Studies in Systems, Decision and Control 131

Paolo Massimo Buscema
Giulia Massini · Marco Breda
Weldon A. Lodwick · Francis Newman
Masoud Asadi-Zeydabadi

Artificial Adaptive Systems Using Auto Contractive Maps

Theory, Applications and Extensions

Springer

Studies in Systems, Decision and Control

Volume 131

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

The series “Studies in Systems, Decision and Control” (SSDC) covers both new developments and advances, as well as the state of the art, in the various areas of broadly perceived systems, decision making and control- quickly, up to date and with a high quality. The intent is to cover the theory, applications, and perspectives on the state of the art and future developments relevant to systems, decision making, control, complex processes and related areas, as embedded in the fields of engineering, computer science, physics, economics, social and life sciences, as well as the paradigms and methodologies behind them. The series contains monographs, textbooks, lecture notes and edited volumes in systems, decision making and control spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

More information about this series at http://www.springer.com/series/13304

Paolo Massimo Buscema
Giulia Massini · Marco Breda
Weldon A. Lodwick · Francis Newman
Masoud Asadi-Zeydabadi

Artificial Adaptive Systems Using Auto Contractive Maps

Theory, Applications and Extensions

Springer

Paolo Massimo Buscema
Semeion Research Center of Sciences of Communication
Rome
Italy

Giulia Massini
Semeion Research Center of Sciences of Communication
Rome
Italy

Marco Breda
Semeion Research Center of Sciences of Communication
Rome
Italy

Weldon A. Lodwick
Department of Mathematical and Statistical Sciences
University of Colorado Denver
Denver, CO
USA

Francis Newman
Department of Radiation Oncology, School of Medicine
University of Colorado Denver
Denver, CO
USA

Masoud Asadi-Zeydabadi
Physics Department
University of Colorado Denver
Denver, CO
USA

ISSN 2198-4182 ISSN 2198-4190 (electronic)
Studies in Systems, Decision and Control
ISBN 978-3-319-75048-4 ISBN 978-3-319-75049-1 (eBook)
https://doi.org/10.1007/978-3-319-75049-1

Library of Congress Control Number: 2018930135

© Springer International Publishing AG, part of Springer Nature 2018
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Chapter 1
An Introduction

> *Nearly all persons look not at theory from the standpoint of established facts, you know, but at established facts from the standpoint of theory; they cannot get beyond an assumed conceptual net they have accepted, …*
> Einstein (letter to Schroedinger, 8 August 1935, in Albrecht Fölsing, *Albert Einstein, Eine Biographie*, page 780, Suhrkamp Verlag, 1993)

Abstract Auto-Contractive Maps (Auto-CM) is a newer approach to artificial adaptive systems (AASs). In turn, AASs encompass the subject of artificial neural networks (ANNs). This chapter is an introduction to AASs.

1.1 Introduction

Artificial adaptive systems (AASs) are data driven systems. This means that AASs try to elicit, extract, from data the underlying model of cause/effect. These methods stand in counter distinction to approaches that impose a model on the data as a means of determining the relationship (cause and effect for example) among the data elements. Informally speaking, for this monograph, we consider *data* as being what is known about the system or problem, its input elements. The philosophy of AAS can be described as and compared to natural/human language. Some parallels are striking. The artificial sciences try to create models of reality, but how well they approximate the "world" determines their effectiveness and usefulness. Human languages, in a similar manner, try to approximate the reality of the subject at hand. There is another similarity between artificial adaptive systems and human languages and that is that both undergo dynamic changes to more closely resemble the "reality" of the entities of interest.

Auto Contractive Maps that is the main topic of this book is an artificial adaptive system, an artificial neural network, that adheres to the approach that Einstein mentions in his letter to Schroedinger. Auto-CM extracts the model from the data as will be apparent in our subsequent presentation beginning with Chap. 3 of this monograph.

