Springer Series in Information Sciences 30

Editor: Teuvo Kohonen

Springer-Verlag Berlin Heidelberg GmbH

Springer Series in Information Sciences

Editors: Thomas S. Huang   Teuvo Kohonen   Manfred R. Schroeder

30 **Self-Organizing Maps**
By T. Kohonen   3rd Edition

31 **Music and Schema Theory**
Cognitive Foundations of Systematic Musicology
By M. Leman

32 **The Maximum Entropy Method**
By N. Wu

33 **A Few Steps Towards 3D Active Vision**
By T. Viéville

34 **Calibration and Orientation of Cameras in Computer Vision**
Editors: A. Grün and T. S. Huang

35 **Computer Speech**
Recognition, Compression, Synthesis
By M. R. Schroeder

Volumes 1–29 are listed at the end of the book.

Teuvo Kohonen

Self-Organizing Maps

Third Edition
With 129 Figures and 22 Tables

Springer

Professor Teuvo Kohonen

Helsinki University of Technology Neural Networks Research Centre
P.O. Box 5400
02015 HUT, Espoo, FINLAND

Series Editors:

Professor Thomas S. Huang

Department of Electrical Engineering and Coordinated Science Laboratory,
University of Illinois, Urbana, IL 61801, USA

Professor Teuvo Kohonen

Helsinki University of Technology Neural Networks Research Centre
P.O. Box 5400
02015 HUT, Espoo, FINLAND

Professor Dr. Manfred R. Schroeder

Drittes Physikalisches Institut, Universität Göttingen, Bürgerstrasse 42-44,
37073 Göttingen, GERMANY

ISSN 0720-678X
ISBN 978-3-540-67921-9

Library of Congress Cataloging-in-Publication Data applied for.

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Kohonen, Teuvo:
Self-organizing maps / Teuvo Kohonen.– 3. ed. – Berlin; Heidelberg; New York; Barcelona; Hong Kong; London; Milan; Paris; Singapore; Tokyo: Springer, 2001
(Springer series in information sciences; 30)
ISBN 978-3-540-67921-9 ISBN 978-3-642-56927-2 (eBook)
DOI 10.1007/978-3-642-56927-2

Originally published by Springer-Verlag Berlin Heidelberg New York in 2001

Typesetting: Data conversion by K. Mattes, Heidelberg
Cover design: *design & production* GmbH, Heidelberg
Printed on acid-free paper 56/3111 - 5 4 3 2

Preface to the Third Edition

Since the second edition of this book came out in early 1997, the number of scientific papers published on the Self-Organizing Map (SOM) has increased from about 1500 to some 4000. Also, two special workshops dedicated to the SOM have been organized, not to mention numerous SOM sessions in neural-network conferences. In view of this growing interest it was felt desirable to make extensive revisions to this book. They are of the following nature.

Statistical pattern analysis has now been approached more carefully than earlier. A more detailed discussion of the eigenvectors and eigenvalues of symmetric matrices, which are the type usually encountered in statistics, has been included in Sect. 1.1.3; also, new probabilistic concepts, such as factor analysis, have been discussed in Sect. 1.3.1. A survey of projection methods (Sect. 1.3.2) has been added, in order to relate the SOM to classical paradigms.

Vector Quantization is now discussed in one main section, and derivation of the point density of the codebook vectors using the calculus of variations has been added, in order to familiarize the reader with this otherwise complicated statistical analysis.

It was also felt that the discussion of the neural-modeling philosophy should include a broader perspective of the main issues. A historical review in Sect. 2.2, and the general philosophy in Sects. 2.3, 2.5 and 2.14 are now expected to especially help newcomers to orient themselves better amongst the profusion of contemporary neural models.

The basic SOM theory in Chap. 3 has now first been approached by a general qualitative introduction in Sect. 3.1. Other completely new concepts discussed in Chap. 3 are the point density of the model vectors (Sect. 3.12) and the interpretation of the SOM mapping (Sect. 3.16).

Only modest revisions have been made to Chap. 4.

Among the new variants in Chap. 5, the SOM of symbol strings (and other nonvectorial items) has been discussed in Sect. 5.7, and a generalization of the SOM in the direction of evolutionary learning has been made in Sect. 5.9.

