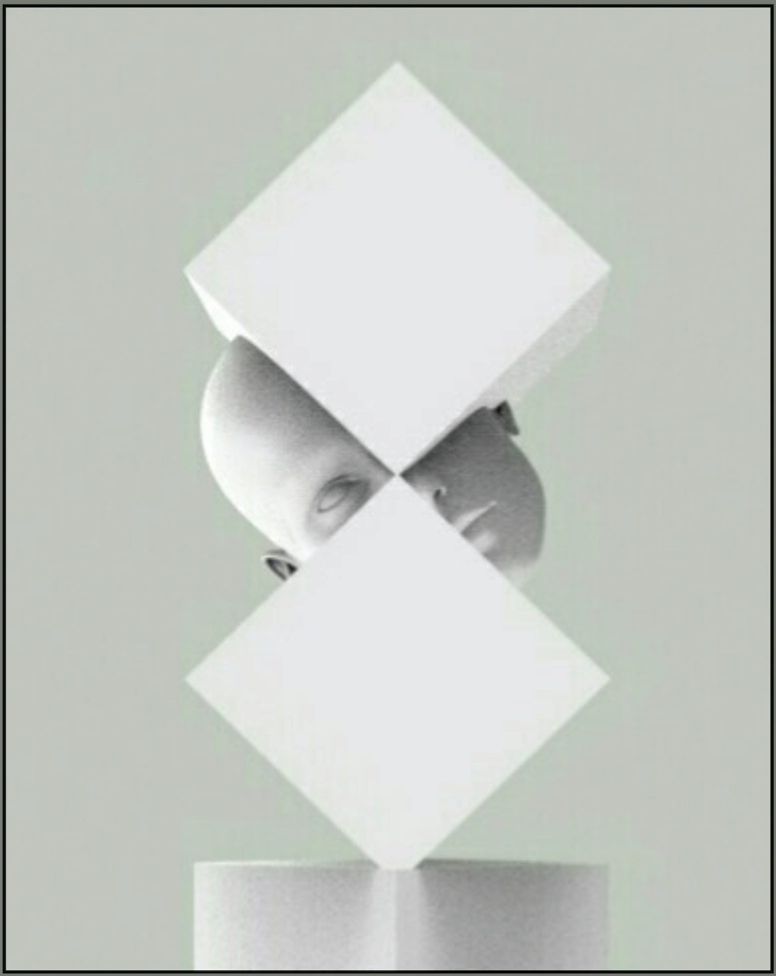
Deep Learning in Time Series Analysis

Arash Gharehbaghi



CRC Press
Taylor & Francis Group

A SCIENCE PUBLISHERS BOOK

Deep Learning in Time Series Analysis

Arash Gharehbaghi
Researcher, School of Information Technology
Halmstad University, Halmstad, Sweden

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A SCIENCE PUBLISHERS BOOK

Cover illustration courtesy of Reza Gharehbaghi

First edition published 2023
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

Library of Congress Cataloging-in-Publication Data (applied for)

ISBN: 978-0-367-32178-9 (hbk)
ISBN: 978-1-032-41886-5 (pbk)
ISBN: 978-0-429-32125-2 (ebk)

DOI: 10.1201/9780429321252

Typeset in Times New Roman
by Radiant Productions

To Shabnam, Anita and Parsa.

Foreword

I am delighted to introduce the first book on deep learning for time series analysis in which analysis of cyclic time series is profoundly addressed along with the theories. This idea was developed within a structure of a hybrid model where the experimental results showed its outperformance against the baselines of neural network-based methods. It was later improved by incorporating deep learning structures of a time growing neural network, the network which was previously introduced by us as a strong alternative to multilayer perceptron and time-delayed neural networks, into a multi-scale learning structure. The idea of cyclic learning is applicable to many natural learning where the phenomena exhibit cyclic behaviours. Physiological characteristics of the human body emanate cyclic activities in many cases such as cardiac and respiratory activities. The idea of cyclic learning has received interest from researchers from various domains of engineering and science.

Realistic validation of machine learning methods is a crucial task. A realistic validation method must provide sufficient outcomes to project capability of a machine learning method in terms of its risks in reproducibility of the results in conjunction with the improvement of the results when the machine learning method is being trained by a richer dataset. These validation capabilities are considered in the A-Test method. As a validation method, A-Test has received recognition from different engineering domains. These methods are likely to become strong machine learning methods, especially for applications with a small size of the learning data.

