Preface

Machine learning (ML) is one of the most fruitful fields of research currently, both in the proposal of new techniques and theoretic algorithms and in their application to real-life problems. From a technological point of view, the world has changed at an unexpected pace; one of the consequences is that it is possible to use high-quality and fast hardware at a relatively cheap price. The development of systems that can be adapted to the environment in a smart way has a huge practical application. Usually, these systems work by optimizing the performance of a certain algorithm/technique according to a certain maximization/minimization criterion but using experimental data rather than a given “program” (in the classical sense). This way of tackling the problem usually stems from characteristics of the task, which might be time-variant or too complex to be solved by a sequence of sequential instructions. Lately, the number of published papers, patents and practical applications related to ML has increased exponentially. One of the most attractive features of ML is that it brings together knowledge from different fields, such as, pattern recognition (neural networks, support vector machines, decision trees, reinforcement learning, …), data mining (time series prediction, modeling, …), statistics (Bayesian methods, Montecarlo methods, bootstrapping, …) or signal processing (Markov models), among others. Therefore, ML takes advantage of the synergy of all these fields thus providing robust solutions that use different fields of knowledge.

Therefore, ML is a multidisciplinary topic and it needs some particular bibliography to gather its different techniques. There are a number of excellent references to go into ML, but the wide range of applications of ML has not been discussed in any reference book. Both theory and practical applications are discussed in this handbook. Different state-of-the-art techniques are analyzed in the first part of the handbook, while a wide and representative range of practical applications are shown in the second part of the book. The editors would like to thank the collaboration of the authors of the first part of the handbook, who accepted the suggestion of including a section of applications in their chapters.

A short introduction to the chapters will be provided in the following. The first part of the handbook consists of eleven chapters. In chapter 1, R. Xu and D. C. Wunsch II do a review of clustering algorithms and provide some real applications; it should be pointed out the adequacy of the references, which show up the deep knowledge of the authors in these techniques, ratified by the book published by the authors recently in 2008. In chapter 2, “Principal Graphs and Manifolds,” Gorban and Zinovyev present the machine learning approaches to the problem of dimensionality reduction with controlled complexity. They start with classical techniques, such as principal component analysis (PCA) and the k-means algorithm. There is a whole universe of approximants between the ‘most rigid’ linear manifolds (principal components) and ‘most soft’ unstructured finite sets of k-means. This chapter gives a brief practical introduction into the methods of construction of general principal objects (i.e., objects embedded in the ‘middle’ of the multidimensional data set). The notions of self-consistency and coarse-grained self-consistency give the general probabilistic framework for construction of principal objects. The family of expectation/maximization algorithms with nearest generalizations is presented. Construction of principal graphs
with controlled complexity is based on the graph grammar approach. In the theory of principal curves and manifolds the penalty functions were introduced to penalize deviation from linear manifolds. For branched principal object the pluriharmonic embeddings (‘pluriharmonic graphs’) serve as ‘ideal objects’ instead of planes, and deviation from this ideal form is penalized. Chapter 3, “Learning Algorithms for RBF Functions and Subspace Based Functions,” by Lei Xu, overviewed advances on normalized radial basis function (RBF) and alternative mixture-of-experts as well as further developments to subspace based functions (SBF) and temporal extensions. These studies are linked to a general statistical learning framework that summarizes not only maximum likelihood learning featured by the EM algorithm but also Bayesian Ying Yang (BYY) harmony and rival penalized competitive learning (RPCL) featured by their automatic model selection nature, with a unified elaboration of their corresponding algorithms. Finally, remarks have also been made on possible trends.

