Intelligent Systems Reference Library

Volume 118

About this Series

The aim of this series is to publish a Reference Library, including novel advances and developments in all aspects of Intelligent Systems in an easily accessible and well structured form. The series includes reference works, handbooks, compendia, textbooks, well-structured monographs, dictionaries, and encyclopedias. It contains well integrated knowledge and current information in the field of Intelligent Systems. The series covers the theory, applications, and design methods of Intelligent Systems. Virtually all disciplines such as engineering, computer science, avionics, business, e-commerce, environment, healthcare, physics and life science are included.

More information about this series at http://www.springer.com/series/8578

Dionisios N. Sotiropoulos · George A. Tsihrintzis

Machine Learning Paradigms

Artificial Immune Systems and their Applications in Software Personalization

Springer

Dionisios N. Sotiropoulos
University of Piraeus
Piraeus
Greece

George A. Tsihrintzis
University of Piraeus
Piraeus
Greece

ISSN 1868-4394 ISSN 1868-4408 (electronic)
Intelligent Systems Reference Library
ISBN 978-3-319-47192-1 ISBN 978-3-319-47194-5 (eBook)
DOI 10.1007/978-3-319-47194-5

Library of Congress Control Number: 2016953915

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

To my beloved family and friends

Dionisios N. Sotiropoulos

To my wife and colleague, Prof.-Dr. Maria Virvou, and our daughters, Evina, Konstantina and Andreani

George A. Tsihrintzis

Foreword

There are many real-world problems of such high complexity that traditional scientific approaches, based on physical and statistical modeling of the data generation mechanism, do not succeed in addressing them efficiently. The cause of inefficiencies often lies in multi-dimensionality, nonlinearities, chaotic phenomena and the presence of a plethora of degrees of freedom and unknown parameters in the mechanism that generates data. As a result, loss of information that is crucial to solve a problem is inherent in the data generation process itself, making a traditional mathematical solution intractable.

At the same time, however, biological systems have evolved to address similar problems in efficient ways. In nature, we observe abundant examples of high level intelligence, such as

- Biological neural networks, i.e., networks of interconnected biological neurons in the nervous system of most multi-cellular animals, are capable of learning, memorizing and recognizing patterns in signals such as images, sounds or odors.
- Ants exhibit a collective, decentralized and self-organized intelligence that allows them to discover the shortest route to food in very efficient ways.

A third example of a biological system that exhibits high level intelligence is the vertebrate immune system. The immune system in vertebrates is a decentralized system of biological structures and processes within an organism that protects against pathogens that threaten the organism and may cause disease. Even though it is virtually impossible for a system to learn and memorize all possible forms of pathogens that may potentially threaten an organism, the immune system is capable of detecting a wide variety of agents, ranging from viruses to parasitic worms, and distinguishing them from the organism's own healthy tissue.

Dionisios N. Sotiropoulos and George A. Tsihrintzis have authored the book at hand on computational aspects of the vertebrate immune system with the intent to promote the use of artificial immune systems in addressing machine learning problems. Artificial immune systems are a class of intelligent systems, inspired by the vertebrate immune system and capable of learning and memorizing. In the

recent years, the discipline of computer science has shown intense research interest into applying artificial immune system techniques in various pattern recognition, clustering and classification problems.

The book at hand is a significant addition to this field. The authors present artificial immune systems from the practical signal processing point of view, emphasizing relevant algorithms for clustering and classification. Additionally, the authors illustrate the use of the proposed algorithms through a number of case studies and application on a variety of real data. Particularly interesting is the authors' proposal to use artificial immune system approaches to tackle classification problems which exhibit high, even extreme, class imbalance (so-called *one-class classification problems*).

A practical application of their approach may be found in the design of recommender systems that require and use only positive examples of the preferences of their users. The authors show, through application on real data, that artificial immune systems may be an efficient way to address such problems.

The book is addressed to any graduate student and researcher in computer science. As such, it is self-contained, with the necessary number of introductory chapters on learning and learning paradigms before specific chapters on artificial immune systems. I believe that the authors have done a good job on addressing the tackled issues. I consider the book a good addition to the areas of learning, bio-inspired computing and artificial immune systems. I am confident that it will help graduate students, researchers and practitioners to understand and expand artificial immune systems and apply them in real-world problems.

Dayton, OH, USA
June 2016

Prof.-Dr. Nikolaos G. Bourbakis
IEEE Fellow, President, Biological and
Artficial Intelligence Foundation (BAIF)
OBR Distinguished Professor of Informatics and
Technology and Director
Assistive Technologies Research Center and Director
Assistive Technologies Research Center

Preface

In the monograph at hand, we explore theoretical and experimental justification of the use of *Artificial Immune Systems* as a *Machine Learning Paradigm*. Our inspiration stems from the fact that vertebrates possess an immune system, consisting of highly complex biological structures and processes, that efficiently protect them against disease. A biological immune system is capable of detecting a wide variety of agents, including viruses, parasites and cancer cells, and distinguishing them from the organism's own healthy tissue. This is achieved in the *adaptive immune* subsystem of the immune system.

More specifically, the adaptive immune (sub)system continuously performs a self/non-self discrimination process. In machine learning terms, the adaptive immune system addresses *a pattern classification problem with extreme class imbalance*. Over the recent years, classification problems with class imbalance have attracted the interest of researchers worldwide. However, little attention has been paid so far to the use of artificial immune systems in addressing classification problems with a high or extreme degree of class imbalance.

We address the fundamental problems of pattern recognition, i.e. (*clustering*, *classification* and *one-class classification*), by developing artificial immune system-based machine learning algorithms. We measure the efficiency of these algorithms against state of the art pattern recognition paradigms such as *support vector machines*. Particular emphasis is placed on pattern classification in the context of the class imbalance problem. In machine learning terms, we address degenerated binary classification problems where the class of interest to be recognized is known through only a limited number of positive training instances. In other words, the target class occupies only a negligible volume of the entire pattern space, while the complementary space of negative patterns remains completely unknown during the training process. A practical application of this approach may be found in the design of recommender systems that require the use of only positive examples of the preferences of their users. We show, through application on real data, that artificial immune systems address such problems efficiently.

The general experimentation framework adopted throughout the current monograph is an open collection of one thousand (1000) pieces from ten (10) classes of

western music. This collection has been extensively used in applications concerning music information retrieval and music genre classification. The experimental results presented in this monograph demonstrate that the general framework of artificial immune system-based classification algorithms constitutes a valid machine learning paradigm for clustering, classification and one-class classification.

In order to make the book as self-contained as possible, we have divided it into two parts. Specifically, the first part of the book presents machine learning fundamentals and paradigms with an emphasis on one-class classification problems, while the second part is devoted to biological and artificial immune systems and their application to one-class classification problems. The reader, depending on his/her previous exposure to machine learning, may choose either to read the book from its beginning or to go directly to its second part. It is our hope that this monograph will help graduate students, researchers and practitioners to understand and expand artificial immune systems and apply them in real-world problems.

Piraeus, Greece
June 2016

Dionisios N. Sotiropoulos
George A. Tsihrintzis

Acknowledgments

We would like to thank Prof.-Dr. Lakhmi C. Jain for agreeing to include this monograph in the Intelligent Systems Reference Library (ISRL) book series of Springer that he edits. We would also like to thank Prof. Nikolaos Bourbakis of Wright State University, USA, for writing a foreword to the monograph. Finally, we would like to thank the Springer staff for their excellent work in typesetting and publishing this monograph.

Contents

Part II Artificial Immune Systems