NEUROCOMPUTING

ELSEVIER Neurocomputing 47 (2002) 305–307

www.elsevier.com/locate/neucom

Book reviews[1]

Learning and Soft Computing, Support Vector Machines, Neural Networks, and Fuzzy Logic Models, Vojislav Kecman; MIT Press, Cambridge, MA, 2001, pp. 578. ISBN 0-262-11255-8, 2001.

There are plenty of books in the broad field of machine learning, computational intelligence, data mining, neural networks, support vector machines, statistical learning theory, knowledge discovery, fuzzy logic, soft computing and other related but variously named areas. However, if one is looking for an excellent textbook that covers all the relevant basics of such a broad field, he or she should read this book first.

The book reviewed here is definitely at the top of a long list of books in this area. It is positioned at the top of the list because it is the first comprehensive textbook providing a thorough unified introduction to the field of *learning from experimental data* (i.e., examples, samples, measurements, records, patterns, observations, images) and *soft computing*. Support vector machines (SVMs) and neural networks (NNs) are presented as the mathematical structures (i.e., models, learning machines, sets of functions implemented in software) that represent learning, while fuzzy logic systems (FLS) are introduced as the tool and the set of techniques for embedding structured human knowledge (expertise, heuristics, experience) into workable algorithms.

Learning and Soft Computing (LearnSC) embodies 268 illustrations, 155 problems, 47 practical examples, 3 extended case studies on NNs based control, financial time series analysis, and computer graphics, as well as many sets of simulated experiments. This is a completely modern book accompanied by the powerful and highly educational MATLAB based, and user friendly, software. The software, as well as the Power Point slides covering the fields of SVMs, NNs and FL models, is downloadable from the book's web site given above. The Solutions Manual that contains the solutions to the book's 155 problems has been prepared for instructors who wish to refer to the author's methods of solution. It is available from the publisher.

Let us quote from the book's Preface that describes the approach taken, format and overall profile of the book LearnSC: 'Each chapter is arranged so that the basic theory and algorithms are illustrated by practical examples and followed by a set of problems and simulation experiments. In the author's experience, this approach is the most accessible, pleasant and useful way to master this material, which contains many new (and potentially difficult) concepts. To some extent, the problems are intended to help the reader acquire technique, but most of them serve to illustrate and develop further the basic subject matter of the chapter. The author feels that this structure is suitable both for a textbook used in a formal course and for self-study'. We entirely agree with such a writing approach, and we believe that the standard set by this volume should be closely followed in all future textbooks on the topics of SVMs, NNs and FLS.

In addition, the detailed book presentation can also be found on its web site `http://ww.support-vector.ws` where, presented in an attractive site design, the following pages cover the broad aspects of such a novel textbook product—Table of contents, Preface, Chapter's survey, Case Studies, Publications, Downloads, Meet the author, Solutions manual, Order the book, Companion web sites of the book, Links, FAQ and more.

The book's basic assumption is, that it is not only useful, but also necessary, to treat SVMs, NNs, and FLS as parts of a connected whole. And, the author has done exactly that, skillfully using an approach that enables the reader to develop SVMs, NNs and FLS in addition to understanding them. The book LearnSC presents these three 'different' approaches, techniques and methods within the Preface, Introduction and nine

[1] Material to be included in this section can be submitted to: Prof. K.J. Cios, Computer Science and Engineering Dept., University of Colorado at Denver, Campus Box 109, 1200 Larimer Street, Denver, CO 80217-3364, USA. E-mail: kcois@carbon.cudenver.edu

PII: S0925-2312(01)00685-3

chapters as follows:

1. Learning and Soft Computing: Rationale, Motivations, Needs, Basics (p. 117).
2. Support Vector Machines (p. 72).
3. Single-Layer Networks (p. 62).
4. Multilayer Perceptrons (p. 58).
5. Radial Basis Function Networks (p. 52).
6. Fuzzy Logic Systems (p. 56).
7. Case Studies (p. 60).
8. Basic Nonlinear Optimization Methods (p. 24).
9. Mathematical Tools of Soft Computing (p. 16).

Selected Abbreviations, Notes, References and a detailed Index are provided in the closing sections of the book. (The number of pages devoted to each chapter reflects the authors' balanced approach to each particular modeling too.) The description of each chapter is an extension of the book web site's chapter's survey:

Chapter 1 gives examples of applications, presents the basic tools of soft computing (NNs, SVMs and FL models), reviews the classical problems of approximation of multivariate functions, and introduces the standard statistical approaches to regression and classification that are based on the knowledge of probability–density functions. Here is where the basic problem of nonlinear (in terms of parameter, and here called the weights to be learned) approximation is presented by using three low-dimensional examples. This enabled the author to introduce and show graphically the character of the nonlinear cost function surface. And, such nonlinearity urges the application of both the classic nonlinear optimization techniques and the novel ones.

Chapter 2 presents the basics of statistical learning theory when there is no information about the probability distribution but only experimental data. The VC dimension and structural risk minimization are introduced. A description is given of the SVM learning algorithm based on quadratic programming that leads to parsimonious SVMs, that is, NNs or SVMs having a small number of hidden layer neurons. The parsimony results from sophisticated learning that matches model capacity to data complexity. In this way, good generalization, meaning the performance of the SVM on previously unseen data, is assured.

