

Advances in Computer Vision and Pattern Recognition

For further volumes:
www.springer.com/series/4205

Marcello Pelillo

Editor

Similarity-Based Pattern Analysis and Recognition

 Springer

Editor

Marcello Pelillo
DAIS
Ca' Foscari University
Venice, Italy

Series Editors

Sameer Singh
Rail Vision Europe Ltd.
Castle Donington
Leicestershire, UK

Sing Bing Kang
Interactive Visual Media Group
Microsoft Research
Redmond, WA, USA

ISSN 2191-6586

Advances in Computer Vision and Pattern Recognition

ISBN 978-1-4471-5627-7

DOI 10.1007/978-1-4471-5628-4

Springer London Heidelberg New York Dordrecht

ISSN 2191-6594 (electronic)

ISBN 978-1-4471-5628-4 (eBook)

Library of Congress Control Number: 2013955585

© Springer-Verlag London 2013

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

For my parents, who made it possible

“Surely there is nothing more basic to thought and language than our sense of similarity. [...] And every reasonable expectation depends on resemblance of circumstances, together with our tendency to expect similar causes to have similar effects.”

Willard V.O. Quine

Foreword

The SIMBAD project was a Future and Emerging Technologies (FET) project funded by the European Commission between 2008 and 2011. It brought together an extraordinary group of talented researchers with a broad spectrum of different perspectives on the central theme of using non-Euclidean similarity functions as the basis for learning. This approach was in contrast with the use of kernel functions that had become the de facto standard at the time of the project's launch in 2008.

The SIMBAD project took a broad view of the problem of so-called non-Euclidean learning: analysing the extent to which this was essential in a particular problem, developing alternative learning strategies that could successfully learn from non-Euclidean similarity functions, developing methods of learning Euclidean representations from probabilistic models and similarity data, and so on. These approaches were not studied just in the abstract but rather were grounded in a series of concrete problems from application domains where it was known or suspected that the Euclidean assumption was unrealistic.

The number and depth of the papers that arose from this research agenda was very impressive, with significant innovations made on all of the fronts listed above. However, the research was not merely a shotgun attack on several divergent fronts, but rather represented the coherent development of the leitmotiv of the project: the use of similarity functions in learning.

Given the breadth of the reach and impact of the research, the project reviewers were fearful that this coherence might be lost in the variety of journals, conferences, and particular problems considered, hence risking that the main message become lost in the plethora of individual results.

It was therefore proposed that a book bringing together the themes of the project and its main results could help champion and communicate the SIMBAD message in one coherent volume. This carefully constructed book is the result of that proposal. It is a distillation of the main themes and results of the project into an accessible and cross-referenced volume. For those interested in learning about the potential and importance of learning from similarity functions, this work is undoubtedly the

key reference from which to begin their study and it is likely to remain so for many years to come.

Virginia Water
June 2013

John Shawe-Taylor

Preface

This book provides a thorough description of a selection of results achieved within SIMBAD, an EU FP7 project which represents the first systematic attempt at bringing to full maturation a paradigm shift that is just emerging within the pattern recognition and machine learning domains, where researchers are becoming increasingly aware of the importance of similarity information *per se*, as opposed to the classical (feature-based) approach.

SIMBAD started in April 2008 and ended in September 2011, and involved the following six partners:

- University of Venice, Italy (*scientific coordinator*)
- University of York, UK
- Delft University of Technology, The Netherlands
- Instituto Superior Tecnico, Lisbon, Portugal
- ETH Zurich, Switzerland
- University of Verona, Italy.

The very end of the project marked also the launch of the SIMBAD workshop series <http://www.dsi.unive.it/~simbad>

whose first edition was held in Venice, in September 2011, in conjunction with the project's final review meeting. These biennial workshops aim to consolidate and promote research efforts in this area and to provide an informal discussion forum for researchers and practitioners.

Within the SIMBAD project we undertook a thorough study of several aspects of purely similarity-based pattern analysis and recognition methods, from the theoretical, computational, and applicative perspective. We covered a wide range of problems and perspectives. We considered both supervised and unsupervised learning paradigms, generative and discriminative models, and our interest ranged from purely theoretical problems to real-world practical applications. The chapters collected in this book aim to provide a coherent overview of our main achievements and to serve as a starting point for graduate students and researchers interested in

this important, yet diverse subject. More details on the project's activities can be found on our website

<http://simbad-fp7.eu>

and in the published papers referenced in this book.

