system generator of Xilinx. Many FPGA chips also contain embedded processors rendering them complete platforms for DSP system design.

SP AND ERROR CORRECTION FOR NONVOLATILE MEMORY DEVICES

Nonvolatile storage devices in the form of NAND flash memories and solid-state drives have become the storage techniques of choice in many mobile and portable devices. The continued density growth of these devices has been mainly driven by aggressive technology scaling and the use of multilevel per-cell techniques. However, bit errors are becoming more severe as memory process technology scales down below 40 nm. Error-control coding techniques have been employed to improve the endurance and performance of NAND flash memories. However, traditional error correction codes [BCH/Reed-Solomon (BCH/RS)] suffer from increased overhead in coding redundancy and read latency as the number of errors increases. In addition, the number of electrons stored in a memory cell is decreasing with every generation of flash memory resulting in weak signals that require enhanced sensing techniques. Research challenges include reduced complexity coding, enhanced threshold sensing, and adaptive interference canceling techniques.

Slides

DSP-ASSISTED ANALOG AND RF CIRCUITS

As complementary metal–oxide–semiconductor (CMOS) process technology keeps shrinking, the analog portion of SoCs is increasingly dominating the silicon area and power consumption because analog components do not scale well with Moore's law, as does digital logic. Traditional matching techniques that compensate for process variations do not work well as CMOS feature size scales down. Thus, to enhance performance and reduce power consumption of analog/RF circuits, it is necessary to utilize DSP techniques that can be realized with essentially free digital logic. Examples include very high performance analog-to-digital convertors with self-calibration, RF power amplifier linearization, and intermediate-frequency sampling receivers.

AUTHORS

Mohammad M. Mansour (mmansour@aub.edu.lb) is an associate professor at the American University of Beirut.

Liang-Gee Chen (lgchen@cc.ee.ntu.edu.tw) is a professor at National Taiwan University.

Wonyong Sung (wysung@snu.ac.kr) is a professor at Seoul National University.

Video

Tülay Adalı, David J. Miller,
Konstantinos I. Diamantaras, and Jan Larsen

Trends in Machine Learning for Signal Processing

By putting the accent on learning from the data and the environment, the Machine Learning for SP (MLSP) Technical Committee (TC) provides the essential bridge between the machine learning and SP communities. While the emphasis in MLSP is on learning and data-driven approaches, SP defines the main applications of interest, and thus the constraints and requirements on solutions, which include computational efficiency, online adaptation, and learning with limited supervision/reference data. While MLSP has always been an active area, it is now converging toward the very center of activity in SP research due primarily to two underlying reasons:

- As data now come in a multitude of forms and natures, it has become evident that solutions must emphasize both learning from the data and minimizing unjustified assumptions about the data generation mechanism. Simplifying assumptions such as Gaussianity, stationarity, and circularity can no longer be easily justified, and nonlinearity plays a more important role in today's problems.
- Almost all of the new application areas in SP emphasize the importance of interdisciplinary research, i.e., the need both to work closely with the target application domain and discipline and the need to leverage suitable ideas and tools from diverse disciplines, to develop the best solutions. Many new applications are also sufficiently complex that they require use of multiple, interacting tools and they may need to meet multiple, simultaneous objectives—e.g., both signal classification and signal enhancement, or simultaneous classification and biomarker discovery in medical applications.

Indeed, these two aspects are at the heart of MLSP research. Since learning is emphasized, MLSP methods have always been more data driven than

Digital Object Identifier 10.1109/MSP.2011.942319
Date of publication: 1 November 2011

model driven; however, the "integrative" nature of MLSP research has also always been emphasized—whenever available, reliable domain information has been integrated, in a principled (e.g., a Bayesian) fashion, into the solutions. Thus, now besides being attractive for many of the traditional SP applications such as pattern recognition, speech, audio, and video processing, MLSP techniques are the primary candidates for a new wave of emerging applications such as brain-computer interface, multimodal data fusion and processing, behavior and emotion recognition, and learning in environments such as social networks and dynamic networks in general.

In what follows, we first discuss current areas of significant activity as well as emerging trends, first in terms of theory, and then applications. Specifically, we discuss the trends in learning theory, and in particular, discuss a major paradigm shift in learning as demonstrated by cognitive information processing. Then, we discuss the role MLSP plays in a number of key emerging application areas.

