Klaus-Robert Müller, Tülay Adali,
Kenji Fukumizu, José C. Principe,
and Sergios Theodoridis

Special Issue on Advances in Kernel-Based Learning for Signal Processing

The importance of learning and adaptation in statistical signal processing creates a symbiotic relationship with machine learning. However, the two disciplines possess different momentum and emphasis, which makes it attractive to periodically review trends and new developments in their overlapping spheres of influence. Looking at the recent trends in machine learning, we see increasing interest in kernel methods, Bayesian reasoning, causality, information theoretic learning, reinforcement learning, and nonnumeric data processing, just to name a few. While some of the machine-learning community trends are clearly visible in signal processing, such as the increased popularity of the Bayesian methods and graphical models, others such as kernel approaches are still less prominent. However, kernel methods offer a number of unique advantages for signal processing, and this special issue aims to review some of those.

KERNEL-BASED LEARNING: BACKGROUND

The application of reproducing kernel Hilbert space (RKHS) methodology in statistical signal processing was proposed by Emmanuel Parzen in the late 1950s, who provided for the first time a functional analysis perspective of random processes defined by second-order moments [1]. Parzen clearly illustrated that the RKHS approach offers an elegant general framework for minimum variance unbiased estimation of regression coefficients, least squares estimation of random variables, detection of known signals in Gaussian noise, among others. Although these problems involve random variables, they can be solved algebraically in the RKHS associated with their covariance functions with all the geometric advantages of the inner product defined in such spaces. In the early 1970s, Kailath presented a series of detailed papers on the RKHS approach for detection and estimation to demonstrate its usefulness in computing likelihood ratios, testing for nonsingularity, bounding signal detectability, and determining detection stability [2]–[4]. Figueiredo [5] took a different approach to apply RKHS in a nonlinear system and signal analysis by building the RKHS bottom-up using arbitrarily weighted Fock spaces. The spaces are defined by Hilbert–Schmidt polynomials or power series in either scalar or multidimensional variables. But this interest was short lived.

KERNEL METHODS OFFER A NUMBER OF UNIQUE ADVANTAGES FOR SIGNAL PROCESSING, AND THIS SPECIAL ISSUE AIMS TO REVIEW SOME OF THOSE.

In machine learning, kernel-based learning has become a state-of-the-art technology within the last two decades. In particular, kernel methods have enriched the spectrum of machine learning and statistical methods with a vast new set of nonlinear algorithms. After the advent of support vector machines [6] some 19 years ago, kernel principal component analysis (kPCA) [7] was established as a blueprint for "kernelizing" linear scalar product-based algorithms, given that a conditionally positive-definite kernel is used [8]. The so-called empirical kernel map [9] allows preprocessing of data by projecting it onto the leading kPCA components; thus, nonlinear variants of algorithms could be constructed via a nonlinear transformation. By virtue of the application of the kernel trick [10], a scale of nonlinearity can be imposed that reflects the nature or our prior knowledge in learning [8].

KERNEL-BASED LEARNING AND SIGNAL PROCESSING

The initial activity in the field included the introduction of a number of novel kernels as well as novel nonlinear algorithm variants. More recent developments, however, emphasized extending the use of kernel-based learning beyond mere classification and regression settings towards more complex scenarios motivated by important signal processing and learning scenarios: analysis of structured data, ranking, the blending of Bayesian methods and kernel methods, semisupervised settings, testing, and causality, just to name a few. A further important direction was the development of large-scale learning methods for establishing fast solvers for the respective optimization problems. An overarching motivation for the fast development in kernel-based learning has primarily come from the application side. On one side, the sciences (e.g., neuroscience, computational biology, natural language processing, and physics) have motivated the work and these fields largely benefited from the novel set of tools. On the other side, kernel-based learning enabled the addressing of emerging industrial problems including social networks, text mining, and the

Digital Object Identifier 10.1109/MSP.2013.2253031
Date of publication: 12 June 2013

general urge to better understand large and complex corpora of data. Thus, the blending of signal processing and kernel-based learning has become more and more seamless and evident over recent years.

CONTRIBUTIONS TO THIS SPECIAL ISSUE

While it is impossible to span the whole space outlined above of the interesting advancements of kernel-based methods for signal processing, this special issue of *IEEE Signal Processing Magazine (SPM)* has focused on a number of timely and interesting topical contributions with review character.

