Studies in Computational Intelligence, Volume 138

Editor-in-Chief
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 117. Da Ruan, Frank Hardeman and Klaas van der Meer (Eds.)
Intelligent Decision and Policy Making Support Systems, 2008
ISBN 978-3-540-78306-0

Vol. 118. Tsau Young Lin, Ying Xie, Anita Wasilewska and Churn-Jung Liau (Eds.)
Data Mining: Foundations and Practice, 2008
ISBN 978-3-540-78487-6

Vol. 119. Slawomir Wiak, Andrzej Krawczyk and Ivo Dolezel (Eds.)
Intelligent Computer Techniques in Applied Electromagnetics, 2008
ISBN 978-3-540-78489-0

Vol. 120. George A. Tsihrintzis and Lakhmi C. Jain (Eds.)
Multimedia Interactive Services in Intelligent Environments, 2008
ISBN 978-3-540-78491-3

Vol. 121. Nadia Nedjah, Leandro dos Santos Coelho and Luiza de Macedo Mourelle (Eds.)
Quantum Inspired Intelligent Systems, 2008
ISBN 978-3-540-78531-6

Vol. 122. Tomasz G. Smolinski, Mariofanna G. Milanova and Aboul-Ella Hassanien (Eds.)
Applications of Computational Intelligence in Biology, 2008
ISBN 978-3-540-78533-0

Vol. 123. Shuichi Iwata, Yukio Ohsawa, Shusaku Tsumoto, Ning Zhong, Yong Shi and Lorenzo Magnani (Eds.)
Communications and Discoveries from Multidisciplinary Data, 2008
ISBN 978-3-540-78732-7

Vol. 124. Ricardo Zavala Yoe
Modelling and Control of Dynamical Systems: Numerical Implementation in a Behavioral Framework, 2008
ISBN 978-3-540-78734-1

Vol. 125. Larry Bull, Bernadó-Mansilla Ester and John Holmes (Eds.)
Learning Classifier Systems in Data Mining, 2008
ISBN 978-3-540-78978-9

Vol. 126. Oleg Okun and Giorgio Valentini (Eds.)
Supervised and Unsupervised Ensemble Methods and their Applications, 2008
ISBN 978-3-540-78980-2

Vol. 127. Régie Gras, Einoshin Suzuki, Fabrice Guillet and Filippo Spagnolo (Eds.)
Statistical Implicative Analysis, 2008
ISBN 978-3-540-78982-6

Vol. 128. Fatos Xhafa and Ajith Abraham (Eds.)
Metaheuristics for Scheduling in Industrial and Manufacturing Applications, 2008
ISBN 978-3-540-78984-0

Vol. 129. Natalio Krasnogor, Giuseppe Nicosia, Mario Pavone and David Pelta (Eds.)
Nature Inspired Cooperative Strategies for Optimization (NICSO 2007), 2008
ISBN 978-3-540-78986-4

Vol. 130. Richi Nayak, Nikhil Ichalkaranje and Lakhmi C. Jain (Eds.)
Evolution of the Web in Artificial Intelligence Environments, 2008
ISBN 978-3-540-79139-3

Vol. 131. Roger Lee and Haeng-Kon Kim (Eds.)
Computer and Information Science, 2008
ISBN 978-3-540-79186-7

Vol. 132. Danil Prokhorov (Ed.)
Computational Intelligence in Automotive Applications, 2008
ISBN 978-3-540-79256-7

Vol. 133. Manuel Graña and Richard J. Duro (Eds.)
Computational Intelligence for Remote Sensing, 2008
ISBN 978-3-540-79352-6

Vol. 134. Ngoc Thanh Nguyen and Radoslaw Katarzyniak (Eds.)
New Challenges in Applied Intelligence Technologies, 2008
ISBN 978-3-540-79354-0

Vol. 135. Hsinchun Chen and Christopher C. Yang (Eds.)
Intelligence and Security Informatics, 2008
ISBN 978-3-540-69207-2

Vol. 136. Carlos Cotta, Marc Sevaux and Kenneth Sörensen (Eds.)
Adaptive and Multilevel Metaheuristics, 2008
ISBN 978-3-540-79437-0

Vol. 137. Lakhmi C. Jain, Mika Sato-Ilic, Maria Virvou, George A. Tsihrintzis, Valentina Emilia Balas and Canicious Abeynayake (Eds.)
Computational Intelligence Paradigms, 2008
ISBN 978-3-540-79473-8

Vol. 138. Bruno Apolloni, Witold Pedrycz, Simone Bassis and Dario Malchiodi
The Puzzle of Granular Computing, 2008
ISBN 978-3-540-79863-7

