic context information into the proposal generation.

2016

Super-Resolution of Compressed Videos Using Convolutional Neural Networks

Kappeler, A.; Yoo, S.; Dai, Q.; Katsaggelos, A.K.

In this paper, for the problem of compressed video superresolution, the authors propose a CNN that is trained on both the spatial and the temporal dimensions of compressed videos to enhance their spatial resolution. Consecutive frames are motion compensated and used as input to a CNN that provides superresolved video frames as output.

2016

Classification of Hyperspectral Image Based on Deep Belief Networks

Li, T.; Zhang, J.; Zhang, Y.

In this paper, deep-learning frameworks, the restricted Boltzmann machine model, and its deep structure deep belief networks are introduced in hyperspectral image processing as the feature extraction and classification approach.

2014

Image Character Recognition Using Deep Convolutional Neural Network Learned from Different Languages

Bai, J.; Chen, Z.; Feng, B.; Xu, B.

This paper proposes a shared-hidden-layer deep convolutional neural network (SHL-CNN) for image character recognition. In SHL-CNN, the hidden layers are made common across characters from different languages, performing a universal feature extraction process that aims at learning common character traits existing in different languages, such as strokes, while the final softmax layer is made language dependent, trained based on characters from the destination language only.

2014

SP

PANEL AND FORUM *(continued from page 13)*

Steve Young (sjy@eng.cam.ac.uk) is a professor of information engineering at the University of Cambridge, United Kingdom, and a senior member of technical staff at Apple. His main research interests lie in the area of statistical spoken language systems, including speech recognition, speech synthesis, and dialog management. He is the recipient of a number of awards including an IEEE Signal Processing Society Technical Achievement Award and the IEEE James L. Flanagan Speech and Audio Processing Award. He is a Fellow of the IEEE and the U.K. Royal Academy of Engineering. In addition to his academic career, he has also founded a number of successful start-ups in the speech technology area.

References

[1] L. Deng and D. Yu, "Deep learning: Methods and applications," in *Foundations and Trends in Signal Processing Series.* Boston, MA: NOW Publishers, 2014.

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* Cambridge, MA: MIT Press, 2016.

[3] L. Deng and X. Li, "Machine learning paradigms for speech recognition: An overview," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, pp. 1060–1089, May 2013.

[4] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, pp. 83–98, May 2013.

[5] A. Sandryhaila and J. M. F. Moura, "Big data processing with signal processing on graphs," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, Sept. 2014.

[6] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.

[7] F.-L. Luo and C. Zhang, Eds. *Signal Processing for 5G: Algorithms and Implementations.* Chichester, U.K.: Wiley-IEEE Press, 2016.

[8] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, Aug. 2011.

[9] L. Deng, "Industrial technology advances: Deep learning: From speech recognition to language and multimodal processing," *APSIPA Trans. Signal Inform. Process.*, vol. 5, pp. 1–15, Jan. 2016.

[10] N. Mrksic, I. Vulic, D. O. Seaghdha, I. Leviant, R. Reichart, M. Gasic, A. Korhonen, and S. Young, "Semantic specialization of distributional word vector spaces using monolingual and cross-lingual constraints," *Trans. Assoc. Computat. Linguistics*, vol. 5, pp. 309–324, Sept. 2017.

[11] J. A. Fessler. (2017). Medical image reconstruction: A brief overview of past milestones and future directions. [Online]. Available: http://arxiv.org/abs/1707.05927

[12] Y. Liu, J. Chen, and L. Deng, "Unsupervised sequence classification using sequential output statistics," in *Proc. NIPS,* Dec. 2017.

[13] G. Hinton, et al., "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Process. Mag.,* vol. 29, no. 6, pp. 82–97, Nov. 2012.

SP