Preface

Learning has been regarded as an important element of development by most of the scientific pursuits including computer sciences in which deep machine learning has recently sounded as an emerging context. Application of deep machine learning methods has been well-received by the researchers and engineers since the last half decade when time series analysis was increasingly regarded as an important topic within different contexts such as biomedical engineering. Although development of strong tools for the implementation of deep learning methods created a breakthrough in computer science and engineering, a shift towards abstract understanding within this context is clearly seen, especially in the younger developers. This can put a negative impression on the general beliefs of deep learning which will be in turn considered as downside of this progress. Nowadays, various deep learning methods are enormously developed and published in the highly reputable references, however, a very low percentage of them entail sufficient quality to make a real impact on the underlying community. One reason can be the lack of sufficient insight into the theoretical foundation as well as into the implementation knowledge. This motivated the author to prepare a textbook on deep learning methods, sophisticated for time series analysis, to bring up fundamentals of the context along with the algorithms for the implementation.

Book Focus

This book focuses on the deep notions of the learning process in general, and deep learning in particular, with more orientation towards essentials of the traditional methods and the modern ones for time series analysis. Although image processing is known as an important topic of deep learning, the authors concluded to exclude this topic from the book and assign it to a separate publication as future work. The rationale behind this conclusion was mainly to avoid extra diversity and losing central attention. The book begins with a smooth transition from the fundamental definitions and speculations toward method formulation. Contents of the book were pedagogically organized in a way to foster and consolidate the essentials of time series analysis. This manner of representation is set to broaden the scope of the readers from the scientific to the engineering aspects. The book also considered to bring up a number of the practical examples of the deep learning methods for time series analysis, with the rise of biomedical engineering and medical informatics applications. Meanwhile, the book represents the deep learning methods mostly in a mathematical manner to help the researchers and the developers in mathematically formulating their own methods. It is evident that mathematical representation of a new method provides better readability compared to the descriptive representation. It is seen that new students show more tendency to learn concepts of deep learning using block diagrams and descriptive methods. Indulging in this learning manner can mislead them from the basic abilities in mathematical representation that can act as a degenerative factor for the learning of deep learning methods. Furthermore, a consistent graphical representation is not seen in many cases.

A number of the new ideas in artificial intelligence are also presented in this book. The A-Test validation method is introduced and compared to the other traditional ones. The readers can easily find out elaborations of this method in providing a more realistic validation as compared to the other two alternatives. In terms of the learning models, the idea of cyclic time series and cyclic learning are the other two new concepts addressed by this book and some of the learning methods such as time growing neural network is also introduced for learning cyclic time series.

Generative models for time series analysis were not addressed by this book. These models fit well into the prediction category which is considered as part of the future work.

Book Readership

This book, as a textbook, has been written in a fashion to establish fundamentals of time series analysis and deep learning methods for the readers. Problem formulation, as well as the methodological representations, have been rendered in a way to address the notional contents with a special focus on the scientific manners, so all the students, scientists, engineers and developers who are interested to learn deep learning methods for time series analysis or building up their own heuristic methods, can find this book interesting to read. The students in engineering, particularly those in artificial intelligence, are rather encouraged to read this book as a textbook.

Contributions

The book title as well as the arrangement of the chapters and the chapter titles were prepared by Arash Gharehbaghi and Ning Xiong. Contents of the first 10 chapters have been completely written by Arash Gharehbaghi. It incorporates all the writings, graphical representations, and tables. Chapter 11 was prepared by Ning Xiong and Johan Holmberg. Chapter 12 was prepared by Elaheh Partovi and Ning Xiong.

Arash Gharehbaghi

Contents

Contributors

Arash Gharehbaghi

School of Information Technology, Halmstad University, Halmstad, Sweden.

Ning Xiong

School of Innovation, Design and Technology, Mälardalen University, Västerås, Sweden.

Johan Holmberg

School of Innovation, Design and Technology, Mälardalen University, Västerås, Sweden.

Elaheh Partovi

Department of Electrical Engineering, Amir Kabir University (Tehran, Polytechnique), Tehran, Iran.

