I.K. Sethi
A.K. Jain
Editors

Artificial Neural Networks and Statistical Pattern Recognition

Machine Intelligence and Pattern Recognition Series, volume 11

North-Holland

ARTIFICIAL NEURAL NETWORKS AND STATISTICAL PATTERN RECOGNITION

Machine Intelligence and Pattern Recognition

Volume 11

Series Editors

L.N. KANAL

and

A. ROSENFELD

University of Maryland
College Park, Maryland, U.S.A.

NORTH-HOLLAND
AMSTERDAM • LONDON • NEW YORK • TOKYO

Artificial Neural Networks and Statistical Pattern Recognition

Old and New Connections

Edited by

Ishwar K. SETHI
Wayne State University
Detroit, Michigan, U.S.A.

Anil K. JAIN
Michigan State University
East Lansing, Michigan, U.S.A.

1991

NORTH-HOLLAND
AMSTERDAM • LONDON • NEW YORK • TOKYO

ELSEVIER SCIENCE PUBLISHERS B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Distributors for the United States and Canada:
ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
655 Avenue of the Americas
New York, N.Y. 10010, U.S.A.

Library of Congress Cataloging-in-Publication Data

Artificial neural networks and statistical pattern recognition : old and new connections / edited by Ishwar K. Sethi, Anil K. Jain.
p. cm. -- (Machine intelligence and pattern recognition ; v. 11)
Includes bibliographical references and index.
ISBN 0-444-88740-7. -- ISBN 0-444-88741-5 (pbk.)
1. Pattern recognition systems. 2. Neural networks (Computer science) I. Sethi, Ishwar K., 1948- . II. Jain, Anil K., 1948- . III. Series.
TK7882.P3A7 1991
006.3--dc20 91-24621
CIP

ISBN: 0 444 88740 7 (hardbound)
0 444 88741 5 (paperback)

Printed in The Netherlands

FOREWORD

The first time I became interested in Neural Nets and Statistical Pattern Recognition was in early 1958 while I was a graduate student in the Moore School of Electrical Engineering of the University of Pennsylvania. My student subscription to the NEW YORKER magazine brought many chuckles from cartoons and stories but the only item from all those many issues that has stayed with me was a column in the December 6, 1958 issue titled "Rival". This covered an interview with Frank Rosenblatt in which he described his hopes for his "artificial intelligences" which would rival humans in perception and problem solving. By the time I read this column I knew a fair amount about Rosenblatt's research on Perceptrons, since as part of a machine learning and recognition research project and a search for a dissertation topic, I had spent much time pouring over his Cornell Aeronautical Laboratory reports. I had also read parts of a book *Stochastic Models for Learning* by Bush and Mosteller (Wiley, 1955) and been studying papers on Statistical Discrimination, in particular papers by C.R. Rao and the chapter on Problems of Discrimination in his book *Advanced Statistical Methods in Biometric Research* (Wiley , 1952). About the same time Robert Bush joined the University of Pennsylvania as chairman of Psychology. I chose Bush as my dissertation advisor, and with some support from R. Duncan Luce did a dissertation (for the Ph.D in electrical engineering!) on the analysis of some stochastic processes arising from Luce's nonlinear "Beta" model for learning. This is how learning models, artificial neural networks, and statistical pattern classification came together in my cognizance.