To Chap. 6, the batch-computation scheme of the LVQ1 has been added.

In Chap. 7, a major revision deals with a new version of WEBSOM, the SOM of large document files, by which it has been possible to implement one of the most extensive ANN applications ever, namely the SOM of seven million patent abstracts. The amount of text thereby mapped is 20 times that of the Encyclopaedia Britannica!

The most visible and most requested addition to the third edition is the new Chap. 8 on software tools which we hope will be useful for practitioners.

It was not possible, however, to extend the survey of new SOM applications much beyond that already published in the second edition. A new hardware implementation has been discussed at the end of Chap. 9, but the main change made to the literature survey in Chap. 10 is its reorganization: the taxonomy and indexing of its contents is now more logical than in the second edition.

In the Preface to the first edition I gave advice for the first reading of this book. Since the numbering of the sections has now changed, the new recommended reading sequence for a short introductory course on the SOM is the following: 2.2, 2.9, 2.10, 2.12, 3.1–3, 3.4.1, 3.5.1, 3.6, 3.7, 3.13, 3.14, 3.15, 6.1, 6.2 (skipping the derivation), 6.4–9, 7.1 and 7.2.

Again I got much help in the revision from my secretary, Mrs. Leila Koivisto; I am very much obliged to her. Some parts of the word processing were done by Mr. Mikko Katajamaa. Naturally I have benefited from the good scientific work done by all my younger colleagues in our university.

Dr. Samuel Kaski has kindly allowed me to use in Sect. 1.3.2 material compiled by him.

After writing all three editions, I want to once again express my gratitude to the Academy of Finland for allowing me to carry out this activity under their auspices.

Espoo, Finland *Teuvo Kohonen*
October, 2000

Preface to the Second Edition

The second, revised edition of this book came out sooner than originally planned. The main reason was that new important results had just been obtained.

The ASSOM (Adaptive-Subspace SOM) is a new architecture in which invariant-feature detectors emerge in an unsupervised learning process. Its basic principle was already introduced in the first edition, but the motivation and theoretical discussion in the second edition is more thorough and consequent. New material has been added to Sect. 5.9 and this section has been rewritten totally. Correspondingly, Sect. 1.4 that deals with adaptive-subspace classifiers in general and constitutes the prerequisite for the ASSOM principle has also been extended and rewritten totally.

Another new SOM development is the WEBSOM, a two-layer architecture intended for the organization of very large collections of full-text documents such as those found in the Internet and World Wide Web. This architecture was published after the first edition came out. The idea and results seemed to be so important that the new Sect. 7.8 has now been added to the second edition.

Another addition that contains new results is Sect. 3.15, which describes acceleration of computing of very large SOMs.

It was also felt that Chapter 7, which deals with SOM applications, had to be extended.

To recapitulate, Chaps. 1, 3, 5, 7, and 9 contain extensive additions and revisions, whereas Chaps. 2, 4, 6, 8, and 10 are identical with those of the first edition, let alone a few minor corrections to their text.

In the editing of this revision I got much help from Mr. Marko Malmberg.

Espoo, Finland
September, 1996

Teuvo Kohonen

Preface to the First Edition

The book we have at hand is the fourth monograph I wrote for Springer-Verlag. The previous one named "Self-Organization and Associative Memory" (Springer Series in Information Sciences, Volume 8) came out in 1984. Since then the self-organizing neural-network algorithms called SOM and LVQ have become very popular, as can be seen from the many works reviewed in Chap. 9. The new results obtained in the past ten years or so have warranted a new monograph. Over these years I have also answered lots of questions; they have influenced the contents of the present book.

I hope it would be of some interest and help to the readers if I now first very briefly describe the various phases that led to my present SOM research, and the reasons underlying each new step.

I became interested in neural networks around 1960, but could not interrupt my graduate studies in physics. After I was appointed Professor of Electronics in 1965, it still took some years to organize teaching at the university. In 1968–69 I was on leave at the University of Washington, and D. Gabor had just published his convolution-correlation model of autoassociative memory. I noticed immediately that there was something not quite right about it: the capacity was very poor and the inherent noise and crosstalk were intolerable. In 1970 I therefore suggested the autoassociative correlation matrix memory model, at the same time as J.A. Anderson and K. Nakano.