Chapter 3 deals with two early learning units—the perceptron and the linear neuron (adaline)—as well as with single-layer networks. Five different learning algorithms for the linear activation function are presented. Despite the fact that the linear neuron appears to be very simple, it is the constitutive part of almost all models treated here and therefore is a very important processing unit. The linear neuron can be looked upon as a graphical (network) representation of classical linear regression and linear classification (discriminant analysis) schemes. Here is where the author stresses the difference between the two faces of the models presented in the LearnSC book. He discusses the representational capabilities of such models as well as their learning capability in detail.

A genuine neural networks (a multilayer perceptron)—one that comprises at least one hidden layer having neurons with nonlinear activation functions—is introduced in Chapter 4. The error-correction type of learning, introduced for single-layer networks in Chapter 3, is generalized, and the gradient-based learning method known as error backpropagation is discussed in detail in this chapter. Also shown are some of the generally accepted heuristics while training multilayer perceptrons. In particular, there is a discussion about whether one, two, or more hidden layers should be used. The discussion about the number of neurons in a hidden layer, necessarily leads to the crucial bias-variance dilemma, i.e., to the problem of under- or over-fitting the data. What type of activation functions in a hidden layer to use, how to initialize weights, what is the geometry of approximation, which error function to use for stopping criterion at learning, and how to choose the learning rate as well as the momentum term, are the other pieces of heuristics discussed and presented in detail at the end of this chapter.

Chapter 5 is concerned with regularization networks, which are better known as radial basis function (RBF) networks. The notion of ill-posed problems is discussed as well as how regularization leads to networks whose activation functions are radially symmetric. Details are provided on how to find a parsimonious radial basis network by applying the orthogonal least squares approach. Also explored is a linear programming (LP) approach to subset (basis function or support vector) selection that, similar to the QP based algorithm for SVMs training, leads to parsimonious NNs and SVMs. This is the first textbook that we know of, that presents the LP in designing sparse NNs and/or SVMs for regression problems. The possible advantages and shortcomings are discussed. Interestingly enough, the LP approach does not require that the hidden layer neurons possess the kernel type of functions. This may be of benefit when used in other data mining models.

Fuzzy logic modeling is the subject of Chapter 6. Basic notions of fuzzy modeling are introduced—fuzzy sets, relations, compositions of fuzzy relations, fuzzy inference, and defuzzification. The union, intersection, and Cartesian product of a family of sets are described, and various properties are established. The similarity between, and sometimes even the equivalence of, RBF networks and fuzzy models are noted in detail. Finally, fuzzy additive models (FAMs) are presented as a simple yet powerful fuzzy modeling technique. FAMs are the most popular type of fuzzy models in today's applications.

Chapter 7 presents three case studies that show the beauty and strength of these modeling tools—neural networks-based control systems, financial time series prediction, and computer graphics by applying neural networks models and these are discussed at length.

Chapter 8 and 9 are the reference part of the book. Chapter 8 focuses on the most popular classical approaches to nonlinear optimization, which is the crucial part of learning from data. It also describes the novel massive search algorithms known as genetic algorithms or evolutionary computing.

Chapter 9 contains specific mathematical topics and tools that might be helpful for understanding the theoretical aspects of soft models, although these concepts and tools are not covered in great detail. It is supposed that the reader has some knowledge of probability theory, linear algebra, and vector calculus. Chapter 9 is designed only for easy reference of properties and notation.

This original and pioneering book is organized around the supervised learning and fuzzy logic modeling. In an information age and in a heavily digitalized world, where terabytes of information are produced daily, we cannot survive unless we possess the ability to learn from that ocean of data. Today, machines should perform such learning in a highly intelligent manner. Otherwise, there will be no sensible use of data. This is exactly the topic of this innovative volume—learn from the data in a clever manner, extract all the relevant information smartly, neglect the less relevant data, discover new pieces of 'truth' and increase knowledge. This is simple to state, but hard to achieve. This book did it. In parallel to this, it describes how to develop the algorithms for transferring existing human knowledge into workable algorithms. Altogether, the book shows the ways in which these ultimate goals of learning and modeling human knowledge may be achieved in a gradual and pleasant manner.

Learning and Soft Computing provides a clearly organized book focusing on a broad range of algorithms and is aimed at senior undergraduate students, graduate students and practicing researchers and scientists who want to use and develop SVMs, NNs and/or FL models rather than simply study them. The book is rich in graphical presentations (268 illustrations). This is another attraction of the book because graphics provide clarity of expression and strengthen the explanation of the material presented. The insight obtained through the simulation experimenting and graphical presentation of the results definitely enables faster and better understanding of initially 'difficult' matrix–vectorial notations present in the field. This also makes the book highly suitable for a self-study.

Ljubo Vlacic
School of Microelectronic Engineering, Griffith University, Nathan Campus, Brisbane, QLD 4111, Australia
E-mail address: l.vlacic@me.gu.edu.au