Cognitive Systems Monographs

Volume 35

About this Series

The Cognitive Systems Monographs (COSMOS) publish new developments and advances in the fields of cognitive systems research, rapidly and informally but with a high quality. The intent is to bridge cognitive brain science and biology with engineering disciplines. It covers all the technical contents, applications, and multidisciplinary aspects of cognitive systems, such as Bionics, System Analysis, System Modelling, System Design, Human Motion, Understanding, Human Activity Understanding, Man-Machine Interaction, Smart and Cognitive Environments, Human and Computer Vision, Neuroinformatics, Humanoids, Biologically motivated systems and artefacts Autonomous Systems, Linguistics, Sports Engineering, Computational Intelligence, Biosignal Processing, or Cognitive Materials as well as the methodologies behind them. Within the scope of the series are monographs, lecture notes, selected contributions from specialized conferences and workshops.

More information about this series at http://www.springer.com/series/8354

Mark Hoogendoorn · Burkhardt Funk

Machine Learning for the Quantified Self

On the Art of Learning from Sensory Data

Springer

Mark Hoogendoorn
Department of Computer Science
Vrije Universiteit Amsterdam
Amsterdam
The Netherlands

Burkhardt Funk
Institut für Wirtschaftsinformatik
Leuphana Universität Lüneburg
Lüneburg, Niedersachsen
Germany

ISSN 1867-4925 ISSN 1867-4933 (electronic)
Cognitive Systems Monographs
ISBN 978-3-319-66307-4 ISBN 978-3-319-66308-1 (eBook)
https://doi.org/10.1007/978-3-319-66308-1

Library of Congress Control Number: 2017949497

Printed on acid-free paper

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

Live as if you were to die tomorrow.
Learn as if you were to live forever.

Mahatma Gandhi

Foreword

Sensors are all around us, and increasingly on us. We carry smartphones and watches, which have the potential to gather enormous quantities of data. These data are often noisy, interrupted, and increasingly high dimensional. A challenge in data science is how to put this veritable fire hose of noisy data to use and extract useful summaries and predictions.

In this timely monograph, Mark Hoogendoorn and Burkhardt Funk face up to the challenge. Their choice of material shows good mastery of the various subfields of machine learning, which they bring to bear on these data. They cover a wide array of techniques for supervised and unsupervised learning, both for cross-sectional and time series data. Ending each chapter with a useful set of thinking and computing problems adds a helpful touch. I am sure this book will be welcomed by a broad audience, and I hope it is a big success.

June 2017

Trevor Hastie
Stanford University, Stanford, CA, USA

Preface

Self-tracking has become part of a modern lifestyle; wearables and smartphones support self-tracking in an easy fashion and change our behavior such as in the health sphere. The amount of data generated by these devices is so overwhelming that it is difficult to get useful insight from it. Luckily, in the domain of artificial intelligence, techniques exist that can help out here: machine learning approaches are well suited to assist and enable one to analyze this type of data. While there are ample books that explain machine learning techniques, self-tracking data comes with its own difficulties that require dedicated techniques such as learning over time and across users. In this book, we will explain the complete loop to effectively use self-tracking data for machine learning; from cleaning the data, the identification of features, finding clusters in the data, algorithms to create predictions of values for the present and future, to learning how to provide feedback to users based on their tracking data. All concepts we explain are drawn from state-of-the-art scientific literature. To illustrate all approaches, we use a case study of a rich self-tracking dataset obtained from the crowdsignals platform. While the book is focused on the self-tracking data, the techniques explained are more widely applicable to sensory data in general, making it useful for a wider audience.

Who should read this book? The book is intended for students, scholars, and practitioners with an interest in analyzing sensory data and user-generated content to build their own algorithms and applications. We will explain the basics of the suitable algorithms, and the underlying mathematics will be explained as far as it is beneficial for the application of the methods. The focus of the book is on the application side. We provide implementation in both Python and R of nearly all algorithms we explain throughout the book and make the code available for all the case studies we present in the book as well.

Additional material is available on the website of the book (ml4qs.org):

- Code examples are available in Python and R
- Datasets used in the book and additional sources to be explored by readers
- Up-to-date list of scientific papers and text books related to the book's theme

We have been researchers in this field for over ten years and would like to thank everybody who formed the body of knowledge that has become the basis for this book. First of all, we would like to thank the people at crowdsignals.io for providing us with the dataset that is used throughout the book, Evan Welbourne in particular. Furthermore, we want to thank the colleagues who contributed to the book: Dennis Becker, Ward van Breda, Vincent Bremer, Gusz Eiben, Eoin Grau, Evert Haasdijk, Ali el Hassouni, Floris den Hengst, and Bart Kamphorst. We also want to thank all the graduate students that participated in the Machine Learning for the Quantified Self course at the Vrije Universiteit Amsterdam in June 2017 and provided feedback on a preliminary version of the book that was used as reader during the course. Mark would like to thank (in the order of appearance in his academic career) Maria Gini, Catholijn Jonker, Jan Treur, Gusz Eiben, and Peter Szolovits for being such great sources of inspiration.

And of course, the writing of this book would not have been possible without our loving family and friends. Mark would specifically like to thank his parents for their continuous support and his friends for helping him in getting the proper relaxation in the busy book-writing period. Burkhardt is very grateful to his family, especially his wife Karen Funk and his two daughters, for allowing him to often work late and to spend almost half a year at the University of Virginia and Stanford University during his sabbatical.

Amsterdam, The Netherlands — Mark Hoogendoorn
Lüneburg, Germany — Burkhardt Funk
August 2017

Contents