The initial experiences of the application of correlation matrix memories to practical pattern recognition (images and speech) were somewhat discouraging. Then, around 1972–73, I tried to invert the problem: If we have a set of pairs of input-output patterns, what might be the *optimal* transfer matrix operator in relation to smallest residual errors? The mathematical solution that ensued was the optimal associative mapping, which involved the Moore-Penrose pseudoinverse of the input observation matrix as a factor. As an associative memory, this mapping has a capacity three times that of the networks discussed by J. Hopfield. The *recognition accuracy* for natural data, however, was essentially not improved, even when we used nonlinear (polynomial) preprocessing! Obviously there was still something wrong with our thinking: associative memory and pattern recognition could not be the same thing!

During 1976–77 I had a new idea. In the theory of associative memory I had worked, among other problems, with the so-called Novelty Filter, which is an adaptive orthogonal projection operator. I was trying to conceive a neuron that would represent a whole *class* of patterns, and I was playing with the idea of a neuron, or rather a small set of interacting neurons describable as a Novelty Filter. If that would work, then an arbitrary linear combination of the stored reference patterns would automatically belong to the same class (or manifold). It turned out that the so-called linear-subspace formalism known in pattern recognition theory was mathematically equivalent to my idea. Then I went one step further: since according to our own experiments the basic subspace method was still too inaccurate for classification, how about trying some *supervised* "training" of the subspaces, or their basis vectors? I soon invented *the first supervised competitive-learning algorithm*, the Learning Subspace Method (LSM), and it worked almost three times more accurately than the previous ones! Its handicap was a slower speed, but we developed a fast co-processor board to cope with "neural network" computations to make the LSM work in real time. We based our first speech-recognition hardware system on that idea, and this algorithm was in use for several years. Half a dozen Ph.D. theses were done on that system in our laboratory.

Our research on the Self-Organizing Map (SOM) did not begin until in early 1981, although I had already jotted down the basic idea into my notebook in 1976. I just wanted an algorithm that would effectively map similar patterns (pattern vectors close to each other in the input signal space) onto contiguous locations in the output space. Ch. v.d. Malsburg had obtained his pioneering results in 1973, but I wanted to generalize and at the same time ultimately simplify his system description. We made numerous similarity (clustering) diagrams by my simplified but at the same time very robust SOM algorithm, including the map of phonemes. When we tentatively tried the SOM for speech recognition in 1983, we at first got no improvement at all over that already achieved by LSM. Then, in 1984, I again thought about supervised learning, and the Supervised SOM described in Sect. 5.8 solved the problem. We had developed an algorithm that was best so far.

During 1985–87 our laboratory had a cooperative project on speech recognition with Asahi Chemical Co., Ltd., the biggest chemical company in Japan. During the first phase I introduced two new algorithms (based on research that I had started a couple of years earlier): the Learning Vector Quantization (LVQ), which is a supervised version of SOM particularly suitable for statistical pattern recognition, and the Dynamically Expanding Context (DEC), both of which will be described in this book. For many years thereafter they formed the basis of our speech recognition systems.

Over the years we worked on numerous other practical applications of the SOM, and these projects will be found among the references.

The present monograph book contains a brand-new, so far unpublished result, the *Adaptive-Subspace SOM (ASSOM)*, which combines the old Learning Subspace Method and the Self-Organizing Map. It does also something more than most artificial neural network (ANN) algorithms do: it detects *invariant* features, and to achieve, e.g., translational invariance to elementary patterns, the "winner take all" function had to be modified fundamentally. The sophisticated solutions described in Sects. 5.9 and 5.10 could not have been invented at once; they had to be acquired during a long course of development from many small steps. The ideas of "representative winner" and "competitive episode learning" had not been conceived earlier; with these ideas the generally known wavelet and Gabor-filter preprocessing can now be made to emerge automatically.

This book contains an extensive mathematical introduction as well as a Glossary of 555 relevant terms or acronyms. As there exists a wealth of literature, at least 1500 papers written on SOM, it was also felt necessary to include a survey of as many works as possible to lead the readers to the newest results and save them from painstaking hours in the library.