Two years later, when I joined General Dynamics/Electronics (GD/E) in Rochester, New York, as Manager of the Machine Intelligence Advanced Development Laboratory, it seemed as though every company and university laboratory was working on perceptron type machines. At GD/E we also implemented our own version of an adaptive pattern recognizer which was soon called APE (Adaptive Pattern Encoder). There were many other learning machines implemented by various organizations, machines with names such as MINOS, SOCRATES, and of course ADALINE and MADALINE. It was a time for catchy names and audacious claims [see Kanal, Proc. IEEE, October 1972]. Clearly PERCEPTRON and ADALINE were the key innovations and they had memorable names, although I have it on good authority that in the 1980' s when the new machine vision company Perceptron was formed, its founders had no idea that the name they had come up with had a previous incarnation. Because of simultaneous exposure to papers on learning models, perceptrons, and statistical discrimination, my attempts at understanding perceptrons and other "bionic" networks were formulated in terms of statistical classification methods, stochastic approximation procedures and stochastic models for learning. "Evaluation of a class of Pattern Recognition Networks" presented at the Bionics conference in Ithaca, N.Y in 1961 and reprinted in this book, summarized some of that understanding. It may seem surprising now, but at that time it had been stated by some of the well known researchers writing in the engineering literature on pattern recognition, that the use of a weighted sum of binary variables as done in the perceptron type classification function limited the variables to being statistically independent.

Rosenblatt had not limited himself to using just a single Threshold Logic Unit but used networks of such units. The problem was how to train multilayer perceptron networks. A paper on the topic written by Block, Knight and Rosenblatt was murky indeed, and did not demonstrate a convergent procedure to train such networks. In 1962-63 at Philco-Ford, seeking a systematic approach to designing layered classification nets, we decided to use a hierarchy of threshold logic units with a first layer of "feature logics" which were threshold logic units on overlapping receptive fields of the image, feeding two additional levels of weighted threshold logic decision units. The weights in each level of the hierarchy were estimated using statistical methods rather than iterative training procedures [L.N.Kanal & N.C. Randall, Recognition System Design by Statistical Analysis, Proc. 19th Conf. A.C.M,1964]. We referred to the networks as two layer networks since we did not count the input as a layer. On a project to recognize tanks in aerial photography, the method worked well enough in practice that the U.S. Army agency sponsoring the project decided to classify the final reports, although previously the project had been unclassified. We were unable to publish the classified results! Then, enamoured by the claimed promise of coherent optical filtering as a parallel implementation for automatic target recognition, the funding we had been promised was diverted away from our electro-optical implementation to a coherent optical filtering group. Some years later we presented the arguments favoring our approach, compared to optical implementations and trainable systems, in an article titled "Systems Considerations for Automatic Imagery Screening" by T.J.Harley, L.N. Kanal and N.C. Randall, which is included in the IEEE Press reprint volume titled *Machine Recognition of Patterns* edited by A. Agrawala. In the years which followed multilevel statistically designed classifiers and A.I. search procedures applied to pattern recognition held my interest, although comments in my 1974 survey,"Patterns In Pattern Recognition: 1968-1974" [IEEE Trans. on IT, 1974], mention papers by Amari and others and show an awareness that neural networks and biologically motivated automata were making a comeback.

In the last few years trainable multilayer neural networks have returned to dominate research in pattern recognition and this time there is potential for gaining much greater insight into their systematic design and performance analysis. Artificial neural networks trained on sample data are nonparametric statistical estimators for densities and classifiers. This leads to many questions about ANN's in comparison to alternate statistical methodologies. Such questions include the information requirements for each approach, the sample sizes for design and test, the robustness to incomplete data and different types of noise, and the generalization capability of competing procedures. Additional points of comparison concern the relations of the sizes of feature vectors for each pattern category; the capability for variable-length vector pattern recognition; the capability for fusion of multiple sources or sensors; the ability to incorporate domain knowledge; the ability to work with other pattern recognition paradigms in an integrated fashion; the ability of the methodology to extend to other types of problem solving, e.g., combinatorial optimization, resource allocation, etc., using the same general network architecture; the suitability for easy mapping to VLSI or other parallel architecture. The capability of neural networks to combine adaptation with parallelism in an easy and natural fashion and the ability of learning continuously while working on a problem in a real environment are of particular interest. Finally, the cost of implementation and of training personnel in the methodology will also be determiners of comparative success.