Sanjoy Das, Bijaya K. Panigrahi and Shyam S. Pattnaik show in chapter 4, “Nature Inspired Methods for Multi-Objective Optimization,” an application of the three basic classes of nature inspired algorithms – evolutionary algorithms, particle swarm optimization, and artificial immune systems, to multi-objective optimization problems. As hybrid algorithms are becoming increasingly popular in optimization, this chapter also includes a brief discussion of hybridization within a multi-objective framework. In Chapter 5, “Artificial Immune Systems for Anomaly Detection,” Eduard Plett, Sanjoy Das, Dapeng Li and Bijaya K. Panigrahi present anomaly detection algorithms analogous to methods employed by the vertebrate immune system, with an emphasis on engineering applications. The chapter also proposes a novel scheme to classify all algorithmic extensions of negative selection into three basic classes: self-organization, evolution, and proliferation. As anomaly detection can be considered as a binary classification problem, in order to further show the usefulness of negative selection, this algorithm is then modified to address a four-category problem, namely the classification of power signals based on the type of disturbance. Chapter 6, written by Antonio Bella, César Ferri, José Hernández-Orallo, and María José Ramírez-Quintana and entitled “Calibration of Machine Learning Models,” reviews the most common calibration techniques and calibration measures. Calibration techniques improve the probability estimation of the model or correct its local (or global) bias, and the degree of calibration is assessed with calibration measures. Both classification and regression tasks are covered in this chapter, and a new taxonomy of calibration techniques is established. Chapter 7, “Classification with Incomplete Data” by Pedro J. García-Laencina, Juan Morales-Sánchez, Rafael Verdú-Monedero, Jorge Larrey-Ruiz, José-Luis Sancho-Gómez and Aníbal R. Figueiras-Vidal, deals with machine learning solutions for incomplete pattern classification; nowadays, data is generated almost everywhere: sensor networks in Mars, submarines in the deepest ocean, opinion polls about any topic, etc. Many of these real-life applications suffer a common drawback, missing or unknown data. The ability of handling missing data has become a fundamental requirement for pattern classification because inappropriate treatment of missing data may cause large errors or false results on classification. Machine learning approaches and methods imported from statistical learning theory have been most intensively studied and used in this subject. The aim of this chapter is to analyze the missing data problem in pattern classification tasks, and to summarize and compare some of the well-known methods used for handling missing values.

Chapter 8 shows that most of the existing research on multivariate time series (MTS) targets supervised prediction and forecasting problems. To date, in fact, little research has been conducted on the exploration of MTS through unsupervised clustering and visualization. Olier and Vellido describe generative topographic mapping through time (GTM-TT), a model with foundations in probability theory that performs such tasks. The standard version of this model has several limitations that limit its applicability, so, in this work, GTM-TT is reformulated within a Bayesian approach using variational techniques. The resulting variational Bayesian GTM-TT is shown to behave very robustly in the presence of noise in the
MTS, helping to avert the problem of data overfitting. Chapter 9, by Todor Ganchev entitled “Locally Recurrent Neural Networks and Their Applications,” offers a review of the various computational models of locally recurrent neurons, and surveys locally recurrent neural network (LRNN) architectures that are based on them. These locally recurrent architectures are capable of identifying and exploiting temporal and spatial correlations (i.e., the context in which events occur). This capability is the main reason for the advantageous performance of LRNN, when compared with their non-recurrent counterparts. Examples of real-world applications that rely on infinite impulse response (IIR) multilayer perceptron (MLP) neural networks, diagonal recurrent neural networks (DRNN), locally recurrent radial basis function neural networks (LRRBFNNs), locally recurrent probabilistic neural networks (LRPNNs), and that involve classification or prediction of temporal sequences, discovering and modeling of spatial and temporal correlations, process identification and control, etc., are briefly outlined. A quantitative assessment of the number of weights in a single layer of neurons, implementing different types of linkage, is presented as well. In conclusion, a brief account of the advantages and disadvantages of LRNNs is performed, and potentially promising research directions are discussed.

Chapter 10, “Nonstationary signal analysis with kernel machines” by Paul Honeine, Cédric Richard and Patrick Flandrin, introduces machine learning for nonstationary signal analysis and classification. The authors show that some specific reproducing kernels allow a pattern recognition algorithm to operate in the time-frequency domain. Furthermore, the authors study the selection of the reproducing kernel for a nonstationary signal classification problem. The last chapter of this theoretic section, chapter 11, “Transfer Learning” by Lisa Torrey and Jude Shavlik, discusses transfer learning, which is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned. While most machine learning algorithms are designed to address single tasks, the development of algorithms that facilitate transfer learning is a topic of ongoing interest in the machine-learning community. This chapter provides an introduction to the goals, formulations, and challenges of transfer learning. It surveys current research in this area, giving an overview of the state of the art and outlining the open problems. The survey covers transfer in both inductive learning and reinforcement learning, and discusses the issues of negative transfer and task mapping in depth.