I have to say quite frankly, however, that the SOM and LVQ algorithms, although being potentially very effective, are not always applied in the correct way, and therefore the results, especially benchmarkings reported even in respectable journals and by respectable scientists, are not always correct. I felt it necessary to point out *how* many details have to be taken into account, before the problem is approached in the proper way. Let me also emphasize the following facts: 1) The SOM has not been meant for statistical pattern recognition; it is a clustering, visualization, and abstraction method. Anybody wishing to implement decision and classification processes should use LVQ in stead of SOM. 2) Preprocessing should not be overlooked. The ANN algorithms are no sausage machines where raw material (data) is input at one end and results come out at the other. Every problem needs a careful selection of feature variables, which so far is mostly done by hand. We are just at the dawn of automated feature extraction, using ANN models such as ASSOM. 3) In benchmarkings and practical problems one should compare the speed of computation, too, not only ultimate accuracies. A relative difference in accuracy of a few per cent can hardly be noticed in practice, whereas tiny speed differences during the actual operation are very visible.

People who are reading about the SOM for the first time may feel slightly uneasy about its theoretical discussion: so sophisticated, and yet only leading to partial results. Therefore, let me quote three well-known French mathematicians, Professors M. Cottrell, J.-C. Fort, and G. Pagès: "Despite the large use and the different implementations in multi-dimensional settings, the Kohonen algorithm is surprisingly resistant to a complete mathematical study." Perhaps the SOM algorithm belongs to the class of "ill posed" problems, but so are many important problems in mathematics. In *practice* people have applied many methods long before any mathematical theory existed for them and even if none may exist at all. Think about walking; theoretically we know that we could not walk at all unless there existed gravity and friction, by virtue of which we can kick the globe and the other creatures on it in the opposite direction. People and animals, however, have always walked without knowing this theory.

This book is supposed to be readable without any tutor, and it may therefore serve as a handbook for people wanting to apply these results in practice. I have tried to anticipate many problems and difficulties that readers may encounter, and then to help them clear the hurdles.

Can this book be used as a university textbook? Yes, many efforts have been made to this end so that it could be useful. It should serve especially well as collateral reading for neural-network courses.

If only a short introductory course on the SOM is given, the teaching material might consist of the following sections: 2.6–8, 2.12, 3.1, 3.2, 3.3.1, 3.4.1, 3.5, 3.9, 3.10, 3.11, 6.1, 6.2 (thereby skipping the derivation of Eq. (6.6)), 6.4–7, 7.1, and 7.2. This is also the recommendable sequence for the first reading of this book.

To a lecturer who is planning a whole special course on the SOM, it might be mentioned that if the audience already has some prior knowledge of linear algebra, vector spaces, matrices, or systems theory, Chapter 1 may be skipped, and one can start with Chapter 2, proceeding till the end of the book (Chap. 8). However, if the audience does not have a sufficient mathematical background, Chapter 1 should be read meticulously, in order to avoid trivial errors in the application of the algorithms. Matrix calculus has plenty of pitfalls!

Acknowledgements. This book would have never been completed without the generous help of many people around me. First of all I would like to mention Mrs. Leila Koivisto who did a very big job, magnificiently typing out my pieces of text and then had the stamina to produce the numerous revisions and the final layout. The nice appearance of this book is very much her achievement.

Dr. Jari Kangas should be mentioned for being responsible for the extensive library of literature references on SOM and LVQ. His efforts made it possible to include the References section in this book.

I would like to thank the following people for their help in making many simulations, part of which have also been published as research papers or theses. Mr. Samuel Kaski simulated the physiological SOM model. Mr. Harri Lappalainen helped me in the Adaptive-Subspace SOM (ASSOM) research.

Mr. Ville Pulkki collected works on hardware implementations of SOM, and I have used some of his illustrations. I have also used the nice pseudocolor image of cloud classification that was obtained in the research project led by Dr. Ari Visa. Some other results have been utilized, too, with due credits given in the corresponding context.

The following people have helped me in word processing: Taneli Harju, Jussi Hynninen, and Mikko Kurimo.

I am also very much obliged to Dr. Bernard Soffer for his valuable comments on this text.

My research was made possible by the generous support given to our projects by the Academy of Finland and Helsinki University of Technology.

Espoo, Finland
December, 1994

Teuvo Kohonen

Contents