Some of the above questions are beginningto be addressed in the literature and the present volume is also a good start in this direction. I am thankful to Professors Anil K. Jain and Ishwar K. Sethi for their initiative in assembling and editing this volume and to the authors of each chapter for their contribution to this volume. The richness of the artificial neural network paradigm for pattern recognition ensures that, despite the many individuals working in this area, much work remains to be done to gain a true understanding of ANN methodologies and their relation to better understood pattern recognition methods. I expect that additional volumes will be assembled and published in this book series on the subject of artificial neural networks and their relation to and interaction with statistical pattern recognition, genetic algorithms, expert systems, and other approaches to the machine recognition of patterns.

Laveen N. Kanal
College Park, MD

This page intentionally left blank

PREFACE

Artificial neural networks (ANNs) are currently enjoying tremendous popularity across many disciplines. The interest in artificial neural networks is not new; it dates back to the work of McCulloch and Pitts, who, about forty years ago, proposed an abstract model of living nerve cells or neurons. Since then, the field of ANN has charted a bumpy course, with expectations running high in the late 50's with the publication of Rosenblatt's Perceptron model, and going down in the late 60's with the publication of Minsky and Pappart's book, "Perceptrons". While isolated interest in ANNs continued thereafter, a resurgence of interest came in the early 80's, with the work of Hopfield and his associates as well as that of Rummelhart and the parallel distributed processing (PDP) research group.

Two broad groups of researchers have been drawn to the study of artificial neural networks. The first group of researchers mainly consists of scientists who are interested in obtaining answers to some fundamental issues of perception, learning, and memory organization in the human brain through the study and development of different neuron models. The artificial neural network models espoused by this group of researchers are required to be as close as possible to the biological neural networks. The second group of ANN researchers is drawn mainly from the engineering community. These researchers are interested in exploiting the learning and parallel processing capabilities of the ANN to build engineering applications. This set of researchers takes a pragmatic approach towards various ANN models; they are not overly concerned about the closeness of the artificial neural systems with their biological counterparts.

Pattern recognition applications have emerged as the main beneficiary of the recent developments in ANNs. Pattern recognition tasks such as recognizing a familiar face or voice, identifying objects around us, or noticing relationships in a set of observations that we perform so effortlessly, have proven to be difficult in unconstrained settings for the traditional algorithmic approach, even using very powerful computers. In this respect, ANNs, with their self-organizing and non-algorithmic learning characteristics, offer a great deal of potential for pattern recognition applications.

The pattern recognition related activities using ANNs can be broadly grouped into two categories. The first group of activities consists of using the discriminatory or self-organizing features of various ANN models, such as multilayer perceptrons, neocognitron, ART series, Kohonen's self-organizing feature maps, etc. to build systems for recognizing different kinds of shapes, sounds and textures. Many such efforts have led to performance levels that are comparable or superior to the existing levels of performance achieved by traditional pattern recognition methods. The second group of pattern recognition related activities centers around mapping traditional pattern classifiers into ANN architectures. The aim of such mappings is to utilize some of the key features of ANN models to obtain better classification performance in terms of speed or error rate or both. Most of the commonly used classifiers, such as linear classifiers, quadratic classifiers, tree classifiers, nearest neighbor classifiers, have been exactly or approximately mapped into layered ANN architectures.

With the growing complexity of pattern recognition related problems which are being solved using ANNs, many ANN researchers are grappling with design issues such as the size of the network, the number of training patterns, and performance assessment and bounds. These researchers are discovering that many of the learning procedures lack the scaling property, i.e. these procedures simply fail or produce unsatisfactory results when applied to problems of bigger size. Phenomena like these are very familiar to researchers in statistical pattern recognition (SPR), where the "curse of dimensionality" is a well-known problem. Issues related to the training and test sample sizes, feature space dimensionality, and the discriminatory power of different classifier types have all been extensively studied in the SPR literature. It appears that many ANN researchers looking at pattern recognition problems are not aware of the ties between their field and SPR, and are therefore unable to successfully exploit the past work that has been done in SPR. Similarly, many pattern recognition and computer vision researchers do not realize the potential of the ANN approach to solve problems such as feature extraction, segmentation, and object recognition.

