

#### 4 - Support of managerial decision making processes by transductive learning

Hubertus Brandner, Universität Hamburg, Germany,  
hubertus.brandner@studium.uni-hamburg.de, Stefan Lessmann

This study analyses to which extent the promising findings of transductive approaches can be transferred to business problems of classification. Different variants of Support Vector Machines are examined to compare the established inductive learning and the transductive technique. To that end a hybrid metaheuristic is implemented to solve the mathematical programming formulations in the same way. Empirical results confirm the potential of transductive inference. Therefore it is advisable to utilize the information of unlabeled data in the context of managerial decision making and planning.

#### WA-21

Wednesday, 10:40-12h00  
6.2.47

### Optimization Algorithms I

Stream: Software for OR/MS

Invited session

Chair: *Simone Garatti*, Dept. of Electronics and Information, Politecnico di Milano, p.zza L. da Vinci 32, 20133, Milan, Italy, sgaratti@elet.polimi.it

#### 1 - Solving uncertain programs via the scenario approach: the FAST algorithm

*Simone Garatti*, Dept. of Electronics and Information, Politecnico di Milano, p.zza L. da Vinci 32, 20133, Milan, Italy, sgaratti@elet.polimi.it, *Algo Carè*, *Marco Campi*

Uncertainty is ubiquitous in decision problems, and this leads naturally to uncertain programs (UP). Robust and chance-constrained solutions to UP can be difficult to obtain in general. In this talk, we discuss the use of the scenario approach, a handy methodology based on random sampling of constraints, to solve UP with a guaranteed degree of approximation. In particular, we introduce FAST (Fast Algorithm for the Scenario Technology), a variant of the standard scenario algorithm with reduced complexity, which improves the applicability of the scenario methodology to a high extent.

#### 2 - Scheduling optimization in virtual enterprises based on the hybridization of a CSP with a genetic algorithm

*Rabah Kassa*, mathematique, Universite Bejaia algerie, universit? de b?jaia 06000 bejaia algerie, 06000, bejaia, Algeria, rabah\_kassa2002@yahoo.fr, *Djamila Boukredera*, *Zaidi Sahnoun*

Production scheduling represents an important manufacturing function whose quality remains an essential stake for virtual enterprises. To optimize its scheduling, a virtual enterprise aims to improve its profitability while minimizing the customer's service costs and respecting manufacturing constraints. This can be formulated as a CSP. We suggest an optimization method of the CSP based on the genetic algorithm. This hybridization aim at better taking over of this kind of problem defined by a large research space and a complex constraint set and finds solutions of good quality.

#### 3 - ParadisEO: a framework for metaheuristics

*El-ghazali Talbi*, University of Lille - INRIA - CNRS, Lille, El-ghazali.Talbi@lifl.fr

We present the ParadisEO white-box object-oriented framework dedicated to the reusable design of metaheuristics. It provides a broad range of features including population based metaheuristics and single-solution metaheuristics. It bases on a conceptual separation of the solution methods from the problems they are intended to solve. The fine-grained nature of the classes allows a high flexibility. ParadisEO is of the rare frameworks providing most common parallel and distributed models; implementation is portable and models can be exploited transparently.

#### WA-23

Wednesday, 10:40-12h00  
6.2.49

### Model Selection in Regression Analysis

Stream: Data Mining in the Financial Sector

Invited session

Chair: *Michael Khachay*, Ural Branch of RAS, Institute of Mathematics and Mechanics, S.Kovalevskoy, 16, 620990, Ekaterinburg, Russian Federation, mkhachay@imm.uran.ru

Chair: *Vadim Strijov*, Computing Center of the Russian Academy of Sciences, Klara Zetkin 13-79A, 127299, Moscow, Russian Federation, strijov@ccas.ru

#### 1 - Model generation and model selection in credit scoring

*Vadim Strijov*, Computing Center of the Russian Academy of Sciences, Klara Zetkin 13-79A, 127299, Moscow, Russian Federation, strijov@ccas.ru

The credit scorecard is the logistic regression model; it maps the feature space to the probability of default of a banking client. A classical scorecard is constructed by an analyst, who manually selects informative features and creates combinations of them. We propose a new technique for the automatic scorecard construction. To develop a scorecard, one must assign a set of primitive functions and model generation rules. The result model is an admissible superposition of the primitive functions and features. The coherent Bayesian inference is used to select features and their superpositions.

#### 2 - Algorithms of feature selection for volatility estimation of European options

*Ekaterina Krymova*, Control/Management and Applied Mathematics, Moscow Institute of Physics and Technology, 9,35 b.3, Nagornaya st., Moscow, 141981, Moscow region, Dubna, Bogolubova 33-304, 117186, Moscow, Russian Federation, ekkrym@gmail.com

The problem of multicollinearity is commonly encountered in regression analysis. This problem may lead to overfitting and result in unstable model parameters. New approach to the feature generation and feature selection was proposed. The feature generation technique is based on Kolmogorov-Gabor polynomial construction. The features are superpositions of primitive functions and free variables. The generated features require reduction of multicollinearity. For this purpose, the LARS modification is developed. Historical data of European options is used as practical example.

#### 3 - A topological approach to formulating conditions of the uniform convergence of frequencies to probabilities

*Michael Khachay*, Ural Branch of RAS, Institute of Mathematics and Mechanics, S.Kovalevskoy, 16, 620990, Ekaterinburg, Russian Federation, mkhachay@imm.uran.ru

Existence of the uniform convergence of frequencies to probabilities over an appropriate events class is a well known sufficient consistency condition of the empirical risk minimization (ERM) in machine learning. The traditional approach for proving such convergence is based on a sublinear growth of entropy of the event class in question and obtaining upper VCD bounds for this class. In this paper, existence of the uniform convergence of frequencies to probabilities over an event class is related to some topological properties of the sigma-algebra, induced by this class.

#### 4 - Benchmarking Framework for Financial Text Mining

*Caslav Bozic*, Institute AIFB, IME Graduate School, Karlsruhe Institute of Technology (KIT), Institute AIFB - 05.20, KIT Campus South, 76128, Karlsruhe, BW, Germany, bozic@kit.edu

Different data mining methods for financial text and various sentiment measures are described in the existing literature, without common benchmark for comparing these approaches. Implemented system (which is a part of FINDS Project) and proposed framework are based on theoretical data integration, and they facilitate combining more sources of financial data into comprehensive integral dataset. The dataset is then used to analyse the candidate measure by regressing it on different returns and other financial indicators that can be defined using the system's novel data transformation approach.