The second part of the book is focused on applications of ML to real-life problems. In chapter 12, “Machine Learning in Personalized Anemia Treatment,” Adam E. Gaweda shows an interesting and ingenious application of reinforcement learning to drug dosing personalization in treatment of chronic conditions. In treatment of chronic illnesses, finding the optimal dose amount for an individual is also a process that is usually based on trial-and-error. This chapter focus on the challenge of personalized anaemia treatment with recombinant human erythropoietin and demonstrate the application of a standard reinforcement learning method, namely Q-learning, to guide the physician in selecting the optimal erythropoietin dose. Finally, the author shows computer simulations to compare the outcomes from reinforcement learning-based anaemia treatment to those achieved by a standard dosing protocol used at a dialysis unit.

Chapter 13, “Deterministic Pattern Mining On Genetic Sequences” by Pedro Gabriel Ferreira and Paulo Jorge Azevedo, presents an overview to the problem of mining deterministic motifs in collections of DNA or protein sequences. The large amount of available biological sequence data requires efficient techniques for motif mining. The authors start by introducing the basic concepts associated with sequence motif discovery. Next, an architecture common to discovery methods is proposed. Each of its blocks are discussed individually. Particular attention is given to different algorithmic approaches proposed in the literature. It finishes with a summary of the characteristics of the presented methods. Chapter 14, “Machine Learning in Natural Language Processing” by Marina Sokolova and Stan Szpakowicz, presents applications of ML in fundamental language-processing and linguistic problems: identify a word’s
part-of-speech, determine its meaning, find relations among parts of a sentence, and so on. Problems of that kind, once solved, are a necessary component of such higher-level language processing tasks as, for example, text summarization, question answering or machine translation. People are usually very good at such tasks; software solutions tend to be inferior. Solutions based on linguistically motivated, manually constructed rules are also quite labour-intensive. The easy availability of large collections of texts points to ML as a method of choice for building solutions based on texts data in amounts beyond the human capacity. Even the quality of those solutions, however, is no match for human performance yet.

In Chapter 15, the same authors also present applications to problems which involve the processing of very large amounts of texts. Problems best served by ML came into focus after the Internet and other computer-based environments acquired the status of the prime medium for text delivery and exchange. That is when the ability to handle extremely high volume of texts, which ML applications had not previously faced, became a major issue. The resulting set of techniques and practices, the so-called name mega-text language processing, are meant to deal with a mass of informally written, loosely edited text. The chapter makes a thorough review of the performance of ML algorithms that help solve knowledge-intensive natural language processing problems where mega-texts are a significant factor. The authors present applications which combine both scientific and economic interests. Chapter 16, “FOL Learning for Knowledge Discovery in Documents” by Stefano Ferilli, Floriana Esposito, Marenglen Biba, Teresa M.A. Basile and Nicola Di Mauro, proposes the application of machine learning techniques, based on first-order logic as a representation language, to the real-world application domain of document processing. First, the tasks and problems involved in document processing are presented, along with the prototypical system DOMINUS and its architecture, whose components are aimed at facing these issues. Some experiments are reported that assess the quality of the proposed approach. The authors like to prove to researchers and practitioners of the field that first-order logic learning can be a viable solution to tackle the domain complexity, and to solve problems such as incremental evolution of the document repository. The field of applications changes in the next chapters.