Part I

Fundamentals of Learning

1

Introduction to Learning

All the components of nature have been set in an everlasting pursue towards evolution through a growing network of mutual effects. Evolution is defined as an absolute point of dimensionless in time when the fully interconnected network of effects is moving in a completely predictable manner. The pursue towards evolution incrementally occurs in the form of an adaptation, not only for brainless materials, but also for any kind of intelligent creatures, such that adaptability to the underlying environment is defined as one of the indicative parameters of intelligence. Adaptation of a component can be expressed by the inherent temporal changes in the behaviour, and ultimately in the structure of the component to exhibit further similarities with the surrounding environment. This important definition will be seen in the section about the learning theory in Chapter 2.

At this point, one should consider time as an important element of adaptation. In fact, adaptation is a dynamic process, through which the adapting system adjusts, itself to the surrounding environment which is, in turn, another dynamic element. In a dynamic environment, an adaptation of a component is seen as a reaction of the component to the environmental variations towards the same objective, exhibiting further similarities with the surrounding environment. However, interpretation of this adaptation varies when it comes to computer science in which decision making is regarded as an important feature of intelligence. Although this point of view might change when it comes to the psychological perspectives, in a broader sense, decision making is a way of adaptation. These two scopes converge to a single point if we consider the process leading to decision making, which is based on learning similarity along with the differences over various groups of data, which is by itself linked to adaptation strongly. Putting such the absolute definitions (which may not be observed within the lifetime of components) into consideration, can be intuitive for real-world scenarios both for the problem formulation, and for the solution methodologies, as like as the development of the deductive sciences in which absolute definitions provide a theoretical foundation of the applicative contents. In analogy, a similar development has happened in mathematics, when the analytical mathematics provided fundamentals of the numerical mathematics to respond to the complicated real-world questions.

Numerical mathematics has been well-embraced since the development of digital computers [127]. It is obvious that many machine learning methods are based on the methods initiated by the scientists from the numerical mathematics domain, without which one could barely imagine such a rapid progress in artificial intelligence. It can be concluded that the natural development of deductive science is typically initiated by the analytic foundation of the theories, and extends towards application, unlike many branches of natural sciences which flow from observation to the theories. It is almost customary in chemistry that a phenomenon is observed first, and then the scientists bring up theories to explain the phenomenon. In other words, the incentive begins from an observation of the theory. In mathematics, the journey sometimes oppositely commenced from a theory to the application. This attitude has been tried to be followed in this book. The principals and concepts are described before methodologies to link the readers to the deep notion of the methods. The authors believe that it is essential for the readers to deeply understand the notions, logic, and reasoning, hidden behind each presenting method.

Nowadays, numerous and varied methods along with the pertinent open source codes for the implementation, have exploded within the community of artificial intelligence, thanks to the new advances in computer engineering [76]. Certainly, a broad range of options are being opened to researchers, which is naturally favourable, however, selecting the most appropriate one among the innumerable options, is not an easy task if one suffices to the practical aspects only. Furthermore, there are always hyper-parameters (sometimes called design parameters), associated with each method, and therefore providing an optimal solution for a research question is almost unpractical without a deep understanding of the theoretical foundation of the method. Many experienced scientists believe scientific studies were averagely deeper in analytic methods before the popularization of the open source implementation of the methods, when the scientists had to develop their codes! Such tough judgement is of course controversial and out of the belief of the authors, but worth considering to conclude; a deep understanding of the deep learning methods is essential prior to any kind of implementation.

Deep learning is indeed a sophistication of the above-mentioned adaptation. Learning theory would be described in more detail in the next chapter, nevertheless, it is worth addressing one important link between learning and adaptation in this instance. Learning is mostly concomitant with decision making, for better compliance with a minimal number of parameters. In contrary, adaptation is mostly a continuous process of system parameters to show better similarity to a specific reference. This point will be expanded in more detail in the section about learning theory, but it is important to note that both learning and adaptation processes depend on the environmental dynamics or the input data. To provide a consistent presumptions for the rest of the chapters, some of the definitions are described in the following sequels. These definitions will be presented descriptively, starting from the point of signal, data, and eventually time series, definition, and landing for analysis. Fundamentals of noise and the existing models for the analysis will be addressed as well. It is important to establish the definitions in a clear way at the introduction, as most of the theories and methodologies are built upon these bases. This chapter will terminate with a brief view of the book organisation.