TRENDS IN LEARNING THEORY

In terms of theory, graphical and kernel methods, Bayesian learning, information-theoretic learning, and sequential learning have always been important areas of activity within MLSP. The need for nonlinear adaptive algorithms for advanced SP and streaming databases has fueled interest in a number of areas, including sequential active learning, which includes as an important subclass: kernel adaptive filters. Sequential learning algorithms are a fundamental tool in adaptive SP and intelligent learning systems as they embody an efficient compromise among constraints such as accuracy, algorithmic simplicity, robustness, low latency, and fast implementation. In addition, by defining an instantaneous information measure on observations, kernel adaptive filters are able to "actively" select training data in online learning scenarios. This active learning mechanism provides a principled framework for knowledge discovery, redundancy removal, and anomaly detection.

DISTRIBUTED LEARNING

With ever-growing data set sizes in real-world applications involving petabytes of information, it is becoming increasingly important to distribute learning tasks by assigning subsets of data to different processors. The processors thus need to communicate and exchange information in such a way that the overall system collectively solves the problem in an optimal manner. There is a wide range of such applications, including multiple agent coordination, estimation and classification/detection problems in sensor networks, and packet routing problems, among many others. The development of learning algorithms for distributed and/or cooperative scenarios, where several nodes have to solve the same or similar classification/estimation/clustering tasks, is therefore becoming an important area of increasing interest within the machine learning community. Typically, algorithms that work in these environments need to conform to limitations in data sharing among the nodes due to either energy/bandwidth constraints or privacy issues. Related applications that are inherently distributed include sensor networks problems and learning in social networks.

SPARSITY-AWARE LEARNING

Sparsity is a natural property of many systems of interest in SP, and sparsity-aware systems have been shown to offer improved performance over sparsity-agnostic ones. Accordingly, this is a topic of growing interest. In mobile communications, for example, MLSP methods can exploit the sparsity present in the network to improve the estimation of channel parameters and timing delays. Sparsity can be due to user inactivity, to the structure of the channel, or to the network topology.

SEMISUPERVISED LEARNING

Numerous SP applications, both traditional and de novo, involve classification and detection, e.g., various speech recognition tasks, music genre classification, entity recognition in video, emotion detection, and network traffic classification based on packet time series, to name just a few. Traditionally, these statistical classification applications have been treated as supervised learning problems, with the classifier designed using a training set of supervised (labeled) examples. However, in many domains, given pervasive sensing and massive data storage capabilities as well as large publicly accessible data repositories, it is both easy and inexpensive to collect a huge "training set" of examples; on the other hand, ground-truth labeling them is both enormously time-consuming as well as expensive, depending on the domain. This labeled/unlabeled data asymmetry motivates semisupervised learning techniques, which generally aim to enhance the (poor) classification performance achievable using a small (deficient) labeled training set by leveraging, for training purposes, many unlabeled samples.

Semisupervised techniques are either generative—modeling the joint density of the feature vector and class label—or discriminative—focusing solely on optimizing the class decision boundary. They may perform either inductive inference—imposing an explicit decision boundary on the feature space—or transductive inference, where the test set itself is effectively treated as part of the unlabeled set, used for joint semisupervised learning and inference. They may assume identical training and test set class distributions. On the other hand, an important recent trend is domain adaptation, where the test set distributions may be different and, thus, where recalibration of the classifier for the test set, albeit an ill-posed problem, may be required. Here a new (test) "domain" may imply a new sensing environment, i.e., changes in where or when data sensing occurs, relative to training. This may also correspond, e.g., to applying a speech recognition system trained on one population to a different population or a network traffic classifier trained on one local network to a different one. Semisupervised domain adaptation is a ubiquitous problem, with many potential (application-specific) factors that may contribute to statistical differences between the training and test domains.

An underlying theme in many recent MLSP approaches is the need to deal with multiple objectives, i.e., to de-emphasize the traditional optimality with respect to a single chosen metric. In addition to the focus on the traditional bias-variance dilemma—always emphasized in MLSP research so that methods will generalize well to unseen data—the set of objectives now also includes robustness, efficiency, and full interaction with the environment for a complete (global) performance assessment. All these considerations, among others, define the cognitive information processing paradigm, which is discussed next.