It is in the context of the above remarks that the idea for this volume originated; we are now delighted to share it with you. There are thirteen chapters in this volume, organized into three groups. The theme of the 5 chapters in the first group revolves around the connections between ANNs and SPR. Familiarity with work in each of these areas can lead to mutual benefit, as the study of ANNs and SPR share many common goals. The first chapter in this group is a paper by Kanal that was originally published in 1961. It is included here because it is one of the first papers, if not the first, to discuss the relationship between perceptrons and statistical classification methods, and also to relate learning algorithms to stochastic approximation methods. The second chapter, by Werbos, provides an overview of artificial neural networks research, especially the back-propagation algorithm for static as well as dynamic situations, and its linkage with statistical pattern recognition. The third chapter in the first group is by Raudys and Jain, who investigate the performance of artificial neural networks designed using only a small set of exemplars. The next two chapters, by Gelfand and Delp, and Sethi, deal with the relationship between tree classifiers and multiple layer perceptron networks.

The second group of 5 chapters is devoted to the application of neural networks to various pattern recognition problems involving image and speech data. The first chapter in this group, by Lee, Srihari and Gaborski, provides a theoretical relationship and an empirical comparison between the Bayes decision rule and the back-propagation network, using the problem of handwritten character recognition. In the second chapter of this group, Khotanzad and Lu examine the use of multiple layer perceptron networks for shape and texture recognition, and compare the performance of neural net classifiers with some conventional classifiers. The third chapter in the application group is by Ghosh and Bovik, who highlight the relationships between conventional and neural techniques for processing of textured images, and suggest discrete 2-dimensional Gabor transforms using a neural network implementation. The next chapter in this group, by Rangarajan, Chellappa and Manjunath looks at the relationship between the Markov random fields and neural networks. This relationship is examined in the context of early or low-level vision processing, suggesting some applications that might benefit from an approach that combines Markov random fields and neural networks. The last chapter in the application group is by Bengio and De Mori. After surveying the application of different neural models to automatic speech recognition, Bengio and De Mori present the details of using radial basis functions network for a particular speech recognition task. One consistent conclusion in all the five chapters of the application group is that neural net classifiers can serve as a good alternative to conventional classifiers. The maximum advantage in the use of neural classifiers occurs when the data are noisy and a large number of training samples are available.

The third section of the book deals with the implementation aspects of artificial neural networks. While the most widely used neural network implementations today are software simulators, it goes without saying that the full advantage of neural network paradigm cannot be reaped without hardware implementations. The first chapter in this section, by Hassoun, deals with the architecture of dynamic associative memories, a class of ANNs which utilizes supervised learning procedures to store information as stable memory states. Optical implementations of ANNs hold great promise. The second chapter, by Vijaya Kumar and Wong, describes such implementations for four associative memory schemes. The third and final chapter of the implementation section is on the VLSI implementation of neural networks. Salam, Choi and Wang provide an overview of the various issues related to the silicon implementation of neural nets. Some notable successes have already been achieved in the silicon implementation of biological devices, the most notable being the silicon retina and cochlea, by Mead and his group, at CalTech. According to Mead, analog VLSI neural networks running at 10 quadrillion operations per second are ultimately achievable.

This book could not have been completed without the whole-hearted cooperation of all the authors; we are thankful to them. We are also grateful to Professor Laveen Kanal for his constant encouragement and suggestions. We are also thankful to the Plenum Publishing Corporation for its kind permission to reprint Laveen Kanal's paper. Our sincere thanks also goes to the staff at Elsevier Science Publishers B.V. for their cooperation and support. We would consider our goal accomplished if this book is able to contribute in some way to greater interaction between the two communities of ANN and SPR researchers. It is to them that this book is dedicated.

Ishwar K. Sethi
Detroit, MI

Anil K. Jain
East Lansing, MI

This page intentionally left blank

TABLE OF CONTENTS

IMPLEMENTATION ASPECTS