Water Science and Technology Library

Taesam Lee
Vijay P. Singh
Kyung Hwa Cho

Deep Learning for Hydrometeorology and Environmental Science

Springer

Water Science and Technology Library

Volume 99

The aim of the *Water Science and Technology Library* is to provide a forum for dissemination of the state-of-the-art of topics of current interest in the area of water science and technology. This is accomplished through publication of reference books and monographs, authored or edited. Occasionally also proceedings volumes are accepted for publication in the series. *Water Science and Technology Library* encompasses a wide range of topics dealing with science as well as socio-economic aspects of water, environment, and ecology. Both the water quantity and quality issues are relevant and are embraced by *Water Science and Technology Library*. The emphasis may be on either the scientific content, or techniques of solution, or both. There is increasing emphasis these days on processes and *Water Science and Technology Library* is committed to promoting this emphasis by publishing books emphasizing scientific discussions of physical, chemical, and/or biological aspects of water resources. Likewise, current or emerging solution techniques receive high priority. Interdisciplinary coverage is encouraged. Case studies contributing to our knowledge of water science and technology are also embraced by the series. Innovative ideas and novel techniques are of particular interest.

Comments or suggestions for future volumes are welcomed.

Vijay P. Singh, Department of Biological and Agricultural Engineering & Zachry Department of Civil and Environment Engineering, Texas A&M University, USA
Email: vsingh@tamu.edu

More information about this series at http://www.springer.com/series/6689

Taesam Lee · Vijay P. Singh ·
Kyung Hwa Cho

Deep Learning for Hydrometeorology and Environmental Science

Springer

Taesam Lee
Department of Civil Engineering
Gyeongsang National University
Jinju, Korea (Republic of)

Kyung Hwa Cho
School of Urban and Environmental Engineering
Ulsan National Institute of Science and Technology
Ulsan, Korea (Republic of)

Vijay P. Singh
Department of Biological and Agricultural Engineering, Zachry Department of Civil and Environmental Engineering
Texas A&M University
College Station, TX, USA

ISSN 0921-092X ISSN 1872-4663 (electronic)
Water Science and Technology Library
ISBN 978-3-030-64776-6 ISBN 978-3-030-64777-3 (eBook)
https://doi.org/10.1007/978-3-030-64777-3

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

Dedicated to

TL: the memory of my mom, Gumran Cho

VPS: my wife Anita, who is no more, son Vinay, daughter Arti, daughter-in-law Sonali, son-in-law Vamsi, and grandchildren Ronin, Kayden, and Davin

KC: Wife Yeonju and Daughter Yuna

Preface

Deep learning is known as part of a machine learning methodology based on an artificial neural network. Increasing data availability and computing power enhance applications of deep learning to hydrometeorological and environmental fields. However, books that specifically focus on the application to these fields are limited. Therefore, this book focuses on the explanation of deep learning techniques and their applications to hydrometeorological and environmental studies.

This book is divided into three parts. The first part is the introduction of the basic neural network, covering the basic concepts of artificial neural network in Chaps. 1–7. Chapter 1 introduces the concept of deep learning, followed by the mathematical background in Chap. 2. In Chap. 3, how to preprocess a dataset before applying a model is presented. Chapter 4 describes the terminology and structure of neural network models. The procedure of training a neural network is discussed in Chap. 5. The approaches to update the weights of a network model are presented in Chap. 6. The techniques to improve the model performance are given in Chap. 7.

The second part introduces advanced techniques in deep learning algorithms from Chaps. 8–10. The advanced neural network algorithms, as Extreme Learning Machine and Autoencoding, are presented in Chap. 8. The temporal deep learning techniques, as Long Short-term Memory (LSTM) and Gated Recurrent Unit (GRU), are discussed in Chap. 9. The spatial deep learning technique, as Convolution Neural Network (CNN), is introduced in Chap. 10.

The third part illustrates how to apply deep learning techniques to real case studies. In Chap. 11, Tensor flow and Keras programming is presented to illustrate how to simply implement deep learning to real datasets. Hydrometeorological and environmental applications of deep learning models are presented in Chaps. 12 and 13, respectively.

The book will be useful to graduate students, college faculty, and researchers in hydrology, meteorology, and environmental sciences. It may also be useful to policymakers in government at local, state, and national levels.

The first author acknowledges his former student, Mrs. Mahsa Moradi, for providing excellent hydrometeorological deep learning examples and the National Research Foundation of Korea for providing partial fund for the current work.