COGNITIVE INFORMATION PROCESSING

Artificial cognitive systems and cognitive information processing are emerging trends and will play an increasingly important role in MLSP in the coming years. The ability to perform cognitive information processing can be seen as a natural progression of MLSP, aiming to revitalize some of the original ideas of Alan Turing's "Theory of Computation" and Nobert Wiener's "Cybernetics" and those subsequently pioneered in the SP community by Bernard Widrow. The grand vision is to design and implement profound cognitive information processing systems for augmented human cognition in real-life environments. The practical imperative of this vision is driven by global megatrends related to pervasive and distributed computation, connectedness of people and systems, and pervasive digital sensing, which just a decade ago would have been impossible.

Cognitive information processing (CIP) involves the ability to perceive, learn, reason, and interact robustly in open-ended changing environments by integrating all available information—from multiple raw information sources and sensor inputs to user-driven feedback, annotations, and descriptions. We suggest using a tiered description of cognitive functionality: from low-level, simple sensing-action processing to high-level processing such as decision making and goal planning. There have been other suggestions setting out a minimal set of conditions for signifying processing as being "cognitive." In Simon Haykin's formulation, a cognitive information processing system would require the presence of four properties: 1) Perception-action cycle processing; 2) memory, to predict consequences in the environment; 3) an attention mechanism for allocation of resources; and 4) intelligence/reasoning for decision making in uncertain and complex environments. The important discussion aiming to fully formalize a definition of cognitive information processing and cognitive systems is ongoing; however, many concrete models, systems, and engineering solutions are already emerging.

Machine learning models that continuously learn from both data and previous knowledge will play an increasingly important and instrumental role in all levels of cognition in the real digital world that consists of large data sets, complex, distributed, interacting systems, and unknown, nonstationary environments—this is all usually too complex to be modeled within a limited set of predefined specifications. In real-life environments , there will be inevitably a need for CIP-based automated robust decisions and behaviors in novel situations, including the handling of conflicts and ambiguities. Hence there is a quest for dynamical learning systems, that continuously adapt to changing environments—one of the central components of machine learning for SP. Further, there is a need, beyond capabilities of current systems with built-in semantic representations, for automatic extraction and organization of meaning, purpose, and intention in interplay with the environment and with entities that include computers, embodied agents (i.e., humanlike artificial systems), and human users. In particular, interactive user systems (users-in-the-loop) models will be of vital importance.

Current examples of the use of machine learning in cognitive information processing include e.g., cognitive radio, personalized information systems, sensor network systems, social dynamics systems, semantic analysis systems, Web 2.0 and beyond, and cognitive components analysis. It is also obvious that the success of such approaches requires a multidisciplinary team effort with mixed competencies in engineering, computer science, statistics, machine learning, psychology, neuroscience, and specific domain knowledge.

TRENDS IN MLSP APPLICATIONS

The integrative nature of MLSP techniques has made them primary candidates for many of the emerging applications—a long list that includes brain-computer interface, behavior and emotion recognition, and learning in environments such as social networks and dynamic networks. Next, we discuss three such applications that have received particular attention within the community.

MULTISET DATA ANALYSIS AND MULTIMODALITY DATA FUSION

Analysis of multiple sets of data, either of the same type—multisubject data, data measured at different (time, space) points or under different environments—or of different types, as in multimodality data (e.g., audio and video, or different medical imaging data) is inherent to many problems in SP.

A good example is biomedical image analysis, which is especially challenging because of the rich nature of the data made available by various imaging modalities. Many biomedical studies collect multiple data sets, such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG), structural MRI (sMRI), and genetic data, in addition to clinical and behavioral data and other subject-based assessment parameters. Efficient use of all this information for inference, while minimizing assumptions made about the underlying nature of the data and relationships, is a difficult task, but one that promises significant gains, both scientific and, in the long run, societal, for challenging and important problems, such as the understanding of the human brain function. The need to minimize strong modeling assumptions is especially evident when studying brain function in natural states such as rest, or when performing tasks such as driving. Data-driven methods such as blind source separation and independent component analysis (ICA), which make minimal modeling assumptions on the data and the underlying processes, are particularly attractive in this context as they can achieve useful decompositions of the multimodal or multiset data without strong assumptions, and can also incorporate reliable prior information whenever available. Along with ICA-based techniques, other latent variable analysis techniques such as tensor decomposition are providing valuable tools for the analysis of multiset data and for fusion of multimodality information.