1.1 Artificial Intelligence

Adaptation of a natural element is affected by a superposition of the surrounding elements through a network of interactions. Many scientists believe in a high level of intelligence holistically governing natural movements. However, a partial insight into the tiny particles shows that intelligence in an element begins when a level of decision making occurs in that element, is not generally true [4]. An approach towards simulating such an intelligence by using mathematical tools, is known as "artificial intelligence". The conventional approach for a deeper understanding of intelligence was firstly inspired by the human brain and concentrated on the learning process only [69]. The presented model became a frontier against statistical methods which were already a popular common, as a more powerful alternative for the learning, however, researchers found out that the two alternatives were intrinsically similar but algorithmically different in terms of the calculation [15]. As a consequence, the gap between the statistical way of learning and artificial intelligence-based fashion became narrower, such that the two alternatives became well merged [115]. In contrast with the traditional view of artificial intelligence which was predominantly about the learning process, modern perspective has broadened the scope of artificial intelligence to an expansive context including three main topics shown in Figure 1.1.

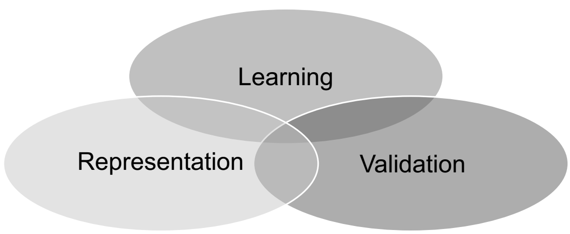


FIGURE 1.1: The three main topics of artificial intelligence.

In this perspective, learning is limited to only one topic of the larger context of artificial intelligence against the two other topics: representation and validation. If the learning is methodologically based on extracting information from the data through procedural machine-based routines, the learning is named "machine learning". Representation addresses methods for demonstration, dimension reduction or quantification of information. Validation is concerned with evaluating the performance of the learning process, which is in turn affected by the representation. Nevertheless, as we will see in the upcoming chapters, in the modern learning perspectives, validation sometimes influences the learning, or even the representation processes.

1.2 Data and Signal Definition

Data is defined as, AN ASSOCIATION TO AN EXISTENCE! This definition contains two keywords: association and existence. Association can be attribution, defined according to the findings, e.g., facial colour of a patient. It is often collected by the measurement, numerical or symbolical label, or obtained through mathematical or statistical mapping. Existence can be a phenomenon, an object, a living object, a signal, or any kind of image. For example, a person's weight is a measuring data, while the facial colour of a newborn baby is a symbolic data collected from the individuals. Weight and colour are the association and the individual from which the data is collected is the existence. In science and engineering, the signal is a sequential registration or/and representation of a phenomenon in time. The phenomenon is often represented in time-value axes. As an example, temporal variation of the electrical potential acquired from a muscular unit can be plotted and considered as a vital signal, called Electromyograph. Figure 1.2 depicts a typical electromyograph.

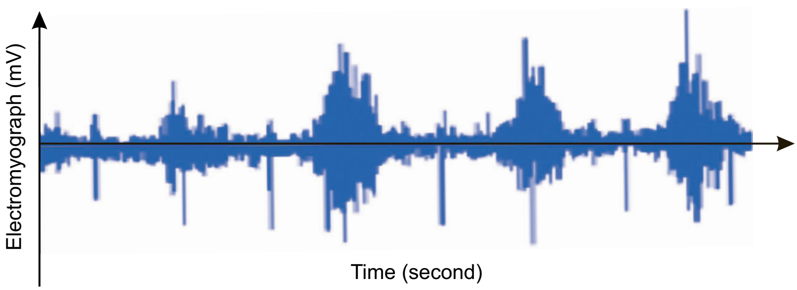


FIGURE 1.2: A rough representation of electromyograph.

