

## Editorial of the special issue data analysis and intelligent optimization with applications

Vadim Strijov<sup>1</sup>  · Gerhard-Wilhelm Weber<sup>2</sup> ·  
Richard Weber<sup>3</sup> · Süreyya Ozogur Akyüz<sup>4</sup>

Received: 30 June 2015 / Accepted: 2 July 2015 / Published online: 4 September 2015  
© The Author(s) 2015

This special issue on “Data Analysis and Intelligent Optimization with Applications” follows a previous special issue of this journal on the interplay of Machine Learning and Optimization, “Model Selection and Optimization in ML” (*Machine Learning* 85:1-2, October 2011). This time we shift our focus to applications of data analysis and optimization techniques. Optimization problems underlie most machine learning approaches. Due to emergence of new practical applications, new problems and challenges for traditional approaches arise. Emergent applications generate new data analysis problems, which, in turn boost new research in optimization. The contribution of machine learning researchers into the field of optimization is of considerable significance and should not be overlooked. This special issue collected solutions, adapted for real world problems, leading to massive and large-scale data sets, online data and imbalanced data. We encouraged submission of papers, devoted to combining machine learning and data analysis techniques with advances in optimization to produce methods of Intelligent Optimization, both theoretical and practical. Our goal for this special issue was to bring together researchers working in different areas, related to analytics and optimization. The special issue includes papers on a variety of topics, including, but not limited to

---

✉ Vadim Strijov  
strijov@gmail.com; strijov@ccas.ru

Gerhard-Wilhelm Weber  
gweber@metu.edu.tr

Richard Weber  
rweber@dii.uchile.cl

Süreyya Ozogur Akyüz  
sureyya.akyuz@eng.bahcesehir.edu.tr

<sup>1</sup> Computing Center of the Russian Academy of Sciences, Vavilova 42, Moscow, Russia 119333

<sup>2</sup> Institute of Applied Mathematics, Middle East Technical University, 06800 Ankara, Turkey

<sup>3</sup> Department of Industrial Engineering, FCFM, Universidad de Chile, Santiago, Chile

<sup>4</sup> Department of Biomedical Engineering, Faculty of Engineering and Natural Sciences, Bahcesehir University, 34353 Istanbul, Turkey

- Intelligent Optimization,
- Model Selection and Model Complexity,
- Predictive Modeling and Optimization,
- Computational Complexity and Optimization,
- Pattern Recognition and Classification,
- Regression and Forecasting,
- Models of Data and Measurements,
- Feature Selection,
- Soft Computing,
- Learning Theory,
- Treatment of Imbalanced Data Sets

with applications in computer vision, health care, information technologies, economics, finance and risk analysis, text mining, information retrieval, and medicine.

Each of the papers was reviewed by three experts in the area of the presented study. The candidates for acceptance had to undergo one or more thorough revisions. In the remainder of this editorial we briefly describe the contribution made by each accepted paper.

*Stochastic Feature Mapping for PAC Bayes Classification* by Xiong Li, Bin Wang, Yuncai Liu and Tai Sing Lee brings together generative and discriminative models. The authors propose a coupling mechanism developed under the PAC-Bayes framework that can fine-tune the generative models and the feature mapping function iteratively to improve the classifier's performance.

*Model Selection in Multivariate Adaptive Regression Splines Using Information Complexity as the Fitness Function* by Elcin Kartal Koc and Hamparsum Bozdogan introduces information-theoretic measure of complexity criterion for model selection in Multivariate Adaptive Regression Splines to tradeoff efficiently between how well the model fits the data and the model complexity. The authors consider the model complexity with respect to the interdependency of parameter estimates, as well as the number of free parameters in the model.

*Analysis of Network Traffic Features for Anomaly Detection* by Felix Iglesias and Tanja Zseby addresses the feature selection problem for network traffic based anomaly detection. The authors propose a multi-stage feature selection method using filters and stepwise regression wrappers, which allowed them to eliminate 13 very costly features and thus reduce the computational effort for on-line feature generation from live traffic observations at network nodes.

*Measuring the Accuracy of Currency Crisis Prediction with Combined Classifiers in Designing Early Warning System* by Nor Azuana Ramli, Mohd Tahir Ismail and Hooy Chee Wooi addresses the question whether the prediction accuracy is affected by the method used in the ensemble of the classifiers. The authors investigate experimentally combinations of classifiers, such as support vector machine with k-nearest neighbor, logistic regression with k-nearest neighbor and LADTree with k-nearest neighbor, for predicting currency crisis of 25 countries.

*Probabilistic combination of classification rules and its application to medical diagnosis* by Jakub M. Tomczak and Maciej Zięba proposes the combination of soft rules in the application to the medical domain. The approach relies on probabilistic decision making with latent relationships among features represented by the conjunctive features and a new manner of estimating probabilities in case of imbalanced data problem. Additionally, the authors propose a method for aggregating sufficient statistics needed to estimate probabilities in a graph-based structure to speed up computations.

*Selective Switching Mechanism in Virtual Machines via Support Vector Machines and Transfer Learning* by Wei Kuang, Laura E. Brown and Zhenlin Wang proposes to learn a decision model for virtual machine paging mode switching between shadow paging (SP) and hardware-assisted paging (HAP). The authors conduct several experiments to test the performance of an SVM-based Adaptive Switching mechanism and demonstrate that if there is a significant gap between HAP and SP, the proposed Adaptive Switching mechanism can match the better one.