Jinju, South Korea — Taesam Lee
College Station, TX, USA — Vijay P. Singh
Ulsan, South Korea — Kyung Hwa Cho
June 2020

Contents

About the Authors

Prof. Taesam Lee, Ph.D. is a Full Professor in the Department of Civil Engineering at Gyeongsang National University in Jinju, South Korea. He got his Ph.D. degree from Colorado State University with a stochastic simulation of stream flow. He specializes in surface-water hydrology, meteorology, machine learning algorithms, and climatic changes in hydrological extremes publishing around 50 technical papers and a statistical downscaling book. He is a member of the American Society of Civil Engineers (ASCE) and the American Geophysical Union (AGU) and the associate editor of the *Journal of Hydrologic Engineering* in ASCE.

Prof. Vijay P. Singh is a University Distinguished Professor, a Regents Professor, and Caroline and William N. Lehrer Distinguished Chair in Water Engineering at Texas A&M University. He received his B.S., M.S., Ph.D. and D.Sc. degrees in engineering. He is a registered professional engineer, a registered professional hydrologist, and an Honorary Diplomate of ASCE-AAWRE. He has published more than 1270 journal articles, 30 textbooks, 70 edited reference books, 105 book chapters, and 315 conference papers in the area of hydrology and water resources. He has received more than 90 national and international awards, including three honorary doctorates. He is a member of 11 international science/engineering academies. He has served as President of the American Institute of Hydrology (AIH), Chair of Watershed Council of American Society of Civil Engineers, and is currently President-Elect of the American Academy of Water Resources Engineers. He has served/serves as editor-in-chief of three journals and two book series and serves on editorial boards of more than 25 journals and three book series.

Prof. Kyung Hwa Cho, Ph.D. is an Associate Professor in the School of Urban and Environmental Engineering at Ulsan National Institute of Science and Technology, South Korea. He obtained his B.S. in chemical engineering and M.S. and Ph.D. in Environmental Engineering. He has published more than 110 journal articles in water and environmental journals such as *Water Research*, *Remote Sensing of Environment*. His expertise lies in modeling water quality, deep learning application for water quality prediction, and using hyperspectral images for water quality monitoring.

Chapter 1
Introduction

Abstract Deep learning has been popularly employed for analysis and forecasting in various fields. In this chapter, a brief introduction of deep learning is presented, including the definition and pros and cons of deep learning, followed by the recent applications of deep learning models in hydrological and environmental fields. The structure of the remaining chapters for this book is also explained.

1.1 What is Deep Learning?

In recent years, deep learning techniques have been developed and employed in a number of fields, such as voice search, automatic text generation, health care, and image recognition. The skyrocketing development and applications of deep learning stem from its capability of image classification and object detection with much more accuracy than a human can do.

Old-fashioned machine learning algorithms using artificial neural networks that have been developed so far are required to have selected features to learn in advance. Automated feature learning is one of the major characteristics of deep learning. Therefore, deep learning can be defined as a subset of machine learning in artificial intelligence with artificial neural networks that are able to learn without supervision from data and is also known as deep neural learning and deep neural network.

More intuitively, it is comparable to shallow learning. Let's assume that one trains a dog or other animals. Shallow learning is just direct intuitive learning by leading its action intentionally, for example, sitting down by pointing a finger and calling. In contrast, deep learning is more like learning a trend or behavior and making a decision by itself, for example, learning numbers by watching them. In order to make deep learning accessible, one must possess two important characteristics as remembering and learning from watching. Therefore, these two deep learning characteristics are defined in this book as temporal deep learning as remembering and spatial deep learning as learning from watching.

One of the major temporal deep learning techniques is recurrent neural network (RNN) models such as Long Short-term Memory and Gated Recurrent Unit. Also,

T. Lee et al., *Deep Learning for Hydrometeorology and Environmental Science*,
Water Science and Technology Library 99,
https://doi.org/10.1007/978-3-030-64777-3_1

Convolutional Neural Network is one of the major spatial deep learning techniques. In this book, these two deep learning techniques are mainly introduced, followed by applications of these deep learning techniques to hydrometeorological and environmental studies. Since those deep learning techniques are based on neural network models, the description of neural network models is given in advance.

1.2 Pros and Cons of Deep Learning

Deep learning has been recognized as the technology that makes artificial intelligence eventually become smart. Deep learning allows prediction by reducing the effort to find feature variables that are mostly time-consuming. With enough amount of data, not much human intervention is needed by deep learning to outperform other models. In other words, it can learn by itself from mimicking a human brain, especially in many layers of neurons in the brain cortex with the given dataset.