P. M. Buscema et al., *Artificial Adaptive Systems Using Auto Contractive Maps*, Studies in Systems, Decision and Control 131, https://doi.org/10.1007/978-3-319-75049-1_1

Artificial adaptive systems include much of artificial intelligence and artificial neural networks (ANNs). This monograph restricts itself to one particular type of unsupervised ANN called Auto-Contractive Map (Auto-CM) and its supervised version called K-Auto CM. The Auto-CM approach is a newer type of ANN that is powerful enough as a stand-alone analytical tool. Auto-CM takes a different approach to ANNs in that its analytical method is more akin to fixed-point algorithms where, in our case, the values at the nodes converge to zero distributing their input dataset values to the weights. The final output of Auto-CM indicates a relationship among the variables of the dataset and these are found in the weights. Traditional ANNs attempt to stabilize the weights via some error minimization techniques. The weights of Auto-CM are also stabilized in the sense that the algorithm stops when all the node values are essentially zero but this occurs because a fixed-point has been reached rather than a minimum of an error function, for example, has been attained.

We not only present Auto-CM but show how to couple Auto-CM to two associated graphical visualization components, the minimal spanning tree (MST) and maximal regular graph (MRG) in order to graphically (in the graph theoretic sense) depict the relationships in the dataset. In addition to these two visualization tools and their interpretation of the underlying relationship in the data, we present methods to transform the output of the neural network (our Auto-CM output) into new datasets that allow for deeper interpretation of the data relationships akin to *deep learning* methods. In particular, the original dataset may be transformed via Auto-CM into new databases that have a richer set of relationships among variables and records; for example, fuzzy relationships and relationships that “collapse” variables and records. Our point of view is that the universe is not random and at the same time it loves to make us work hard to understand her. Thus, ferreting out the “patterns” that are in data is not easy. The ferreting out of patterns is what Auto-CM attempts to do. Auto-CM attempts to uncover nature’s secrets.

What we present targets users, researchers and students of ANNs who wish to see what a more ample approach to ANNs might look like. Only basic notions of mathematics are assumed as well as an understanding of neural networks, though we derive key features. Our level of presentation is at the upper level undergraduate to lower level graduate student in computer science, mathematics, physics or engineering.

This monograph continues this section with a general discussion of where Auto-CM fits within the field of artificial intelligence. Chapter 2 has an overview of the standard ANNs in order to clearly contrast the features of Auto-CM. Chapter 3 contains the core of our monograph, Auto-CM, an unsupervised type of ANN. Chapter 4 presents the graphical component of Auto-CM. Chapter 5 looks at how the Auto-CM output can be further transformed into new datasets that contain a wider view of the relationships that exist among data elements. Chapter 6 presents a supervised version of Auto-CM called K-Contractive Map. Chapter 7 compares Auto-CM with various ANNs. We end this monograph with Chap. 8 which has more advanced notions of how to add dynamic changes in time and spaces, to Auto-CM and hence ANNs.

1.2 What Are Artificial Adaptive Systems

Our introduction begins with the differences between deterministic/stochastic and adaptive approaches. Deterministic/stochastic systems are often called complicated systems whereas adaptive systems are often called complex systems.

A system is a connected region of space-time, divided into components, whose local and parallel interactions determine the functioning of the system itself. A system, whose operation (interactions between its components) dynamically changes the structure and/or the status of its members, while maintaining its space-time cohesion, is what we call a *complex system*. A human cell that becomes a human or a social organism, such as a group of scholars that becomes an academic department, is a complex system.

A system whose complete permutation of all of its components allows the definition of all possible states of the system is a *complicated system*, the more complicated the more numerous are its possible states. For example, a jumbo jet aircraft, with all of its 200,000 small components, is complicated but is not a complex system. A movie is a complicated system. A complicated system is a system whose operation does not generate new information since every possible trajectory is defined a priori.

A complicated system, which works over time, ages since its performance remains static. In a complex system the global behavior of the system is not inferred from the simple sum of the behaviors of all of its components, or by linear interpolation techniques. The overall behavior of a complex system is behavior that emerges from the operation of the system *over time*. This process is therefore highly non-linear, which in many cases is not formalized by equations in closed form. A complex system is one whose operation generates new information over time, and over time changes the structure and/or the state of its components, which we call an *adaptive system*.

An artificial system is a model of a part of reality that is encoded in mathematical relationships and implemented on a computer. Our view of an artificial adaptive system is depicted in Fig. 1.1 below where evolutionary programming and ANNs fall under the umbrella of AASs. Most of the features of Fig. 1.1 will be discussed in Chap. 2.

