

## Guest Editorial

**O. Erhun Kundakcioglu · Marcello Sanguineti ·  
Theodore B. Trafalis**

© Springer-Verlag 2008

In the last few decades, the analysis of data has become a crucial task which helps shape nearly every aspect of modern life. From a research perspective, data has become a critical characteristic of our world. We witness the ubiquity of data analysis and management through numerous disciplines such as biology, finance, manufacturing, medicine, and communications. On one side, numerous applications produce data that have to be processed; on the other side, many applications are based on data that have already been processed. We are living in a society that is a huge data-producer and a huge data-consumer at the same time.

Nowadays, learning from data is a melting pot involving scientists from a large variety of areas such as Computer Science, Operations Research, Engineering, Mathematics, Physics, Statistics, Biology, and Medicine. The multidisciplinary approach to data analysis provides challenging data mining problems and promotes innovative solution methodologies. Some major underlying themes of modern data-driven and data-producing applications are:

---

O. E. Kundakcioglu  
Department of Industrial and Systems Engineering, University of Florida,  
Gainesville, FL 32611, USA  
e-mail: erhun@ufl.edu

M. Sanguineti  
Department of Communications, Computer and System Sciences,  
University of Genova, 16145 Genova, Italy  
e-mail: marcello@dist.unige.it

T. B. Trafalis (✉)  
School of Industrial Engineering, The University of Oklahoma,  
Norman, OK 73019, USA  
e-mail: ttrafal@ou.edu

- data richness (elaborate and/or produce huge amounts of data);
- on-line adaptation (require data-processing algorithms that must be quickly updated as soon as new observations become available);
- high-dimensionality (each single datum often has a large dimension);
- structured nature (some types of data, e.g., images, bio-sequences, and weighted graphs, cannot be represented simply as vectors of real numbers).

Classical statistical methods are not designed to cope with problems exhibiting such features. On the other hand, the last few decades have seen major advances in optimization: the development of new mathematical tools has kept up with the availability of powerful modeling languages and the dramatic improvement of computational capabilities. New methods of learning from data can strongly benefit from optimization-based approaches and can exploit them to deal with the data types arising in different areas of science and technology of major interest.

The articles in this special issue highlight the role that can be played by optimization theory and techniques in developing powerful solution methodologies and efficient algorithms for learning from data. These articles also point out challenging problems from data-intensive applications, which can be attacked via optimization tools.

The first paper, by Angelo Alessandri, Lucia Cassettari, and Roberto Mosca, addresses the estimation of nonparametric models using a finite data sample. The task consists of finding a nonparametric mapping that interpolates the data according to the empirical-risk-minimization approach. The paper compares, in terms of complexity and experimental performances, the solution provided by classical regression approaches, based on polynomial fitting, with the solution given by one-hidden-layer feedforward neural networks. Two simulation case studies are analyzed to understand the practical issues that arise in solving such problems.

The second paper is by Augusto Destrero, Sofia Mosci, Christine De Mol, Alessandro Verri, and Francesca Odone. The authors focus on feature selection methods for problems with high-dimensional data. They present a regularized approach to feature selection, which looks for the solution to least squares problems with sparsity-enforcing constraints. They discuss the benefits of adopting a regularized approach with  $L_1$  or  $L_1 - L_2$  penalties in two applications: microarray data analysis in computational biology and object detection in computer vision. They describe general algorithmic aspects, adapt them to the two above-mentioned case studies, and discuss architecture issues specific to these two domains.

In the third contribution, Peng Du, Jiming Peng, and Tamás Terlaky propose a bi-level optimization formulation for the problems of model selection and feature selection in Support Vector Machines (SVMs). The idea exploited to select the best model consists in embedding the standard convex quadratic problem of SVM training into a lower-level problem in a bi-level optimization model. The objective of the quadratic SVM problem is minimized over the feasible range of the kernel parameters at the upper level of the bi-level model. Feature selection is handled in the same framework, extending the kernel functions by the introduction of independent kernel parameters for all the features in the original space. Experimental results for both model and feature selection are presented.

The fourth paper, by Giorgio Gnecco and Marcello Sanguineti, deals with the weight-decay technique, introduced into learning to improve the generalization capability of connectionistic models expressed as a linear combination of a set of computational units. Weight decay is investigated using tools from regularization theory and is compared with Tikhonov's regularization of the learning problem and with a mixed regularized technique. The learning techniques are compared in terms of filtering factors and accuracy of suboptimal solutions. The accuracy is estimated for connectionistic models with a-priori fixed numbers of computational units and flexible choices of the parameters.

The fifth paper is by Olutayo O. Oladunni and Theodore B. Trafalis. The authors develop a nonlinear kernel approach based on the Tikhonov's regularization scheme for knowledge-based multi-classification discrimination. The linear multi-classification Tikhonov's-regularization knowledge-based Support Vector Machine (SVM) model is used. A kernel function is employed to find a nonlinear classifier capable of discriminating future points into an appropriate class. As to the prior knowledge, it is expressed in the form of multiple polyhedral sets belonging to one or more categories or classes, and is introduced as additional constraints into the formulation of the regularized nonlinear kernel least-squares multi-class SVM model. Data and prior knowledge from the two-phase flow regimes in pipes are used to train and test the proposed formulation.

In the sixth contribution, by Carlotta Orsenigo and Carlo Vercellis, the extension of Discrete Support Vector Machines (DSVM) to multicategory classification is addressed by three different methods. In two of them, the authors frame binary perceptrons based on a variant of DSVM, within one-against-all and round-robin schemes, respectively. The third method directly addresses multicategory discrimination by building a dichotomic decision tree. In such a tree, an optimal split to separate classes is derived at each node by a new extended formulation of DSVM. Computational experiments are performed on publicly available datasets.