AUDIO AND MUSIC PROCESSING

Audio SP has always been a central part of SP, with many applications, ranging from sound recording and reproduction systems to advanced speech recognition. Machine learning has also been a central component when it comes to understanding and extracting audio information, even in spite of the fact that machine learning models and algorithms have often been developed without any special attention given to physical modeling of the production mechanisms for audio signals. Online streaming and distributions of audio, and in particular music, has opened a new avenue of possibilities for systems that enable interpretation (semantic audio), music organization, interaction, and sharing. This has indeed already revolutionized the way we consume music and in fact has created new global market opportunities. The special issue on music SP in *IEEE Journal of Selected Topics in Signal Processing* (fall 2011) set the stage for current activities in this field. A key component will be the interplay of SP, which enables the extraction of relevant features, and machine learning, which assists with interpretation and representation of results to users.

Slides

SYSTEMS BIOLOGY

At the beginning of the 21st century, the establishment of high-throughput screening methods and the completion of the human genome mapping marked the beginning of a new era for biological research. The contribution of SP to the acquisition and analysis of these data was crucial since biomolecular signals—e.g., microarray-based gene expression profiles, protein spots in gels, mass spectra, biomolecular images—had to be filtered, accurately detected, normalized, and analyzed. In addition to the SP, machine learning methods also began to be used to unravel the biological meaning of the signals and to categorize the evidence in meaningful ways. It was at that stage that well-established supervised and unsupervised machine learning methods started to provide useful answers to even clinically relevant questions, such as finding gene expression profiles that could be used as biomarkers for certain types of leukemia.

Today, about a decade later, despite the continual development of high throughput methods, producing terabytes of data on a daily basis, many important biological questions still cannot be well addressed, and it is widely accepted that bioinformatics data analysis alone is not sufficient to capture the dynamics and emerging properties of living cells, tissues, and organisms. A new field is thus rapidly emerging, that of systems biology, with the main objective to integrate all qualitative and quantitative biological knowledge, extracted either by biological research or analysis, within holistic and useful models that can capture biological system dynamics at different scales (cell, tissue, whole organism) but also across multiple scales.

Hence, although grounded in biology, biochemistry, and mathematics, the contribution of informatics and especially machine learning in systems biology is more requisite than ever. To build integrative dynamical models while extracting the network of pairwise interactions among molecular species and their possibly causal relations, we need powerful MLSP methods to discover as many true interactions as available data sets (of given size) may allow. A challenging problem systems biology is facing today in many different contexts is the joint estimation of parameters and model structures from sparse and noisy time-course data. Although some of the most sophisticated inference methods have already been tried, we are not yet able to train models of sufficient size, commensurate with the (large) number of molecular units. Many derived models thus have to be tuned either exclusively manually or only partially through parameter estimation methods. It is evident that more efficient MLSP methods are needed to learn, from the available data, dynamical system models of high complexity and accuracy, able to be used in realistic simulations for in-silico experimentation, leading to formation of new hypotheses that could drive new biological research.

ACKNOWLEDGMENTS

We would like to thank all the past and present members of the MLSP/NNSP TC who provided feedback and participated in the discussion on "Trends in MLSP." In particular, we would like to acknowledge the feedback from Sergios Theodoridis, Elias Manolakos (on systems biology), Jeronimo Arenas-Garcia, Gustavo Camps-Valls, and Ignacio Santamaria.

AUTHORS

Tülay Adalı (adali@umbc.edu) is a professor at the University of Maryland, Baltimore County.

David J. Miller (djmiller@engr.psu.edu) is a professor at the Pennsylvania State University.

Konstantinos I. Diamantaras (kdiamant@it.teithe.gr) is a professor at the TEI of Thessaloniki.

Jan Larsen (jl@imm.dtu.dk) is an associate professor at the Technical University of Denmark.

Video