The dynamics of an adaptive system necessitates a mathematics that does not impose linear assumptions on the data. The mathematics that is needed works on data using a "Socratic" style. Ideally, it works in a way in which the overall behavior of the system emerges spontaneously by the local interaction of its components represented by *data and equations*. This is the method of a "natural" operation, from the identification of the targeted problem up to the interpretation of the results. To achieve this, we look at problems as research questions on which we undertake experiments via computer algorithms, simulations. This is the method of "Natural Computation" (Bottom-Up) and it is the best way for the study of natural adaptive systems (Buscema 2010). AASs are the mathematical expression of the Natural Computation method. Their purpose is to bring out the overall operation of a natural adaptive system, whose data represent discrete portions of the functioning system.

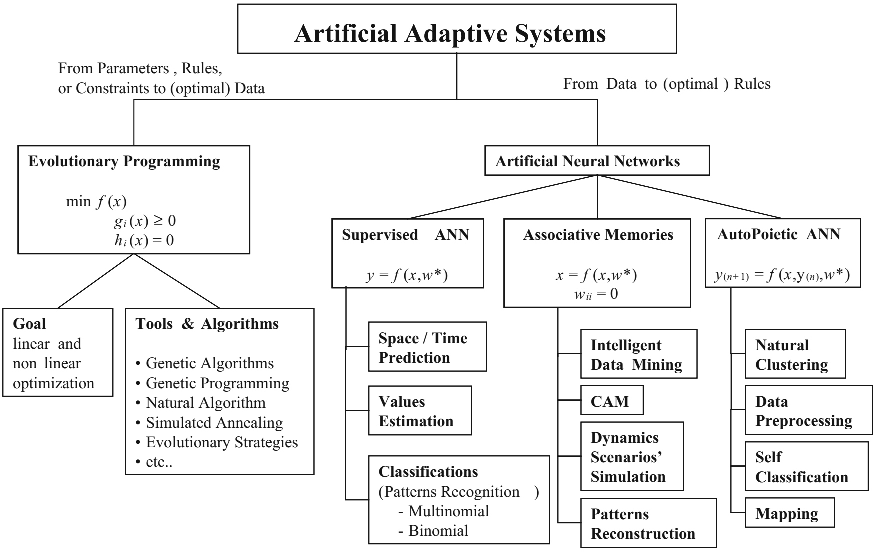

Fig. 1.1 Artificial adaptive systems

We will use the expression "cultural system" to mean a natural system whose operation is filtered in a non-linear interaction between human brains. In natural systems and, therefore, also cultural ones, there is hidden information that lies within the data and via AASs can be disclosed and become perceptible among the relationships discovered in the data, and its information which is more visible and easily measureable. The hidden information reveals the "secret plan" that the natural adaptive system is going to pursue.

Artificial Adaptive Systems look for traces of the hidden agenda of natural adaptive systems. The range of these applications is also related to Intelligent Data Mining ([1] and see Fig. 1.2).

AASs emerge by subjecting the system to experiments. An "experiment" is the design, execution, and completion of data processing from the initial treatment of a selected or a complete set of available data including the interpretation of the local results. By "simulation" we mean the process that starts from the application of the different artificial adaptive algorithms to the data, up to the generation of results (see Fig. 1.3). The category "research" in Fig. 1.3 may be considered to be related to the same problem as different facets or they may be viewed as independent.

Figure 1.3 illustrates what we consider to be a robust research protocol and it is how we validated all our algorithms and examples. An AAS can be thought of being implemented as follows. We assume we have identified the problem, which means we have collected significant and non-trivial questions and facts about a specific natural and/or cultural system of interest. It also means that we have speculated about the possible operation of such a system, that is, we have formed hypotheses which can be methodologically verified in the following steps.

