Foundations of Computational Intelligence
Volume 5: Function Approximation and Classification

Approximation theory is that area of analysis which is concerned with the ability to approximate functions by simpler and more easily calculated functions. It is an area which, like many other fields of analysis, has its primary roots in the mathematics.The need for function approximation and classification arises in many branches of applied mathematics, computer science and data mining in particular.

Data mining, the extraction of hidden predictive information from large volume of data, is a powerful methodology with great potential in several aspects of our daily life. Data mining tools helps to understand hidden relationships, predict future trends and behaviors, allowing businesses/users to make proactive, knowledge-driven decisions. This Volume encouraged submission of original research articles and expository papers on theory and algorithms of function approximation and classification, and real world applications.

This edited volume comprises of 14 chapters, including several overview Chapters, which provides an up-to-date and state-of-the art research covering the theory and algorithms of function approximation and classification.

Chapter 1 "Feature Selection for Partial Least Square Based Dimension Reduction" by *Li and Zeng* investigate a systematic feature reduction framework by combing dimension reduction with feature selection. To evaluate the proposed framework authors used four typical data sets. Experimental results illustrate that the proposed method improves the performance on the gene expression microarray data in terms of accuracy.

In Chapter 2, "Classification by the Use of Decomposition of Correlation Integral" *Jirina and Jirina Jr.* illustrate that the correlation integral can be decomposed into functions each related to a particular point of data space. For these functions, one can use similar polynomial approximations such as the correlation integral. The essential difference is that the value of the exponent, which would correspond to the correlation dimension, differs in accordance to the position of the point in question.

Chapter 3, "Novel biomarkers for prostate cancer revealed by (α,β)) - k-feature sets" by *Ravetti et al.* present a method based on the (α,β) - k- feature set problem for identifying relevant attributes in high-dimensional datasets for classification purposes. Using the gene expression of thousands of genes, authors illustrate that the method can give a reduced set that can identify samples as belonging to prostate cancer tumors or not.

In **Chapter 4**, "Use of the q-Gaussian Function in Radial Basis Function Networks" *Tinos and Murta Jr.* deploy q-Gaussian function as a radial basis function in RBF

Networks for pattern recognition problems. The use of q-Gaussian RBFs allows to modify the shape of the RBF by changing the real parameter q, and to employ radial units with different RBF shapes in a same RBF Network.

Most classification problems associate a single class to each example or instance. However, there are many classification tasks where each instance can be associated with one or more classes. This group of problems represents an area known as multi-label classification. In **Chapter 5**, "A Tutorial on Multi-Label Classification Techniques**"** *Carvalho and Freitas* present the most frequently used techniques to deal with these problems in a pedagogical manner, with examples illustrating the main techniques and proposing a taxonomy of multi-label techniques that highlights the similarities and differences between these techniques.

In **Chapter 6**, "Computational Intelligence in Biomedical Image Processing" *Bollenbeck and Seiffert* show that the segmentation of biological images, characterized by non-uniform image features, significantly benefits from combining global physical models and local feature-based supervised classification. Authors used an entropy-based voting of optimal feed-forward networks by cross-validation architecture selection and global registration-based segmentation.

Chapter 7, "Investigating Neighborhood Graphs for Inducing Density Based Clusters" by *Kasahara and Nicoletti* investigate the impact of seven different ways of defining a neighborhood region between two points, when identifying clusters in a neighborhood graph, particularly focusing on density based clusters. On the one hand results show that the neighborhood definitions that do not employ parameters are not suitable for inducing density based clusters. On the other hand authors also illustrate that although essential for successfully separating density based clusters, the parameters employed by some of the definitions need to have their values tuned by the user.

Chapter 8, "A Comparative Study of Three Graph Edit Distance Algorithms" by *Gao et al.* propose two cost function-free GED algorithms. In the edge direction histogram (EDH)-based method, edit operations are involved in graph structure difference characterized by EDH and GED is converted into earth mover's distance (EMD) of EDHs, while edit operations are involved in node distribution difference characterized by HMM and GED is converted into KLD of HMMs in the HMM-based method. With respect to two cost function free algorithms, HMM-based method excels EDH-based method in classification and clustering rate, and efficiency.

In **Chapter 9**, "Some Issues on Extensions of Information and Dynamic Information Systems" *Pancerz* discusses some issues on extensions of information and dynamic information systems. Consistent and partially consistent extensions of information and dynamic information systems are helpful in prediction problems. On the basis of those extensions author determine possibility distributions of states and transitions between states over a universe of discourse related to a given system of processes.

Chapter 10, "A Probabilistic Approach to the Evaluation and Combination of Preferences" *Parracho and Anna* propose a model for the process of evaluating and

combining preferences. After determining the preferences according to each criterion, these partial preferences are combined into global ones. Different combination rules are set, in a probabilistic framework. Attention is centered to the case of indicators measured in different levels of aggregation.

In **Chapter 11**, "Classification of Complex Molecules" by *Torrens and Castellano* introduce algorithms for classification and taxonomy based on criteria, e.g., information entropy and its production. In molecular classification, the feasibility of replacing a given molecule by similar ones in the composition of a complex drug is studied.

In **Chapter 12** "Intelligent finite element method and application to simulation of behavior of soils under cyclic loading" *Javad et al.* present a neural network-based finite element method for modeling of the behavior of soils under cyclic loading. The methodology is based on the integration of a neural network in a finite element framework. In this method, a neural network is trained using experimental data representing the mechanical response of material to applied load. The trained network is then incorporated in the finite element analysis to predict the constitutive relationships for the material.

Chapter 13, "An Empirical Evaluation of the Effectiveness of Different Types of Predictor Attributes in Protein Function Prediction" *Otero et al.* present an empirical evaluation of different protein representations for protein function prediction in terms of maximizing predictive accuracy, investigating which type of representation is more suitable for different levels of hierarchy.

In the last Chapter, "Genetic Selection Algorithm and Cloning for Data Mining with GMDH Method" *Jirina and Jirina Jr.* generalize the idea of the genetically modified GMDH neural network for processing multivariate data appearing in data mining problems and to extend this type of network by cloning. Clones are close, but not identical copies of original individuals. The new genetically modified GMDH method with cloning (GMC GMDH) has no tough layered structure.

We are very much grateful to the authors of this volume and to the reviewers for their great efforts by reviewing and providing interesting feedback to authors of the chapter. The editors would like to thank Dr. Thomas Ditzinger (Springer Engineering Inhouse Editor, Studies in Computational Intelligence Series), Professor Janusz Kacprzyk (Editor-in-Chief, Springer Studies in Computational Intelligence Series) and Ms. Heather King (Editorial Assistant, Springer Verlag, Heidelberg) for the editorial assistance and excellent cooperative collaboration to produce this important scientific work. We hope that the reader will share our joy and will find it useful!

December 2008 Editors

Ajith Abraham, Trondheim, Norway
Aboul Ella Hassanien, Cairo, Egypt
Václav Snášel, Ostrava, Czech Republic