Thus, Chapter 17, “Machine Learning and Financial Investing” by Jie Du and Roy Rada, begins with a model for financial investing and then reviews the literature as regards knowledge-based and machine-learning based methods for financial investing. The claim is that knowledge bases can be incorporated into an evolutionary computation approach to financial investing to support adaptive investing, and the design of a system that does this is presented. In Chapter 18, “Applications Of Evolutionary Neural Networks For Sales Forecasting of Fashionable Products” by Yong Yu, Tsan-Ming Choi and Kin-Fan Au and Zhan-Li Sun, a theoretical framework is proposed in which the details on how an evolutionary computation approach can be applied in searching for a desirable network structure for the forecasting task are discussed. The optimized ENN structure for sales forecasting is then developed. With the use of real sales data, the performances of the proposed ENN forecasting scheme are compared with several other traditional methods which include artificial neural network and SARIMA. Insights regarding the applications of ENN for forecasting sales of fashionable products are generated.

In Chapter 19, “Support Vector Machine based Hybrid Classifiers and Rule Extraction Thereof: Application to Bankruptcy Prediction in Banks,” Farquad, Ravi and Bapi propose a hybrid rule extraction approach using support vector machine in the first phase and one of the intelligent techniques such as fuzzy rule based systems (FRBS), adaptive network based fuzzy inference system (ANFIS), decision tree (DT) and radial basis function networks (RBF) in the second phase within the framework of soft computing. They applied these hybrid classifiers to problem of bankruptcy prediction in banks. In the proposed hybrids, first phase extracts the support vectors using the training set and these support vectors are used to train FRBS, ANFIS, DT and RBF to generate rules. Empirical study is conducted using three datasets viz., Spanish banks, Turkish banks and US banks. It is concluded that the proposed hybrid
rule extraction procedure outperformed the stand-alone classifiers. Chapter 20, “Data Mining Experiences in Steel Industry” by Joaquín Ordieres-Meré, Ana González-Marcos, Manuel Castejón-Limas and Francisco J. Martínez-de-Pisón, reports five experiences in successfully applying different data mining techniques in a hot-dip galvanizing line. The work was aimed at extracting hidden knowledge from massive data bases in order to improve the existing control systems. The results obtained, though small at first glance, lead to huge savings at such high volume production environment. Fortunately, the industry has already recognized the benefits of data mining and is eager to exploit its advantages. Some editors of this handbook, together with Carlos Fernández and Juan Guerrero present the application of neural networks, specifically Multilayer Perceptrons and Self-Organizing Maps to Animal Science in Chapter 21, “Application of neural networks in Animal Science.” Two different applications are shown; first, milk yield prediction in goat herds, and second, knowledge extraction from surveys in different farms that is then used to improve the management of the farms.

Chapter 22, “Statistical Machine Learning Approaches for Sports Video Mining using Hidden Markov Models” by Guoliang Fan and Yi Ding, discusses the application of statistical machine learning approaches to sports video mining. Specifically, the authors advocate the concept of semantic space where video mining is formulated as an inference problem so that semantic computing can be accomplished in an explicit way. Several existing hidden Markov models (HMMs) are studied and compared regarding their performance. Particularly, it is proposed a new extended HMM that incorporates advantages from existing HMMs and offer more capacity, functionality and flexibility than its precedents. This chapter discusses the application of statistical machine learning approaches to sports video mining. Specifically, it is advocated the concept of semantic space where video mining is formulated as an inference problem so that semantic computing can be accomplished in an explicit way. Several existing hidden Markov models (HMMs) are studied and compared regarding their performance. Particularly, the authors propose a new extended HMM that incorporates advantages from existing HMMs and offer more capacity, functionality and flexibility than its precedents. José Blasco, Nuria Aleixos, Juan Gómez-Sanchis, Juan F. Guerrero and Enrique Moltó, in Chapter 23, “A Survey of Bayesian Techniques in Computer Vision,” show some applications of Bayesian learning techniques in computer vision. Agriculture, inspection and classification of fruit and vegetables, robotics, insect identification and process automation are some of the examples shown. Problems related to the natural variability of color, sizes and shapes of biological products, and natural illuminants are also discussed. Finally, implementations that lead to real-time implementation are explained.