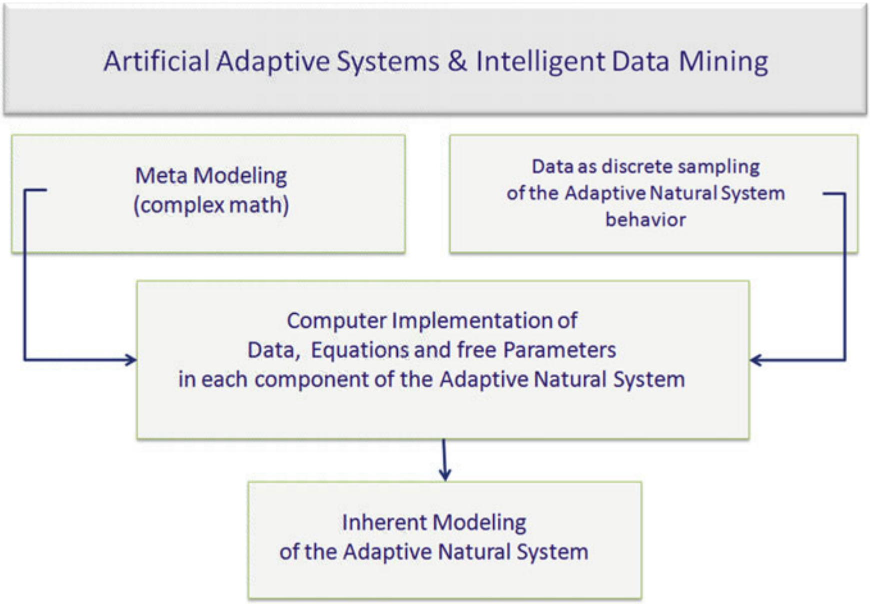

Fig. 1.2 The main components of Intelligent Data Mining schematized in order to understand natural adaptive systems

1. **The Data**. The study's questions and/or hypotheses need to be translated into structured input entities or data. That is, data is what we know about the system. Such data must be statistically representative of the process that we want to examine. The data must be collected with appropriate statistical criteria and/or must be generated by special experiments/trials. In all cases, the data that is available will be the only input for the next steps of the research process. All of the research has to be completely data driven.
2. **Data Pre-processing**. The data available for research should be organized and structured with appropriate algorithms, in order to create a database presenting all the information in an explicit format. Consequently, adaptive algorithms that will be used later, will facilitate the discovery of distributed and/or hidden information and structures in the database.
3. **Adaptive Algorithms**. Choose the AAS algorithms that are appropriate for the research and arrange them in sequential and/or parallel structures. The choice and the organization of the algorithms in a framework define research design.
4. **Blind Validation Methods**. The individual algorithms and the research design must be validated in a blind fashion, splitting the data numerous times into training and testing data subsets. There are various protocols available to implement the research's validation phase, and the most appropriate protocol has to be chosen for the assigned research. In addition, a suitable cost function that measures the

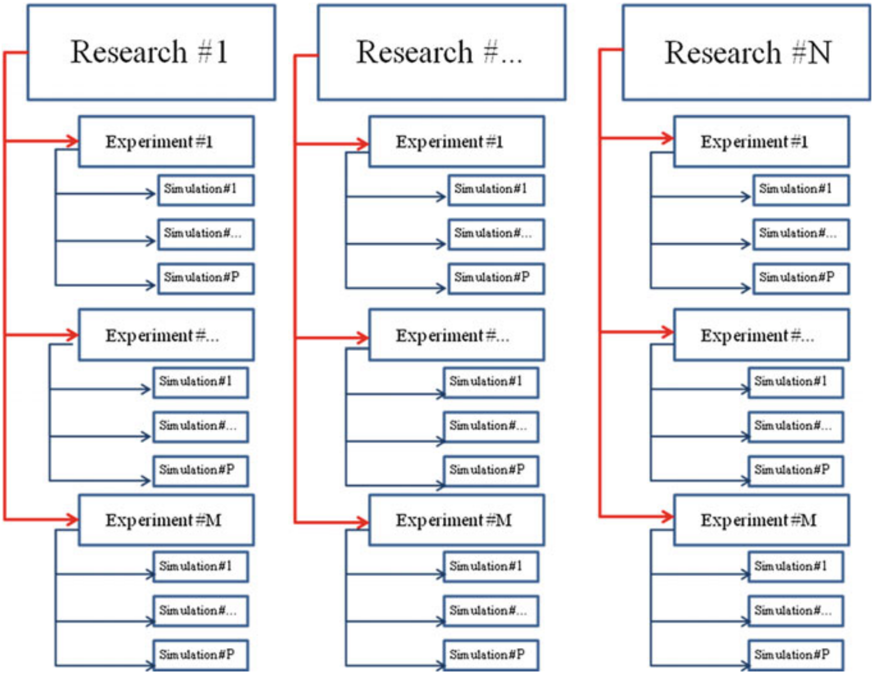

Fig. 1.3 Research, experiment and computer simulation dynamics

results of the different simulations has to be selected, according to statistical criteria.