Deep learning is still pricy and resource-consuming and also requires a large amount of data. Since it is a branch of neural networks, it still lacks a strong theoretical foundation and the output result cannot generally provide theoretical reasoning and explanation. This unexplainable theoretical foundation and the requirement of a large dataset limits the application of deep learning models to hydrometeorology and environmental sciences.

1.3 Recent Applications of Deep Learning in Hydrometeorological and Environmental Studies

There are a number of recent developments and applications of deep learning models, especially to hydrometeorological and environmental studies. Some of the selected studies are discussed to present how deep learning models have been applied in the fields of hydrometeorology and environmental science.

In the hydrological field, temporal deep learning algorithms have also been applied in recent years. For example, Kratzert et al. (2018) applied a recent Recurrent Neural Network (RNN) model, named Long Short-Term Memory (LSTM), for rainfall-runoff modeling and compared it with Sacramento Soil Moisture Accounting Model (SAC-SMA) coupled with Snow-17 snow routine. Their results showed that LSTM had a competitive performance in comparison to the physical model of SAC-SMA, especially with regional scaling. Hu et al. (2018) compared artificial neural network (ANN) and LSTM for forecasting floods. Their results indicated that LSTM outperformed the conventional ANN and also showed that the model was more stable. Lee et al. (2020) proposed a stochastic simulation model with the LSTM model and their results indicated that the LSTM stochastic simulation model reproduced long-term variability as well as short-term memory.

In the meteorological field, Pan et al. (2019) employed the spatial deep learning model of Convolution Neural Network (CNN) to improve precipitation estimation in statistical downscaling. They trained the model to learn precipitation-related dynamical features from the surrounding dynamical fields and their results showed that the proposed model improved the precipitation-related parameterization scheme with CNN. Miao et al. (2019) applied the combined model of CNN and LSTM to forecast monsoon precipitation and compared it with ECMWF-Interim reanalysis precipitation. Their results showed that the combined deep learning model was superior to the physical model in forecasting precipitation more accurately from 1 day to 2 weeks in advance.

In environmental applications, Park et al. (2019) applied a deep neural network to model membrane fouling mechanisms and compared them with mathematical models. Their results indicated that the deep neural network showed better predictive performance in the fouling growth simulation and the flux decline simulation. Oga et al. (2019) applied CNN to estimate the water quality of a river. Using monitoring images, water quality can be estimated by training CNN, and it was observed that the proposed deep learning model outperformed the existing method in terms of accuracy.

1.4 Organization of Chapters

The chapters of this book are divided into three parts. The first part covers the introduction of the basic concepts of neural network models from Chaps. 1–7. Mathematical background and data preprocessing are given in Chaps. 2 and 3, respectively. Chapter 4 discusses the basic concept of a neural network and its training procedure is explained in Chap. 5. Chapter 6 explains how to update the weights of the neural network, followed by techniques to improve the model performance in Chap. 7.

The second part describes the advanced techniques in deep learning algorithms from Chaps. 8–10. In Chap. 8, advanced neural network algorithms, such as Extreme Learning Machine and Autoencoding, are discussed in Chap. 8. Chapter 9 deals with deep learning techniques, such as a recurrent neural network (RNN), long short-term memory (LSTM), and Gated Recurrent Unit (GRU), for time series data. Deep learning techniques for spatial datasets, as convolution neural network, are presented in Chap. 10.

The third part demonstrates the application procedure for deep learning techniques. Tensorflow and Keras programming are discussed in Chap. 11. Hydrometeorological and environmental applications to deep learning algorithms are presented in Chaps. 12 and 13, respectively.

1.5 Summary and Conclusion

Nowadays, there is a great deal of interest in the development and application of deep learning techniques in a number of fields. It is hoped that this book helps appreciate deep learning techniques, improve their understanding, and advance their application to hydrometeorological and environmental fields. Undergraduate and graduate students who are interested in deep learning algorithms in hydrometeorological and environmental fields may find the book to be useful. For beginners in governmental and educational sectors who need to apply deep learning techniques, this book might serve as a guide to become familiar with computational procedures for deep learning.

