Namrata Vaswani

A Feature Article Cluster on Exploiting Structure in Data Analytics: Low-Rank and Sparse Structures

Individual feature articles and special issues are two major mechanisms of full-length tutorial surveys of *IEEE Signal Processing Magazine* (*SPM*). Since the May 2016 feature article cluster by Jane Wang et al. on brain signal analytics, *SPM*'s current and past editors-in-chief and their teams have been exploring a different way to complement this existing structure—a feature article cluster (or mini special issue) that allows for a set of three to five solicited articles on a current topic, instead of just one (feature article) or ten to 11 (a special issue). This issue of *SPM* offers a feature article cluster on exploiting structure in data analytics: low-rank and sparse structures and is the second such cluster. It is the first in *SPM*'s planned yearly series on data science and includes the following four articles:

- "Harnessing Structures in Big Data via Guaranteed Low-Rank Matrix Estimation" by Chen and Chi
- "Robust Subspace Learning" by Vaswani et al.
- "Correlation-Awareness in Low-Rank Models" by Pal
- "Theoretical Foundations of Deep Learning via Sparse Representations" by Papyan et al.

In today's big data age, data is generated everywhere around us. Examples include texts, Tweets, network traffic, changing Facebook connections, or video surveillance feeds coming in from one or multiple cameras. Much of this data is high dimensional, and a lot of it is also highly noisy, outlier corrupted, or incomplete. Thus, the first step before processing such data is noise/outlier removal and/or filling in the missing entries, along with dimension reduction. All of these tasks are hard and ill-posed without any structural (prior) assumptions on the data. Two of the most commonly used and practically valid structural assumptions for high-dimensional data sets are sparsity and low rank. When sparsity is exploited, it is now well known that the signal can be recovered from a highly undersampled set of its linear projection measurements, under mild assumptions. This idea, referred to as *sparse recovery* or *compressive sensing* and its various extensions, is now well known, well studied, and well reviewed.

In this issue

This feature article cluster focuses on data recovery methods that exploit the other most commonly valid structural assumption on data sets—low rank, as well as on those that exploit both low-rank and sparse structures for different parts of a data set.

The first article by Chen and Chi focuses on low-rank matrix recovery from incomplete data, with and without the use of additional constraints. Low-rank modeling plays a pivotal role in signal processing and machine learning, with applications ranging from collaborative filtering to dimensionality reduction and adaptive filtering. In recent years, progress has been made in understanding how to exploit the low-rank assumption while still obtaining computationally efficient and provably correct solutions. This article reviews the literature on convex relaxation approaches such as nuclear norm minimization, as well as more recent nonconvex procedures, such as projected gradient descent or alternating minimization, which are much faster, while needing only a little more measurements, provably.

Vaswani et al.'s article focuses on the related topic of low-rank matrix recovery from outlier-corrupted data and simultaneously exploits low-rank and sparsity assumptions on different components of a data set. It reviews both provably correct (and useful) as well as empirically useful solutions to robust principal component analysis (PCA) while also covering the closely related topics of robust subspace recovery and robust subspace tracking. PCA is one of the most widely used dimension reduction techniques. Given relatively clean data, it easily solved via singular value decomposition (SVD). The problem of subspace learning or PCA in the presence of outliers is called *robust subspace learning* or *robust PCA* (*RPCA*). When the outliers are modeled as sparse corruptions (corruptions in only a few entries), we call the problem *RPCA*, while when an entire data vector is assumed to be either an outlier or an inlier, we call the problem *robust subspace recovery*. For long data sequences, if one tries to use a single lower-dimensional subspace to represent the data, the required subspace dimension may end up being quite large. For such data, a better model is to assume that it lies in a low-dimensional subspace

Digital Object Identifier 10.1109/MSP.2018.2828358
Date of publication: 27 June 2018

that can change over time, albeit gradually. The problem of tracking such data (and the subspaces) while being robust to outliers is called *robust subspace tracking*.

The third article in this four-part series is by Pal and reviews new deterministic sampling techniques, reconstruction algorithms, and performance guarantees for low-rank correlation-driven estimation problems. In many problems, the information of interest (often captured in suitable parameters) is embedded in the correlation of the data. In such cases, it is possible to compress the raw data and yet perfectly recover its correlation matrix. Unlike conventional compressed sensing, which uses a sparse representation of the data to achieve compression, here compression can be achieved without sparsity, simply by exploiting certain specific correlation structure of the data, such as Toeplitz structure, that arise in a large number of statistical signal processing and superresolution imaging problems. The article shows how to optimally compress low-rank Toeplitz covariance matrices using structured sampling ideas, along with robust recovery algorithms that do not require regularization. Fundamental lower bounds on source localization from such compressed covariance matrices are also discussed. Finally, by considering the sparse Bayesian learning framework, it is illustrated that the ability to exploit correlation structure in addition to sparsity can enable significantly higher levels of compression compared to what can be attained by using sparsity alone.

All four articles focus on provable solutions that are also fast and practically useful, while also briefly reviewing everything else that exists. Detailed experimental comparisons help demonstrate the practical implications of the theoretical guarantees as well as make it easy for a practitioner to pick the most suitable approach. Detailed discussions of open questions are also provided.

We end the series with the article by Papyan et al. The great success of deep learning in the past decade has been mostly empirically based. Indeed, a solid theory explaining the proposed architectures, the algorithms used, and the superb performance obtained by this field has been lagging behind. This article presents a systematic theoretical framework for explaining deep learning based on data modeling via sparse representation. It proposes a multilayer sparse model that describes the data's inner structure, showing that decomposing these signals into their building atoms amounts to various deep convolutional neural network architectures. This observation is accompanied by a new and exciting ability to theoretically analyze the performance of these networks, posing clear conditions on terms for their success. The article offers a gradual and stand-alone description of this story, starting from the general need for models, passing through the story of sparse representation theory and convolutional dictionaries, then turning to the main message of tying these to the realm of deep learning. Time will tell whether this model-based view will be adopted by our community as the gateway to the much-needed theory for deep learning.

About the guest editor

Namrata Vaswani (namrata@iastate.edu) received her B.Tech. degree from the Indian Institute of Technology (IIT-Delhi) in 1999 and her Ph.D. degree in 2004 from the University of Maryland, College Park. She is a professor of electrical and computer engineering, and (by courtesy) of mathematics, at Iowa State University. Her research interests lie at the intersection of statistical machine learning/data science, computer vision, and signal processing. Her recent research has focused on practically useful and provably correct solutions for dynamic high-dimensional data problems such as dynamic compressive sensing (CS) and robust principal component analysis. She received the 2014 IEEE Signal Processing Society Best Paper Award for her modified-CS work that appeared in *IEEE Transactions on Signal Processing* in 2010. She is the special issues area editor of *IEEE Signal Processing Magazine*, has served twice as an associate editor of *IEEE Transactions on Signal Processing*, and is the lead guest editor of a *Proceedings of the IEEE* Special Issue on Rethinking PCA for Modern Data Sets.

SP