5. **Results Interpretation**. This phase is very delicate. We need to interpret the different outputs of the experiment in relation to the questions and assumptions we made in the initial phase of the research when the experimental simulation proves itself to be significantly correct. This phase should lead to new hypotheses and/or questions that allow the experimental cycle to start from the beginning. These new hypotheses might be based on comparisons with known results or on predictions that were made requiring re-evaluation of the assumptions.
6. **Storage of New Information**. The storage of the results obtained, and their interpretation, is new knowledge that was not present in the initial phase of the research. This knowledge should be structured and organized in a new database, upgradeable over time, and can possibly be the object of further experiments.

The content of any research should not to be constrained a priori for two reasons. First, data must be the only source of knowledge of the targeted problem for the adaptive algorithms. In other words, any constraints and/or extra knowledge about the problem, has to be expressed in the form of quantitative and/or qualitative data. Second, artificial adaptive systems are equipped with a specific capacity to learn the complex dynamics of any process only from the data. This data have to be provided in a statistically meaningful manner. AASs, therefore, are algorithms that do not follow specific a priori rules. They do not optimize the relationships between linear data. Artificial adaptive systems *learn* directly from the data and they *evolve* in order to define the mathematical function that interpolates the assigned data in an optimal or near-optimal way. In this sense, artificial adaptive systems can be called "Meta-Models", because they learn the *intrinsic model* from the data.

1.3 The Components

The components of a research problem involving AASs can be briefly described by three components (see Fig. 1.4):

1. Databases;
2. Artificial Adaptive System Algorithms;
3. Validation Protocols.

Some recommended strategies in carrying out research from the point of view of an AAS process are:

1. Carry out many different experimentations that fail, in order to "grow up"; failure is a necessary condition and an opportunity to be embraced for needed development, innovation and changes;
2. Learn to consider and to look at any process from the same point of view that nature has adopted in order to generate that process;
3. Pay attention to the relationships among the objects, because relationships among objects are perceived before the objects, for example, when you see two points

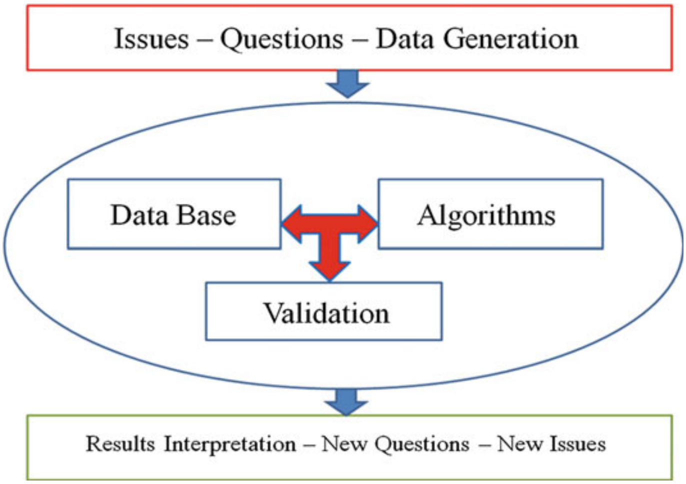

Fig. 1.4 The research process using artificial adaptive system methodology

in the space, your gaze is the relationship *between* the two points, without your gaze, the two points may not be linked together;

4. "Listen" to the weak links among things because the weak links often explain the strong(er) ones, for example, there seems to be a weak link between the head and the feet. But many headaches depend upon the way in which we move our feet and our posture;
5. Use mathematics to transform any random "scribble" into a possible pattern, because nature may in certain circumstances try to ignore noise. An addative inverse relationship between variance and entropy may exist in certain circumstances [2].

In short, what we invite the reader to consider and to explore in the following pages is achieved by using only three aspects:

a. Data of the process targeted to be investigated, the "problem";
b. The use of AAS algorithms in computer simulations of the problem;
c. Blind validation tests.

Another aspect that helps during experiments or research trials is the capability to be completely *coherent* inside every experiment and completely *incoherent* going from one experiment to another. That is, to test a theory one needs to be coherent and analytic within each experiment, then "jump" to a completely different experiment in order to test the same theory.

We end this section by reiterating that the topic of interest to this monograph is the unsupervised ANN called Auto-CM and its supervised version together with their connections to graphs and database transformations. Therefore, we will, in the next chapter, briefly review ANNs.