A project like SIMBAD could not have been done without the help and support of many people and institutions, and it is a pleasure to take this opportunity to express my gratitude to them. In the first place, I'd like to acknowledge the Future and Emerging Technology (FET) Programme of the 7th Framework Programme for Research of the European Commission which funded the SIMBAD project, and I am very grateful to our project officer, Teresa De Martino, and to the reviewers, Georgios Sakas, Christoph Schnörr and John Shawe-Taylor, whose insightful suggestions and constant encouragement have been instrumental to make SIMBAD a better project.

It has been my good fortune to collaborate for almost four years with a fantastic group of people, whose genuine enthusiasm and exceptional professional competence made SIMBAD a unique, intellectually stimulating experience. In particular, I'm grateful to my fellow principal investigators who coordinated the activities of the various research units: Joachim Buhmann, Bob Duin, Mario Figueiredo, Edwin Hancock, and Vittorio Murino; to their deputies: Manuele Bicego, Umberto Castellani, Ana Fred, Marco Loog, Volker Roth, and Richard Wilson; and to all PhD students and postdocs who have worked within the project.

In Venice, I've been helped by many people in my group, and I'd like to thank them all for their support. In particular, I wish to thank Andrea Torsello for the assistance he gave me at various stages of the project, and Veronica Giove for her valuable work concerning all administrative aspects. Special thanks are due to Samuel Rota Bulò for his constant support throughout the project and for helping me assemble this book.

I'd like to thank the editorial staff at Springer, in particular Wayne Wheeler for supporting the idea of publishing this book, and Simon Rees for his advice throughout the production of the volume and for gently tolerating my procrastinations.

My deepest gratitude, however, goes to my wife, Rosanna, and my children, Claudia and Valerio, without whose endless patience and understanding the SIMBAD project, and hence this book, would have not seen the light.

Venice
July 2013

Marcello Pelillo

Contents

| | | |
|---|--|------------|
| 1 | Introduction: The SIMBAD Project | 1 |
| | Marcello Pelillo | |
| Part I Foundational Issues | | |
| 2 | Non-Euclidean Dissimilarities: Causes, Embedding and Informativeness | 13 |
| | Robert P.W. Duin, Elżbieta Pełkalska, and Marco Loog | |
| 3 | SIMBAD: Emergence of Pattern Similarity | 45 |
| | Joachim M. Buhmann | |
| Part II Deriving Similarities for Non-vectorial Data | | |
| 4 | On the Combination of Information-Theoretic Kernels with Generative Embeddings | 67 |
| | Pedro M.Q. Aguiar, Manuele Bicego, Umberto Castellani, Mário A.T. Figueiredo, André T. Martins, Vittorio Murino, Alessandro Perina, and Aydın Ulaş | |
| 5 | Learning Similarities from Examples Under the Evidence Accumulation Clustering Paradigm | 85 |
| | Ana L.N. Fred, André Lourenço, Helena Aidos, Samuel Rota Bulò, Nicola Rebagliati, Mário A.T. Figueiredo, and Marcello Pelillo | |
| Part III Embedding and Beyond | | |
| 6 | Geometricity and Embedding | 121 |
| | Peng Ren, Furqan Aziz, Lin Han, Eliza Xu, Richard C. Wilson, and Edwin R. Hancock | |
| 7 | Structure Preserving Embedding of Dissimilarity Data | 157 |
| | Volker Roth, Thomas J. Fuchs, Julia E. Vogt, Sandhya Prabhakaran, and Joachim M. Buhmann | |

| | | |
|-----------------------------|---|-----|
| 8 | A Game-Theoretic Approach to Pairwise Clustering and Matching | 179 |
| | Marcello Pelillo, Samuel Rota Bulò, Andrea Torsello, Andrea Albarelli, and Emanuele Rodolà | |
| Part IV Applications | | |
| 9 | Automated Analysis of Tissue Micro-Array Images on the Example of Renal Cell Carcinoma | 219 |
| | Peter J. Schüffler, Thomas J. Fuchs, Cheng Soon Ong, Volker Roth, and Joachim M. Buhmann | |
| 10 | Analysis of Brain Magnetic Resonance (MR) Scans for the Diagnosis of Mental Illness | 247 |
| | Aydın Ulaş, Umberto Castellani, Manuele Bicego, Vittorio Murino, Marcella Bellani, Michele Tansella, and Paolo Brambilla | |
| Index | | 289 |