Chapter 24, “Software Cost Estimation using Soft Computing Approaches” by K. Vinaykumar, V. Ravi and Mahil Carr, shows a different application. Predicting the cost of software products to be developed is the main objective of the software engineering practitioners. No model has proved to be effective, efficient and consistent in predicting the software development cost. In this chapter, the authors investigated the use of soft computing approaches for estimating software development effort. Further using intelligent techniques, linear ensembles and non-linear ensembles are developed within the framework of soft computing. The developed ensembles are tested on COCOMO’81 data. It is clear from empirical results that non-linear ensemble using radial basis function network as an arbitrator outperformed all other ensembles. Chapter 25 shows an application related to that described in the previous chapter. “Counting the Hidden Defects in Software Documents” by Frank Padberg shows the use of neural networks to estimate how many defects are hidden in a software document. Inputs for the models are metrics that get collected when effecting a standard quality assurance technique on the document, a software inspection. The author adapts the size, complexity, and input dimension of the networks to the amount of information available for training; and using Bayesian techniques instead of cross-validation for determining model parameters and selecting the final model. For inspections, the machine learning approach is highly suc-
successful and outperforms the previously existing defect estimation methods in software engineering by a factor of 4 in accuracy on the standard benchmark. This approach is well applicable in other contexts that are subject to small training data sets. Chapter 26, “Machine Learning for Biometrics” by Albert Ali Salah, deals with an application within the field of biometrics. The recently growing field of biometrics involves matching human biometric signals in a fast and reliable manner for identifying individuals. Depending on the purpose of the particular biometric application, security or user convenience may be emphasized, resulting in a wide range of operating conditions. ML methods are heavily employed for biometric template construction and matching, classification of biometric traits with temporal dynamics, and for information fusion with the purpose of integrating multiple biometrics. The chapter on ML for biometrics reviews the hot issues in biometrics research, describes the most influential ML approaches to these issues, and identifies best practices. In particular, the chapter covers distance and similarity functions for biometric signals, subspace-based classification methods, unsupervised biometric feature learning, methods to deal with dynamic temporal signals, classifier combination and fusion for multiple biometric signals. Links to important biometric databases and code repositories are provided.

Chapter 27, “Neural Networks For Modeling The Contact Foot-Shoe Upper” by M. J. Rupérez, J. D. Martín, C. Monserrat and M. Alcañiz, shows that important advances in virtual reality make possible real improvements in footwear computer aided design. To simulate the interaction between the shoe and the foot surface, several tests are carried out to evaluate the materials used as the footwear components. This chapter shows a procedure based on artificial neural networks (ANNs) to reduce the number of tests that are needed for a comfortable shoe design. Using the ANN, it is possible to find a neural model that provides a unique equation for the characteristic curve of the materials used as shoe uppers instead of a different characteristic curve for each material. Chapter 28, “Evolutionary Multi-objective Optimization of Autonomous Mobile Robots in Neural-Based Cognition for Behavioral Robustness” by Chin Kim On, Jason Teo, and Azali Saudi, shows the utilization of a multi-objective approach for evolving artificial neural networks that act as the controllers for phototaxis and radio frequency (RF) localization behaviors of a virtual Khepera robot simulated in a 3D physics-based environment. It explains the comparison of performances between the elitism without archive and elitism with archive used in the evolutionary multi-objective optimization (EMO) algorithm in an evolutionary robotics perspective. Furthermore, the controllers’ moving performances, tracking ability and robustness have also been demonstrated and tested with four different levels of environments. The experimentation results show the controllers enable the robots to navigate successfully, hence demonstrating the EMO algorithm can be practically used to automatically generate controllers for phototaxis and RF-localization behaviors, respectively. Understanding the underlying assumptions and theoretical constructs through the utilization of EMO will allow the robotics researchers to better design autonomous robot controllers that require minimal levels of human-designed elements.

The last chapter, chapter 29, “Improving Automated Planning with Machine Learning” by Susana Fernández Arregui, Sergio Jiménez Celorio and Tomás de la Rosa Turbidez, reports the last machine learning techniques for the assistance of automated planning. Recent discoveries in automated planning have opened the scope of planners, from toy problems to real-world applications, making new challenges come into focus. The chapter collects the last machine learning techniques for assisting automated planners. For each technique, the chapter provides an in-depth analysis of their domain, advantages and disadvantages; finally, the chapter draws the outline of the new promising avenues for research in learning for planning systems.