References

Hu C et al (2018) Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water (Switzerland), 10(11). https://doi.org/10.3390/w10111543

Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M (2018) Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrol Earth Syst Sci 22(11): https://doi.org/10.5194/hess-22-6005-2018

Lee T, Shin JY, Kim JS, Singh VP (2020) Stochastic simulation on reproducing long-term memory of hydroclimatological variables using deep learning model. J Hydrol 582. https://doi.org/10.1016/j.jhydrol.2019.124540

Miao Q, Pan B, Wang H, Hsu K, Sorooshian S (2019) Improving monsoon precipitation prediction using combined convolutional and long short term memory neural network. Water (Switzerland) 11(5). DOI:https://doi.org/10.3390/w11050977

Oga T, Umeki Y, Iwahashi M, Matsuda Y (2019) River water quality estimation based on convolutional neural network, 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2018—Proceedings, pp 1305–1308. https://doi.org/10.23919/APSIPA.2018.8659554

Pan B, Hsu K, AghaKouchak A, Sorooshian S (2019) Improving precipitation estimation using convolutional neural network. Water Resour Res 55(3): https://doi.org/10.1029/2018WR024090

Park S et al (2019) Deep neural networks for modeling fouling growth and flux decline during NF/RO membrane filtration. J Membrane Sci 587. https://doi.org/10.1016/j.memsci.2019.06.004

Chapter 2
Mathematical Background

Abstract In this current chapter, the fundamental mathematical background is presented for a deep learning model. Linear simple and multiple regression models are explained, including the definition of error terms and parameter estimation procedure, since they are similarly used in deep learning models. Also, the basic concept of the time series model is also explained and this part is mainly referred to in the LSTM model chapter.

In the current chapter, the fundamental mathematical background, including linear regression and time series model, is presented.

2.1 Linear Regression Model

2.1.1 Simple Linear Regression

A simple linear model can be described as $y = \beta_0 + \beta_1 x$ with the actual observed value of y as a linear function of x with parameters β_0 and β_1. Here, x is called as a predictor, explanatory variable, independent variable or input variable, while y is as predictand, response variable, dependent variable or output variable. This linear model can be generalized to a probabilistic model for the random variable Y as

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{2.1}$$

Note that Y is capitalized because the output of the model includes a random noise (ε) assumed to be normally distributed with mean $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma_\varepsilon^2$ and the output is now a random variable. Also, X is called a predictor (or an explanatory variable), while Y is a predictand (or a response variable).

T. Lee et al., *Deep Learning for Hydrometeorology and Environmental Science*,
Water Science and Technology Library 99,
https://doi.org/10.1007/978-3-030-64777-3_2

For a sample of observed data pairs of size n, that is, (x_i, y_i) for $i = 1, \ldots, N$, the sum of squares of errors (SSE) is defined as

$$SSE = \sum_{i=1}^{N} [y_i - \hat{y}_i]^2 = \sum_{i=1}^{N} [y_i - (\beta_0 + \beta_1 x_i)]^2 = \sum_{i=1}^{N} \varepsilon_i^2 \tag{2.2}$$

The parameters can be estimated by finding the parameter set with the criterion minimizing the sum of errors (i.e., SSE), called least-square estimate, which is analogous to taking the derivatives of SSE with respect to the parameters separately and equating the derivatives to zero as

$$\frac{\partial SSE}{\partial \beta_0} = -2\sum_{i=1}^{N} [y_i - (\beta_0 + \beta_1 x_i)] = 0 \tag{2.3}$$

$$\frac{\partial SSE}{\partial \beta_1} = -2\sum_{i=1}^{N} [y_i - (\beta_0 + \beta_1 x_i)](x_i) = 0 \tag{2.4}$$

From Eq. (2.3),

$$\sum_{i=1}^{N} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i = 0 \tag{2.5}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \tag{2.6}$$

From Eq. (2.4),

$$\sum_{i=1}^{N} [y_i - (\overline{y} - \beta_1 \overline{x} + \beta_1 x_i)]x_i = 0 \tag{2.7}$$

$$\sum_{i=1}^{N} [(y_i - \overline{y}) - \beta_1 (x_i - \overline{x})]x_i = 0 \tag{2.8}$$

Since $\sum_{i=1}^{N} [(y_i - \overline{y}) - \beta_1 (x_i - \overline{x})]\overline{x} = 0$,

$$\sum_{i=1}^{N} [(y_i - \overline{y}) - \beta_1 (x_i - \overline{x})]x_i - \sum_{i=1}^{N} [(y_i - \overline{y}) - \beta_1 (x_i - \overline{x})]\overline{x} = 0 \tag{2.9}$$

$$\sum_{i=1}^{N} [(y_i - \overline{y}) - \beta_1 (x_i - \overline{x})](x_i - \overline{x}) = 0 \tag{2.10}$$

Then,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N} (x_i - \overline{x})^2} \tag{2.11}$$

Note that the least-square estimate of β_0 and β_1 is now denoted as $\hat{\beta}_0$ and $\hat{\beta}_1$.

