Soft Computing (2021) 25:12565–12567
https://doi.org/10.1007/s00500-021-06144-y

EDITORIAL

Special Issue Optimization for Machine Learning Guest Editorial

Marcello Sanguineti[1] · Elena Tànfani[2]

Published online: 28 August 2021

Since the very beginning, there has been a fruitful exchange between machine learning (ML) and optimization. While ML exploits optimization models and algorithms, it simultaneously poses problems that often constitute optimization challenges. This cross-fertilization is particularly evident nowadays. Many applications produce data that have to be processed, and many other applications are based on data that have already been processed. In society, data producers and data consumers continuously exchange their roles. At the same time, scientists generate and collect huge amounts of data and need to develop methodologies for data analysis. The interplay between ML and optimization continues to develop, and its benefits are there for all to see. This requires the use of ever more powerful approaches and techniques. Among such improvements, those coming from the field of operations research (OR) play a basic role. Thus, machine learning has become a very active research area in OR.

This special issue is associated with the 2019 International Conference on Optimization and Decision Science (ODS), which hosted the XLIX Annual Conference of the Italian Operations Research Society (AIRO). The ODS conference aims at being an opportunity for researchers and practitioners from various fields (applied mathematics, computer science, engineering, economics), private and public companies, industries, and policy makers, to present and share ideas, experiences, and knowledge, as well as to create a point of contact among them. The 2019 edition of ODS was held in Genova, Italy, and was organized by AIRO, DIEC (Department of Economics and Business Studies), DIBRIS (Department of Informatics, Bioengineering, Robotics, and Systems Engineering), and DIMA (Departments of Mathematics) of the University of Genoa and attracted a total of 172 presentations (52 short papers and 120 abstracts).

This special issue was launched to welcome both theoretical investigations and algorithmic developments on optimization in machine learning, as well as contributions on practical implementations and numerical studies. It consists of a selection of papers that present innovative results in classical optimization frameworks, advances in more sophisticated models and techniques, and new optimization-based learning paradigms. They contribute insights into the understanding of advanced modelling techniques and new optimization-based learning paradigms.

We hope that this special issue will stimulate interaction among scientists from optimization and machine learning and promote fruitful scientific collaborations, thus improving the applicability of optimization methods to challenging problems in learning from data.

Finally, we would like to express our thanks to the referees for their careful work and to the authors who submitted their research results.

To provide readers a quick overview of the articles collected, a brief summary for each of them is presented below.

The work "Deep reinforcement learning for multi-objective placement of virtual machines in cloud data centers", by Luca Caviglione, Mauro Gaggero, Massimo Paolucci, and Robert Ronco, addresses the problem, arising in cloud computing, of designing suitable management policies to face the workload while guaranteeing quality constraints and mitigating costs. Virtual Machines (VMs) placement is an interplay of different objectives, constraints, and technological domains. This problem can benefit from machine learning techniques, thanks to their capability of finding "hidden" relationships among the available data and therefore generate placement actions. To

✉ Marcello Sanguineti
marcello.sanguineti@unige.it

Elena Tànfani
etanfani@economia.unige.it

[1] Department of Informatics, Bioengineering, Robotics, and Systems Engineering (DIBRIS), University of Genova, Genova, Italy

[2] Department of Economics and Business Studies (DIEC), University of Genova, Genova, Italy

address the trade-off between the used power and the adherence to a service-level metric subscribed by customers, the authors propose an optimization-based placement mechanism to select the servers where to deploy VMs. The proposed mechanism for VM placement is based on Deep Reinforcement Learning (DRL). A decision maker is considered that, after proper training, is able to select the most suitable heuristic for computing the placement for each VM requested by end users. The effectiveness of the approach is evaluated via a comparison against solutions widely adopted in the literature and real-world scenarios, including the use of workload traces collected in a production-quality cloud datacenter.

In "Green machine learning via augmented Gaussian processes and multi-information source optimization", by Antonio Candelieri, Riccardo Perego, and Francesco Archetti, the problem of Hyper-Parameter Optimization (HPO) is addressed. The problem can be regarded as an optimization outer loop on the top of ML model learning (inner loop). Hyper-parameters are all the parameters of a ML model which are not updated during the learning and are used to configure either the model (e.g. the number of layers of a deep network) or characterize the algorithm used in the training phase (e.g. the learning rate). Sometimes, they also include the choice of optimization algorithm itself and the data features which are fed into the model. The availability of various information sources, with different computational costs and different "fidelity", is exploited. The multisource optimization strategy adopted in the paper fits into the scheme of Gaussian process-based Bayesian optimization. An Augmented Gaussian Process Method exploiting Multiple Information SOurces (AGP-MISO) is proposed, where the process is trained using only "reliable" information among available sources. The authors report computational results related to the HPO of a support vector machine classifier using two sources: a large dataset—the most expensive one—and a smaller portion of it.

