Book Reviews

Combining Pattern Classifiers: Methods and Algorithms—Ludmila I. Kuncheva (Hoboken, NJ: Wiley, 2004, pp. 350, ISBN: 0-471-21078-1). *Reviewed by Ethem Alpaydin*

The first time I heard someone mention the idea of combining multiple learners was in spring 1991. The speaker was M. Warmuth of the University of California at Santa Cruz (UCSC), Santa Cruz. The venue was Stanford University, Stanford, CA. I was a postdoctoral researcher at the International Computer Science Institute (ICSI), University of California, Berkeley, and this was a one-day meeting of connectionists from ICSI, Stanford University, and UCSC. From that day, I remember people discussing the advantage of taking a vote over three separately trained multilayer perceptrons (MLP) versus combining the three MLPs with a layer on top and training it all as one large network.

It is no coincidence that work on multiple learners started in early 1990s. At that time, we had enough computation and memory to train/test/store multiple learners. It was also then that low-cost multiprocessor systems started to appear and what better and easier way is there to program 64 processors than running 64 separate MLPs on them?

The theory and practice of combining multiple learners has gone a long way since then. Today, any machine learning/pattern recognition textbook devotes at least a chapter to model combination and it was a high time for a book wholly devoted to the subject. Kuncheva, herself the author of many influential papers on model combination, has done a very good job and the book is a must-read for researchers and practitioners.

The book is divided into ten chapters. The first two chapters are introductory. Chapter 1 introduces the basic concepts of a class, feature, classifier, discriminant, confusion matrix, cross validation, statistical testing, and Bayes decision theory. Chapter 2 discusses basic algorithms that are used as base classifiers in the following chapters, such as linear and quadratic classifiers, nonparametric classifiers (Parzen windows and k-nearest neighbor), decision trees, and MLPs.

Chapter 3 introduces the reader to multiple classifier systems and discusses dimensions along which to compare different schemes: fusion versus selection, fixed versus trained, etc. The next two chapters review methods for combination where the model outputs are discrete and continuous, respectively. Chapter 4 discusses combination methods when the base classifier outputs are 0/1. Among those that are reviewed are voting, naive combination, and multinomial methods (where, basically, the base classifier outputs are seen as binary inputs to a second-level classifier, as in stacking). Chapter 5 similarly discusses methods when the outputs are continuous, e.g., posterior probabilities. There is a distinction made between class-conscious and class-indifferent combination; however, the most important part of this chapter is the discussion of fixed combination rules: average, median, product, minimum, and maximum.

Chapter 6 reviews methods for classifier selection; that is, given an input, we do not consult all the models in the ensemble but use one (or few). One way to do this is to have a calculation of local competence for each base model; another is to have a gating learner which chooses one of the local models given the input, which is called a mixture of experts.

The reviewer is with the Department of Computer Engineering, Boğaziçi University, Istanbul TR-34342, Turkey (e-mail: alpaydin@boun.edu.tr).

Digital Object Identifier 10.1109/TNN.2007.897478

Chapter 7 discusses bagging and boosting (adaboost), two popular combination methods. This chapter also contains a nice discussion of bias–variance–noise decomposition of error and various ways to estimate bias and variance for classification problems.

Chapter 8 is a catch-all. It contains a discussion of methods where different base classifiers use different subsets of the features, as in sensor fusion. There is a section on error-correcting output codes, which I would have put in chapter 4. There is also a very informative section on methods for combining clustering methods. However, this chapter lacks a section on combining regressors.

Chapter 9 is on the theory of fixed combination rules and analyzes their behavior under different assumptions about the distribution of base classifier outputs.

Chapter 10 discusses the concept of diversity, the elusive measure telling us which models best complement each other. The chapter discusses how diversity can be estimated, its relation to accuracy, and how it can be used in practice (from among a set of candidate models, combine only those which are diverse), and concludes very aptly with a section entitled "Diversity of diversity."

Overall, this is a very well-written monograph. It is not a collection of dry summaries of articles but a book that explains and analyzes different approaches comparatively so that the reader can see how they are similar and how they differ. The literature survey is extensive. It contains 312 references, covering not only pattern recognition, machine learning, and statistics, but also economics and social choice (an author index would have been great). In many cases, there are handworked examples or results shown on 1-D or 2-D example data sets, which enhance understanding. The MATLAB code for many methods is given in chapter appendices allowing readers to play with the explained methods or apply them quickly to their own data; and that is just one more reason why this book makes a great contribution to the machine learning/pattern recognition literature.

Introduction to Stochastic Search and Optimization. Estimation, Simulation, and Control—James C. Spall (Hoboken, NJ: Wiley, 2003, pp. 595, ISBN: 0-471-33052-3). *Reviewed by W. Nowak*

Stochastic events fill our lives. It is hard to name any field of science, medicine, industry, or economy that does not built on extensive use of stochastic search and optimization algorithms. The reviewed book by J. C. Spell is an excellent graduate level introduction to this important part of applied mathematics and computer science. I was impressed by the truly general and interdisciplinary approach taken by the author. The stochastic methods are presented in the practical context of applications to numerous real-word problems: refining the design of aircraft, checking effectiveness of a new drug, making decisions to increase company profit, optimizing water-treatment plant, traffic management, etc.

The book is comprehensive—the main text has 504 pages divided into 17 chapters. In addition, five very useful (especially for newcomers in the field) and clearly written appendices are provided, on multivariate analysis, basic tests in statistics, probability theory and

The reviewer is with the Institute of Physics, Nicolaus Copernicus University, Torun 87-100, Poland (e-mail: wiesiek@fizyka.umk.pl).

Digital Object Identifier 10.1109/TNN.2007.897481