Bruno Apolloni
Witold Pedrycz
Simone Bassis
Dario Malchiodi

The Puzzle of Granular Computing

Springer

Bruno Apolloni
Dip. Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico, 39/41 20135 Milano
Italy

Witold Pedrycz
University of Alberta
Dept. Electrial & Computer Engineering
9107 116 Street
Edmonton AB T6G 2V4 Canada
& Systems Research Institute
Polish Academy of Sciences
Warsaw, Poland

Simone Bassis
Dip. Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico, 39/41
20135 Milano
Italy

Dario Malchiodi
Dip. Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico, 39/41
20135 Milano
Italy

ISBN 978-3-540-79863-7 e-ISBN 978-3-540-79864-4

DOI 10.1007/978-3-540-79864-4

Studies in Computational Intelligence ISSN 1860949X

Library of Congress Control Number: 2008926081

© 2008 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, 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.

Typeset & *Cover Design:* Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

9 8 7 6 5 4 3 2 1
springer.com

To those who are called upon to make decisions, practically the whole of mankind, politicians included. Faced with necessarily granular information, we dont expect people to arrive at the optimum decision. But we demand that they make reasonable choices.

Image created by Guglielmo Apolloni

Preface

The computer era is characterized by a massive shift in problem solving from the search of a suitable solving procedure to the careful exploitation of data information. If the most common recipe to solve a problem forty years ago was "look for the right formula", the common today's direction is "consider what suggestions data tell you about the problem solution". This data driven goal intensively exploited in the eighties in the connectionist paradigm, but even early adopted in the sixties by the fuzzy set paradigm, opened the door of an ideal megastore where a plenty of methods are promoted, each claiming high percentage of success in a vast variety of problems. Depending on the operational context and on the peculiarities that are enhanced of, they are alternatively declared methods of computational intelligence, to denote loose axiomatic premises compensated by the analogies with procedures that are supposed to be followed by intelligent (human or animal) beings and the appeal of the results they promise, or more recently methods of granular computing to root their rationale on the data information content. These looseness and approaching bivalence are not a defect *per se*, but very often produce an overlap in solving procedures leaving the user dubious on the solution way he has to follow in a specific problem.

Aim of this monograph is to help the reader to compose the puzzle pieces represented by these procedures into a high diversified mosaic of the modern methods for solving computational problems, within which the localization of a single tile is highly informative of its proper use. The strategy we adopted is to privilege the information management aspects in respect to the algorithmic sophistication. This is why we opted in the title for the second categorization of the methods in terms of information granules. Since our goal is rather ambitious, we will start with a very fundamental knowledge phenomenology with the commitment of

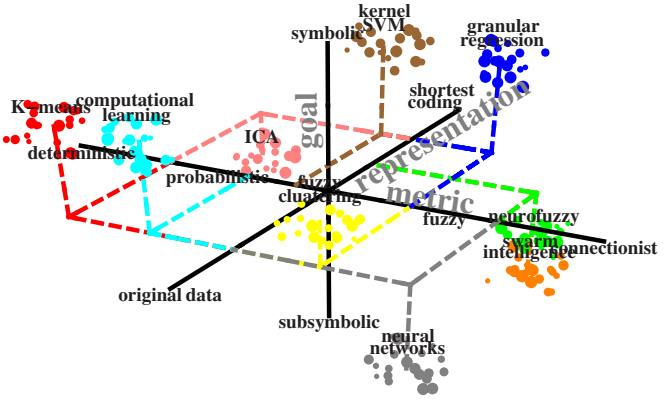

Fig. 0.1. Three dimensional framework representing an inference task. Clusters: some among the tasks afforded in the monography.

moving to more complex phenomena just as it becomes needed. Nevertheless the book is hard to read. It does not mean that it requires the reader to posses an extensive training in mathematics or in computer science. To a significant extent the book is self contained, apart some elementary mathematics and probability notions. Nevertheless, a coherent line of reasoning is developed along all the book demanding an accurate consideration on the part of the reader. With this caveat and guidelines, we expect the book to be beneficial both to the intellectually curious student and to the researcher open to modern conceptual paradigms.