In addition, the coefficient of determination, the proportion of the variance in the predictand variable that is predictable from a predictor, is denoted as R^2 and can be estimated as follows:

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{N}\left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{2.12}$$

where the sum of squares of the total (SST) is $\sum_{i=1}^{n}(y_i - \overline{y})^2$ and the regression sum of squares (SSR) is $\sum_{i=1}^{N}\left(\hat{y}_i - \overline{y}\right)^2$. This statistic is a measure of how well the predictor observations can be replaced by the model predictions according to the portion of the variation, as in Eq. (2.12). A higher value of R2 indicates better performance of the linear regression model.

Example 2.1 Estimate the parameters of β_0 and β_1 with the least-square estimate as shown in Eqs. (2.6) and (2.11), respectively, for the dataset in the second column (i.e., x_1) for a predictor (x) and the fourth column for a predictand (y) in Table 2.1.

Solution:

As shown in Table 2.2,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2} = \frac{1.88}{3.65} = 0.51$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = -0.62 - 0.51 \times 0.51 = -0.8$$

Its scatterplot is shown in Fig. 2.1 with its estimated equation. Note that the line indicates the estimated value from $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -0.8 + 0.51x$.

Also, the coefficient of determination (R^2) can be estimated from Table 2.2 as

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = \frac{0.97}{2.77} = 0.35$$

2.1.2 Multiple Linear Regression

The multiple regression for multiple variables of ${x_i}^1, {x_i}^2, \ldots., {x_i}^S$ can be described as

Table 2.1 Example dataset for simple and multiple linear regression

Index	$\mathbf{x_1}$	$\mathbf{x_2}$	**Y**
1	−0.33	−0.49	−1.83
2	0.69	0.25	−0.23
3	0.26	−0.69	−1.25
4	−0.22	−0.54	−0.70
5	0.50	0.16	0.03
6	0.28	−0.16	−0.31
7	1.76	−0.18	−0.23
8	0.35	0.45	−0.79
9	0.47	0.49	−0.49
10	1.33	0.42	−0.44
Average	0.51	−0.03	−0.62

Table 2.2 Simple linear regression example to estimate the parameters and determination of coefficient (R^2)

Index	$x_i - \overline{x}$	$y_i - \overline{y}$	$(x_i - \overline{x}) \times (y_i - \overline{y})$	$(x_i - \overline{x})^2$	$\hat{y}_i$	$\hat{y}_i - \overline{y}$	$\left(\hat{y}_i - \overline{y}\right)^2$	$(y_i - \overline{y})^2$
1	−0.84	−1.21	1.01	0.70	−1.06	−0.43	0.19	1.45
2	0.18	0.39	0.07	0.03	−0.53	0.09	0.01	0.16
3	−0.25	−0.63	0.16	0.06	−0.75	−0.13	0.02	0.39
4	−0.73	−0.08	0.06	0.53	−1.00	−0.38	0.14	0.01
5	−0.01	0.65	−0.01	0.00	−0.63	0.00	0.00	0.43
6	−0.23	0.31	−0.07	0.05	−0.74	−0.12	0.01	0.10
7	1.25	0.39	0.49	1.57	0.02	0.65	0.42	0.16
8	−0.16	−0.17	0.03	0.03	−0.71	−0.08	0.01	0.03
9	−0.04	0.13	−0.01	0.00	−0.64	−0.02	0.00	0.02
10	0.82	0.18	0.15	0.67	−0.20	0.42	0.18	0.03
		Sum	**1.88**	**3.65**		Sum	**0.97**	**2.77**

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \cdots + \beta_s x_i^s + \varepsilon_i \tag{2.13}$$

where S is the number of predictors of interest. In matrix form, the linear regression model can be expressed by

$$Y_i = \vec{\mathbf{x}}_i^T \boldsymbol{\beta} + \varepsilon_i \tag{2.14}$$

where $\overrightarrow{\mathbf{x}}_i = [1, x_i^1, x_i^2, \ldots, x_i^S]^T$ and $\boldsymbol{\beta} = [\beta_0, \beta_1, .., \beta_s]^T$. Note that 1 is added in $\overrightarrow{\mathbf{x}}_i$ to include the intercept term of β_0 in this matrix form.