Temporal profile of fluid pressure is also a signal, that is regarded as a vital signal if the fluid is blood. Regardless of the type of signal, there is always a variable of time associated with a variable of value representing the intensity of the phenomenon represented by the signal, or one can conclude that a signal is the variation profile of a phenomenon over time. Several related publications exist in the literature that categorise images either as signal or as data [105]. They introduce the processing commonly for images and signals, where an image is treated as a two-dimensional signal. In contradiction, this book considers image a separate class, which is not dealt with in terms of the methodologies. There are several motivations for this categorisation. Although some of the processing methods are essentially common for images and signals, a large number of the methods have been sophisticatedly introduced for signals only, in terms of both theoretical foundation and the implementation details. Furthermore, each pixel of an image may, or may not, have a physical dimension, depending on the image contents. Nevertheless, it is not decisive that a physical dimension is always associated with an image. On the other hand, an image is always captured or registered, like as a signal, and consequently cannot be classified as data. Deep learning methods for time series of the image is beyond the scope of this book and will not be addressed in the sequels of the book.

1.3 Data Versus Signal

The first difference that one can realise after the above descriptions, is an association of time with signal, in contrast with data where this association is not necessitated. The signal is often recorded in a previously known time interval, but data is not necessarily collected in a certain time sequence. For example, the facial colour of a baby might be collected a few hours after the birth, however, not all newborn babies must be attributed by their facial colour in some of the clinical routines. In contrary, signal is a registration in time with a certain order of time sequence. The data is, therefore, recorded or collected, whereas the signal is merely recorded. Another discriminating aspect of signal against data, is the dimension of signal. To clarify this point, a signal always corresponds to a phenomenon, represented by using a measuring technique, that provides a link between the phenomenon and the measuring technique with a known physical dimension. Data is not necessarily associated with a physical dimension, as not for the facial colour of a newborn baby. Moreover, the link to the physical phenomenon is not generally seen for data, but is so for the signal clearly. Sometimes data is a result of applying a processing algorithm on a signal and the resulting features extracted from the signal, constitute multi-dimensional data, from which there might be features that are not physically interpretative. Such data is often obtained by introducing a mathematical, or even a statistical mapping, applied to a set of certain types of signals. You will see the case studies in the following sections, in particular when the time series analysis is described.

1.4 Signal Models

Scientific studies toward understanding natural phenomena are mostly based on the pertinent models, capable of justifying different behaviors of the phenomena. A good model is the one that can explore different actions, interactions and also behaviors of the phenomenon. It is important to emphasize the fact that a model is indeed a way that we see a phenomenon, by assuming a set of the presumptions along with the range of the model applicability. The phenomenon by itself might show other sides in different conditions, which were not foreseen by the model. In general, there are three ways to model signals: deterministic, chaotic, and stochastic. Although similarities are seen in the processing methodologies, fundamental differences in the theories make such the classification necessary. In deterministic models, a signal is modelled by a closed mathematical formula, implying that behavior of the signal is well-recognized, and described by the formula. It is obvious that if a signal is completely modelled by a closed mathematical formula, then the value of the signal for the future as well as the past time, can be accurately predicted using the formula, which is clearly unpractical in real-world scenarios. In fact, deterministic models mostly correspond to the absolute definitions, and their applicability is mostly limited only to opening up a theoretical foundation. In deterministic methodologies, parameters like amplitude, phase, energy, and frequency are of importance for the processing. There is always a gap between deterministic models and applied solutions to respond practical questions, caused by variation of the signal behavior in time, called non-stationary behaviour of the signal, and also signal variation over subjects. This gap is mainly covered by further expansion of the deterministic models for specific practical questions, in a mathematical and sometimes statistical manner. Stochastic models come to practice when statistical methods are employed for modeling the signals. Stochastic models are based on the fact that amplitude of the signal is not strictly known in time, and is considered a random variable. Stochastic signals are attributed by the statistical parameters, like average and variance of the signals, in addition to those for the deterministic models. Many practical questions are better resolved thanks to the stochastic models. Chaotic models are basically deterministic with an initial value which is a random variable. Such models are sometimes called, disordered deterministic.

1.5 Noise and Interference

Noise is an unwanted, random-valued content, affecting a signal, with different unknown sources. So, there are always uncertainties associated with noise. Nevertheless, a kind of categorization is attributed to noise in sense of the possible source, to allow the engineers and researchers to justify behaviour of random processes appearing in the studies, and to select an appropriate statistical method to model the process. Thermal noise is the most common type of noise, seen in almost all electronic circuits. It results from any motion of the electrical charge carrier inside a semiconductor part of an electronic component. This is a random process the unavoidable occurs when the tem-