Specific References

1. Buscema, M., William J. Tastle. 2013. Data Mining Applications Using Artificial Adaptive Systems, Springer.
2. Swanson, R., and S.M. Swanson. 1993. The Effect of Noise on Entropy. *Acta Crystallographica. Section D, Biological Crystallography* 49 (Pt 1): 182–185.

General References

Anderson J.A., and E. Rosenfeld (eds.). 1988. *Neurocomputing Foundations of Research*, Cambridge, MA: The MIT Press.

Arbib, M.A. (ed.). 1995. *The Handbook of Brain Theory and Neural Networks, A Bradford Book*. Cambridge, Massachusetts, London, England: The MIT Press.

Bengio, Y. 2009. Learning Deep Architectures for AI. *Machine Learning* 2 (1): 1–127.

Buscema, M. 1998. Artificial Neural Networks and Complex Social Systems. In *Substance Use & Misuse*, vol. 33(1): Theory, vol. 33(2): Models, vol. 33(3): Applications, New York: Marcel Dekker.

Buscema, M. 1999. Semeion Group. In *Reti Neurali Artificiali e Sistemi Sociali Complessi*, vol. I: Teoria e Modelli, vol. II: Applicazioni, Franco Angeli, Milano.

Carpenter, G.A., and S. Grossberg. 1991. *Pattern Recognition by Self-Organizing Neural Network*. Cambridge, MA: MIT Press.

Grossberg, S. 1978. How Does the Brain Build a Cognitive Code? *Psychological Rewiew* 87.

Hinton, G.E., S. Osindero, and Y-W. Teh. 2006. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation* 18 (7): 1527–1554 (MIT Press Journal).

Hinton, G.E., and R.R. Salakhutdinov. 2006. Reducing the Dimensionality of Data with Neural Networks. *Science* 313: 504–507.

Hopfield, J.J. 1987. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. In *Proceedings of the National Academy of Sciences* 79, in Anderson 1988.

Hopfield, J.J. 1984. Neurons with Graded Response have Collective Computational Properties Like Those of Two-State Neurons. In *Proceedings of the National Academy of Sciences USA, Bioscience*, 81.

Hopfield, J.J., and D.W. Tank. 1985. Neural Computation of Decisions in Optimization Problems. *Biological Cybernetics* 52.

Hopfield, J.J., and D.W. Tank. 1986. Computing with Neural Circuits: A Model. *Articles Science* 233: 8.

Kohonen T. 1972. Correlation Matrix Memories. *IEEE Transactions on Computers C-21*, in Anderson 1988.

Larochelle, H., and Y. Bengio. 2008. Classification using Discriminative Restricted Boltzmann Machines. In *Proceedings of the 25-th International Conference on Machine Learning*, Helsinki, Finland.

Le, Q.V., M.C. Ranzato, R. Monga, M. Devin, K. Chen, G.S. Corrado, J. Dean, and A.Y. Ng. 2012. Building High-level Features Using Large Scale Unsupervised Learning, In *Proceedings of the 29-th International Conference on Machine Learning*, Edinburgh, Scotland, UK.

Maturana, H., and F. Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living*, Springer Science.

Minsky M. 1954. Neural Nets and the Brain-Model Problem, *Doctoral dissertation*, Princeton University.

Minsky, M., and S. Papert. 1988. *Perceptrons, extended ed*. Cambridge, MA: The MIT Press.

Raiko, T., H. Valpola, and Y. LeCun. 2012. Deep Learning Made Easier by Linear Transformations in Perceptrons. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS) 2012*, vol. XX of JMLR:W and CP XXLa, Canary Islands: Palma.

Raina, R., A. Madhavan, and A.Y. Ng. 2009. Large-Scale Deep Unsupervised Learning using Graphics Processors. In *Proceedings of the 26th International Conference on Machine Learning*, Montreal, Canada.

Reetz, B. 1993. Greedy Solution to the Travelling Sales Person Problem. *ATD 2*.

Rosenblatt, F. 1962. *Principles of Neurodynamics*. N.Y.: Spartan.

Rumelhart D.E., J. L. Mcclelland (eds.). 1986. *Parallel Distributed Processing, vol. 1 Foundations, Explorations in the Microstructure of Cognition, vol. 2 Psychological and Biological Models*. Cambridge, MA, London, England: The MIT Press.