The paper "Polyhedral Separation via Difference of Convex (DC) programming", by Annabella Astorino, Massimo Di Francesco, Manlio Gaudioso, Enrico Gorgone, and Benedetto Manca, deals with binary classification based on the use of a polyhedral surface. In particular, they address the numerical treatment of the optimization problem to be solved in order to get a polyhedral separation surface. To this end, DC nature of the objective function is exploited and an algorithm designed to treat DC functions is adopted, called DCA. A decomposition of the error function for the polyhedral separability problem as the difference of two convex functions is used. The DCA algorithm is then applied to carry out an extensive experimentation on several classes of benchmark instances able to show the good performance of the approach both in terms of classification correctness and computation time.

In the article "On the convergence of a block-coordinate incremental gradient method", Laura Palagi and Ruggiero Seccia investigate the convergence of a Block-coordinate Incremental Gradient (BIG) method that arises in ML, for instance in regularized empirical risk minimization. The proposed algorithm generalizes a Block Layer Incremental Gradient (BLInG) algorithm, available in the literature to train deep networks by exploiting their layered structure, which uses an incremental approach for updating the weights over each single layer. Under some assumptions on the objective function, they prove that BIG method can be seen as a gradient method with errors. Its convergence is investigated by showing that the error at each iteration satisfies some standard conditions. Taking the hint from some available deterministic convergence results for gradient methods with errors, they prove convergence of BIG towards stationary points and to an ε-approximate solution, respectively, when a diminishing and a bounded away from zero stepsizes are employed.

In the work "On the enumeration of Boolean functions with distinguished variables", by Josep Freixas, insights into ML models are obtained by studying the problem of enumerating Boolean functions and some subclasses of them. The study of Boolean functions and, in particular, their enumerations contribute to the theoretical development of neural networks and are useful for the design of circuits and real-world voting systems that fulfil some desirable properties. Since such enumeration problem is highly complex, it is of interest to study subclasses that escape this limitation and can be enumerated by means of sequences depending on the number of variables. In the paper, new formulas are proven corresponding to enumerations of some subclasses of Boolean functions. The versatility of these functions makes the problem of interest in several different fields such as game theory, hypergraphs, reliability, cryptography, and logic gates. Inductive reasoning, which plays a prominent role in soft computing, is the main method used to achieve the results.

The paper "Correlations of random classifiers on large data sets", by Věra Kůrková and Marcello Sanguineti, investigates capabilities of efficient approximation of randomly chosen functions by some classes of feedforward neural networks. The authors explore approximation errors measured in the l_2-norm in terms of correlations of random classifiers with input–output neural mapping. A probabilistic model of relevance of computational tasks is proposed, which includes distributions that may not satisfy the naive Bayes assumption (i.e. when classes are not assigned to network inputs independently). Effects of increasing sizes of sets of data to be classified are analysed by exploiting geometrical properties of high-dimensional spaces. The authors prove that on large domains, correlations of randomly chosen classifiers with any fixed binary-valued function are concentrated around their mean value. They show that the critical factor for suitability of a class of networks for computing randomly chosen classifiers is the maximum of sizes of the mean values of their correlations with network input–output functions.

Finally, in the work "Optimal trade-off between sample size, precision of supervision, and selection probabilities for the unbalanced fixed effects panel data model", Giorgio Gnecco, Federico Nutarelli, and Daniela Selvi analyse the optimal trade-off between sample size, precision of supervision, and selection probabilities on a linear model of the input-output relationship, namely the unbalanced fixed effects panel data model. In the fixed effects model, applied in the econometric analysis of microeconomic and macroeconomic data, biostatistics, educational research, engineering, neuroscience, political science, and sociology, the observations related to different observational units (individuals) are associated with possibly different constants, which are able to represent unobserved heterogeneity in the data. The authors address the case in which the model includes the additional possibility of controlling the conditional variance of the output given the input and the selection probabilities of the different units per unit time. An optimization problem is formulated, based on a large-sample upper bound on the generalization error associated with the estimates of the parameters of the unbalanced fixed effects panel data model, conditioned on the training input dataset. It is shown that under suitable assumptions, in some cases "many but bad" examples provide a smaller large-sample upper bound on the conditional generalization error than "few but good" ones, whereas in other cases the opposite occurs.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.