Arenas-García et al. consider kernel-based methods for multivariate feature extraction, while Jenssen outlines the interplay of Renyi entropy and nonlinear dimensionality reduction. Pérez-Cruz et al. discuss Gaussian processes, their use and implementation introducing gently and relating to Wiener filtering, and their use in signal processing. Särkkä et al. present methods, using the Kalman filter formalism, for converting spatiotemporal Gaussian process regression into infinite dimensional state-space models, which makes Gaussian processes computationally feasible and opens one more door to combine machine-learning and signal processing methodologies. Nonlinear kernel methods are difficult to interpret and analyze, thus, Montavon et al. provide a comprehensive framework to quantify the effective dimensionality, signal-to-noise ratio, and local error bars for a learning problem; in addition, techniques to provide a local explanation of the nonlinear model are reviewed.

Along a related line of thought describing the manifold characteristics of learning, Talmon et al. present the concept of diffusion maps for signal processing and relate it to kPCA and nonlinear filtering. A recent and very active field has been the use of kernel-based methods for hypothesis testing. Harchaoui et al. describe RKHS embeddings of probability distributions; then tests are merely divergences between these embedded distributions. Song et al. discuss a novel RKHS-embedding approach to inference with graphical modeling, including belief propagation and general Bayes' rule. Bazerque and Giannakis also take the RKHS view, however, taking the slant of sparse modeling in conjunction with nonparametrics. Equipped in this manner, sparse kernel-based learning becomes a powerful tool in a number of signal processing applications, such as cognitive radio.

Focusing in particular on the open issues in cognitive radio networks, Ding et al. demonstrate effective signal processing solutions by kernel-based learning. Zhao et al. present two effective models for kernel-based extension of tensor decompositions, which enable investigation of multiway nonlinear dependencies in structured data. Applications examples include reconstruction of three-dimensional movement trajectories from electrocorticography signals recorded from a monkey's brain and human action classification based on video sequences. The application of RKHS methodology to neural signal processing is addressed by Park et al. This application exemplifies how the footprint of statistical signal processing can be enlarged to include the spike train space, where an algebra cannot be defined. The approach defines an injective mapping from this space to a RKHS by means of kernels, which allows the conventional signal processing algorithms of optimal filtering to be readily applied to multiple spike trains, preserving the structure of event timing.

THE BLENDING OF SIGNAL PROCESSING AND KERNEL-BASED LEARNING HAS BECOME MORE AND MORE SEAMLESS AND EVIDENT OVER RECENT YEARS.

Finally, we would like to note that the articles in this special issue have undergone a very careful reviewing process; most of them going through multiple revisions. We would like to cordially thank the many reviewers, who have helped us improve the quality of the final articles, as well as Special Issues Area Editor Fulvio Gini and Coordinator for Society Publications Rebecca Wollman for their great support. We hope that the readers of *IEEE Signal Processing Magazine* will find this collection of review articles useful, interesting, and stimulating.

ACKNOWLEDGMENTS

Klaus-Robert Müller acknowledges funding by the German Science Foundation (DFG), the German Ministry of Research (BMBF), the European Union, and the World Class University Program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology (grant R31-10008). Tülay Adali acknowledges funding from the National Science Foundation and the National Institutes of Health.

REFERENCES

[1] E. Parzen, "Statistical inference on time series by Hilbert space methods," Stat. Dept., Stanford Univ., CA, Tech. Rep. 23, 1959.

[2] T. Kailath, "RKHS approach to detection and estimation problems—Part I: Deterministic signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-17, no. 5, pp. 530–549, Sept. 1971.

[3] T. Kailath and D. Duttweiler, "An RKHS approach to detection and estimation problems—Part III: Generalized innovations representations and a likelihood-ratio formula," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 6, pp. 30–45, Nov. 1972.

[4] T. Kailath and H. Weinert, "An RKHS approach to detection and estimation problems—Part II: Gaussian signal detection," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 1, pp. 15–23, Jan. 1975.

[5] R. DeFigueiredo, "A generalized Fock space framework for nonlinear system and signal analysis," *IEEE Trans. Circuits Syst.*, vol. CAS-30, no. 9, pp. 637–647, Sept. 1983.

[6] C. Cortes and V. N. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[7] B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[8] A. J. Smola, B. Schölkopf, and K.-R. Müller, "The connection between regularization operators and support vector kernels," *Neural Netw.*, vol. 11, no. 4, pp. 637–649, 1998.

[9] B. Schölkopf, S. Mika, C. J. C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch, and A. J. Smola, "Input space versus feature space in kernel-based methods," *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 1000–1017, Sept. 1999.

[10] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Ann. ACM Workshop on Computational Learning Theory*, 1992, pp. 144–152.

[SP]