Werbos, P. 1974. Beyond Regression: New Tools for Prediction and Analysis in Behavioral Sciences, *PhD thesis*, Cambridge, MA: Harvard.

Chapter 2
Artificial Neural Networks

Abstract Artificial Adaptive Systems include Artificial Neural Networks (ANNs or simply neural networks as they are commonly known). The philosophy of neural networks is to extract from data the underlying model that relates this data as an input/output (domain/range) pair. This is quite different from the way most mathematical modeling processes operate. Most mathematical modeling processes normally impose on the given data a model from which the input to output relationship is obtained. For example, a linear model that is a "best fit" in some sense, that relates the input to the output is such a model. What is imposed on the data by artificial neural networks is an a priori architecture rather than an a priori model. From the architecture, a model is extracted. It is clear, from any process that seeks to relate input to output (domain to range), requires a representation of the relationships among data. The advantage of imposing an architecture rather than a data model, is that it allows for the model to adapt. Fundamentally, a neural network is represented by its architecture. Thus, we look at the architecture first followed by a brief introduction of the two types of approaches for implementing the architecture—supervised and unsupervised neural networks. Recall that Auto-CM, which we discuss in Chap. 3, is an unsupervised ANN while K-CM, discussed in Chap. 6, is a supervised version of Auto-CM. However, in this chapter, we show that, in fact, supervised and unsupervised neural networks can be viewed within one framework in the case of the linear perceptron. The chapter ends with a brief look at some theoretical considerations.

2.1 Introduction

We begin with the anatomy of neural networks, its architecture. ANNs are a family of methods inspired by the human brain's learning capability. ANNs are scientifically used in three different epistemological ways:

1. To understand, the function of the brain by computer simulations;
2. To reproduce in computer algorithms, the way the brain functions in its relationship with the environment, for example, in problem solving, driving a car, and so on (human brain emulation);

P. M. Buscema et al., *Artificial Adaptive Systems Using Auto Contractive Maps*, Studies in Systems, Decision and Control 131, https://doi.org/10.1007/978-3-319-75049-1_2

3. To understand the transition from individual to collective behavior (data analysis, data mining and the research on complex systems are part of this).

Currently ANNs comprise a range of very different models, but they all share the following characteristics.

- The fundamental elements of ANNs are the *nodes*, also known as processing elements and their *connections*.
- Each node in an ANN has its own *input* through which it receives communications from the other nodes or from the environment, and its own *output*, through which it communicates with other nodes or with the environment. In addition, it has an internal function, $f(\cdot)$ which transforms its global input into an output.
- Each connection may poses an internal relationship or “force” between pairs of nodes that excite or inhibit each other. Positive values indicate excitatory connections and negative ones indicate inhibitory connections.
- Connections between nodes may change over time. This dynamic (time dimension) triggers a *learning process* throughout the entire ANN. The way the law by which the connections change in time is called the “learning equation”.
  In order for the connections of the ANN to properly change, the environment must act on the ANN several times.
- When ANNs are used to process data, the data are ANN’s environment. Thus, in order to process data, the data is subjected to the ANN several times.
- The overall dynamic of an ANN depends exclusively on the local interaction of its nodes. The final state of the ANN “evolves” “spontaneously” from the interaction of all of its components (nodes).
- Communications between nodes in every ANN tend to occur in *parallel*. This parallelism may be *synchronous* or *asynchronous* and each ANN may emphasize this parallelism in a different way. In synchronous ANNs, all nodes simultaneously update their state variables whereas in the asynchronous regime a random node is chosen to update and the other nodes likewise are updated in a random fashion. However, an ANN always has some form of parallelism in the activity of its nodes. From a theoretical viewpoint this parallelism does not depend on the hardware on which the ANNs are implemented.

The architecture of every ANN is composed of the following five components:

1. Type and number of *nodes* and their corresponding properties;
2. Type and number of *connections* and their corresponding location;
3. Type and number of *layers*;
4. Type of *signal flow* strategy;
5. Type of *learning strategy*.

In short, ANNs have *nodes*, *connections*, *layers, signal flow,* and *learning strategy*. These five aspects are discussed next.

The Nodes

There are three types of ANN nodes, depending on the position they occupy within the ANN.