*Feature Selection in machine learning: an exact penalty approach using a Different of Convex function Algorithm* by Le Thi Hoai An, Le Hoai Minh and Pham Dinh Tao develops an exact penalty approach for feature selection in machine learning via the zero-norm-regularization problem, which allows to consider all the existing convex and non-convex approximation approaches to treat the zero-norm in a unified view within DC programming and DCA framework.

*Day trading profit maximization with multi-task learning and technical analysis* by Zsolt Bitvai and Trevor Cohn aims to demonstrate that stock market price movements are predictable, and patterns of market movements can be exploited to realize excess profits over passive trading strategies. The authors show this empirically by developing a novel stochastic trading algorithm in the form of a linear model with a profit maximization objective. Using this method the authors demonstrate improvements over the competitive buy-and-hold baseline over a decade of stock market data for several companies.

*A Computational Approach to Nonparametric Regression: Bootstrapping CMARS Method* by Ceyda Yazic, Fatma Yerlikaya-Özkurt and İnci Batmaz aims to reduce the complexity of conic multivariate adaptive regression splines models without degrading its performance. The authors apply three different bootstrap methods to CMARS, empirically evaluate their performance and compare these methods with respect to criteria such as accuracy, complexity, stability and robustness, using four data sets that represent the small and medium sample size and scales.

The paper *Committee Polyhedral Separability. Complexity and Polynomial Approximation* by Michael Khachay addresses the approximability issues of the MASC-GP(n) NP-hard combinatorial optimization problem formalizing the machine learning strategy based on the structural risk minimization principle in the class of majority voting piecewise linear classifiers. Along with its contribution to computational complexity theory, the paper provides a new approach to efficient learning for ensembles of affine classifiers having high generalization ability.

*Pruning of Error Correcting Output Codes by Optimization of Accuracy-Diversity Trade off* by Sureyya Ozogur, Terry Windeatt and Raymond Smith proposes a new algorithm that prunes the set of base classifiers by optimizing the accuracy and diversity simultaneously with the proposed cost function. The new algorithms result in a fast and more efficient pruning method for Error Correcting Output Codes.

*Triadic Formal Concept Analysis and Triclustering: Searching for Optimal Patterns* by Dmitry I. Ignatov, Dmitry V. Gnatyshak, Sergei O. Kuznetsov and Boris G. Mirkin presents several definitions of “optimal patterns” in triadic data and results of experimental comparison of five triclustering algorithms on real-world and synthetic datasets. The multicriteria choice allows an expert to decide which of the criteria are most important in a specific case and make a choice. The experimentation on both synthetic and real data shows that there is no one winning method according to the introduced criteria.

*Additive Regularization of Topic Models* by Konstantin Vorontsov and Anna Potapenko introduces a new semi-probabilistic approach to topic modeling. Instead of building a purely probabilistic generative model of text the authors regularize an ill-posed problem of sto-

chastic matrix factorization by maximizing the weighted sum of the log-likelihood and additional criteria. This approach allows to combine probabilistic assumptions with linguistic and problem-specific requirements in a single multi-objective topic model.

*Two Level Quantile Regression Forests for Bias Correction in Range Prediction* by Nguyen Thanh Tung, Thuy Thi Nguyen and Joshua Z. Huang proposes a new bias correction method, called bcQRF that uses bias correction in Quantile Regression Forests (QRF) for range prediction. The authors propose a new feature weighting subspace sampling method to build the first level QRF model. The residual term of the first level QRF model is then used as the response feature to train the second level QRF model for bias correction.

*Random Drift Particle Swarm Optimization Algorithm: Convergence Analysis and Parameter Selection* by Jun Sun, Xiaojun Wu, Vasile Palade, Wei Fang and Yuhui Shi provides systematical analysis and empirical study of the random drift particle swarm optimization algorithm, inspired by the free electron model in metal conductors placed in an external electric field.

*Using Causal Discovery for Feature Selection in Multivariate Numerical Time Series* by Youqiang Sun, Jiuyong Li, Jixue Liu, Christopher Chow, Bingyu Sun and Rujing Wang addresses feature selection in multivariate time series. The authors present a method for causal features identification with effective sliding window sizes in multivariate numerical time series based on Granger causality discovery. The proposed method considers the influence of lagged observations of features on the target time series.

*An incremental piecewise linear classifier based on polyhedral conic separation* by Gurkan Ozturk, Adil M. Bagirov and Refail Kasimbeyli develops a piecewise linear classifier based on polyhedral conic separation. The proposed classifier depends on few polyhedral conic functions and therefore requires low memory. The authors propose a special procedure of generating starting points based on the incremental approach to minimize the error function.

*DEARank: A Data-envelopment-analysis-based Ranking Method* by Chunheng Jiang and Wenbin Lin introduces data envelopment analysis into the field of learning to rank and propose DEARank algorithm. Making use of DEA's powerful potential in capturing the intrinsic characteristics of documents, the authors construct the weak ranker candidates using the optimal weights of features for units.

**Acknowledgments** The Guest Editors are grateful to Peter Flach, the Editor-in-Chief of the Machine Learning Journal, and Dragos Margineantu, Editor for Special Issues, for encouraging this special issue and for their guidance, help and close advice at every stage.