Editorial

Special Issue on Business Analytics and Intelligent Optimization

During the last decades, the disciplines of Data Mining and Operations Research have been working mostly independent of each other. The increasing complexity of today's applications in areas such as business, medicine, and science requires, however, more and more interaction between both disciplines. On the one hand, many data mining algorithms rely on optimization methods, and incorporating state-of-the-art optimization techniques may improve the efficiency and effectiveness of data mining algorithms. On the other hand, there are many optimization problems that would benefit from an integration of data mining and knowledge discovery processes (KDD). Furthermore, the combined use of both disciplines would provide superior solutions in many application domains, particularly in an attempt to take into account the entire decision making process.

This special issue presents several works from the area of Business Analytics and Intelligent Optimization thus highlighting some recent developments in this fast evolving research field and at the same time showing the need for future research.

Verbraken *et al.* analyze the phenomenon of customer churn in the telecommunication industry. They compare different Bayesian network classifiers combined with a feature selection method based on the Markov Blanket. Model selection is performed by classical approaches, such as e.g. the Area Under the Curve (AUC) as well as by using the recently proposed Maximum Profit Criterion, which optimizes not just a statistical measure but the profit generated by the particular retention campaign. This profit is calculated as function of the fraction of customers to be contacted. The authors suggest optimizing this fraction to determine the best retention campaign for each model, respectively. Results show that the Bayesian network models provided better insights into the churn phenomenon than a classical logistic regression approach.

Wilcox *et al.* contribute to the development of automated, fast, and cost-effective systems for animal feed quality monitoring. Data acquired from the respective hyperspectral image (HSI) cubes are classified by a Support Vector Machine (SVM) into the different classes. Their original contributions are two methods for selecting spectral band sets, one based on an evolutionary search method and another one using a divergence-based recursive feature elimination approach. They show that both feature selection techniques provide a significant advantage for selecting small spectral band sets over the respective traditional method.

The application area studied by Román *et al.* is web mining, in particular the identification of user sessions. They propose Mixed Integer Programming problems (MIP) to reconstruct sessions using just data from a web log file; in particular they present a Sessionization Integer Programming problem (SIP) and an optimization model that is a novel use of the well-known Bipartite Cardinality Matching (BCM) problem. On data generated by visits to a web site over 15 months they could show that their models outperform the traditionally used time-out heuristic and also gave more insights into visitor behavior.

Jovanovic *et al.* present a methodological work which is motivated by the difficulties of constructing suitable classifiers for a given data mining task. They point out that manual model selection is a time-consuming and complex task and develop a framework to automatically design component-based decision tree algorithms using an evolutionary algorithm. Test results from two experimental studies underline the potential of the proposed framework. The authors provide a list of future research challenges regarding their work thus inspiring readers to contribute to this interesting research direction.

Taylan *et al.* study the outlier problem in linear regression where they propose a Tikhonov regularization problem and treat it using convex optimization techniques. They employ conic quadratic programming which permits the use of interior point methods. The new method is called CTMSOM (Mean Shift Outlier Model based the Conic Programming and Tikhonov Regularization) which outperformed a linear model (LM) and a traditional Mean Shift Outlier Model (MSOM) on seven data sets.

Maldonado and Montecinos develop multi-classifiers based on Support Vector Machines (SVM) for robust classification of imbalanced data sets. Motivated by the problem of credit card customer churn prediction they propose a framework combining several sampling techniques with ensembles of different classifiers improving predictive performance and gaining insight into the problem of classifying imbalanced data sets. These effects are shown for several artificial data sets and a real-world set from a financial entity. Finally, they identify several venues for future research in classification of imbalanced data sets.

We hope the readers of this special issue find interesting results they could apply in their respective area and/or get inspired for their own future research thus contributing to a continuous development of ideas presented here.

The editors would like to thank the Editor-in-Chief, Professor Fazel Famili, IOS Press, and the authors for their work. Special thanks go to the anonymous reviewers for their excellent collaboration.

Kate Smith-Miles and Richard Weber[1]