We frame the matter of the book within the general goal of inferring rules from a suitable reading of a sample of data the rules refer to. The whole relies on an intelligent interplay between the properties we may observe in the data and the task we are called to solve on the basis of them, hence in a selection and management of data features that are highly dependent on their operational context and processing goal. The entire study can be cast in a three-dimensional framework (see Fig. 0.1), whose individual axes come with a well-defined semantics:

- The level of capability we have of stating a *metric* in terms of some order, or at least a certain degree of similarity between the data. This gives rise to a hierarchy of the methods in terms of available information content. Here the richest to poorest direction moves from deterministic, to probabilistic, then fuzzy and finally connectionist methods. We will follow this broadly defined hierarchy allowing for any kind of inversions and contaminations.

- The amount of structure we have embedded in these data, i.e., the features forming a *representation* of them. We will mainly quantify this information in terms of their distribution law or, dually, of the degree of structural connections existing therein.
- The typology of operational results we are expecting in face of the data at hand, i.e., the *goal* we aim to achieve. The taxonomy concerns both the kind of rule we want to discover and its description. As for the former, we distinguish between tasks such as classification, regression, etc. As for the latter, we mainly distinguish the cases where we are interested in a formal description of the rule, say through a formula or a procedure, from those where we are satisfied of having a computational device, for instance a set of instructions and parameters, producing correct solutions to a vast majority of problem instances.

Note that this framework is actually not far from the framework being used for clusterizing human emotional states, which is spanned by the three axes: coping, valence and novelty. Both frames are constituted of axes representing fuzzy variables to be suitably handled in view of a final goal rather than used to produce a definite label to emotions – there – or set of data in our case. Thus it is not surprising to realize that, once the tight cage of axiomatic theories has been abandoned, we are driven in our computations by complex thought paradigms that are not far from those ruling our emotions. Hence, on the one hand, it is not shocking to say that our computational framework may be viewed as a particular specification of the Affective Computing framework. On the other hand, let us reassure the reader that the key subject of our discourse will be the information granule, whose location in the above three dimensional framework specializes at level of the cluster of observed data items. In this setting, rules represent a suitable processing of granules.

Our underlying didactic strategy is to isolate fundamental conceptual bricks of Granular Computing in terms of some key problems and underlying methods, and discuss their rigorous implementation. Namely we will focus on the following problems:

1. classifying data through
   - decision rules, and
   - Support Vector Machines
2. extracting features denoting Relevant Components from data (ICA, kernels, etc.),
3. learning functions regressing data in
   - symbolic form, and
   - subsymbolic way.

What we have said so far is the logic description of the matter. In the book, however, we will proceed along the chapters with groups of homogeneous topics that should be better suitable to the reader thanks to the compression of the results we may deal with, that is a general goal for any learning system – human brain included. Therefore, we formerly devote one chapter to introduce the information granule as atomic unit of our procedures and declining its modalities. Then, we will devote a collection of chapters to the fundamentals. Namely, Part 1 is focused on the statistical ways of processing data, with chapters dedicated to: i) understanding what a sample really represents, ii) devising tools for exploiting it in terms of algorithmic inference, and iii) the most typical implementations of these tools, learning algorithms included. A second part is focused on the essentials of fuzzy sets, with chapters on : i) transferring information from the user to granules around data, and ii) refining granules from data evidence. Then, in order to prepare the reader to move to real applications, we toss the bases on three fundamental slots. Thus in Part III we expound the problem of classifying data through the two main sections on feature extraction and their classification, where the boundary between these two subjects is quite often delineated in a not clearly defined manner. Assuming that after these chapters the reader is able to isolate the atoms of his reasoning (*what she/he is speaking about*), we next introduce tools for combining these atoms/granules into structured sentences representing the rules of the ontologies he is setting up. They represent the logical and operational tools we may consider for interacting with the environment. It is commonly recognized that most rules are inherently subsymbolic: connectionist according to the eighties cognitive paradigms, or the product of social computations to a wider extent, within the recent threads. This does not mean that are meaningless rules. On the contrary, they spring from a rigorously cultivated thinking attitude finding a synthesis level that is wider in respect to strictly formal theories. With this perspective, we discuss in the fourth Part some key paradigms of social computing. We include a chapter on evolutionary computing and neural networks with some successful training algorithms, and a chapter on swarm intelligence strongholds. A compromise way between algorithms mainly aimed at the dry successful results and human brain need of understanding methods to get success is treated in fifth Part. It runs in terms structured information granules leading to fuzzy rules as a synthesis of deductions from user experience and inductions from data observations. We devote a chapter to the conceptual aspects and another to computational methods. Finally in the

last chapter of this book we envision the role of knowledge engineers engaged to use all what we can of the tools assessed in the previous chapters in order to solve complex problems. The core is represented by some general fusion strategies. In particular we will deepen two case studies on support vector machines specially featured to deal with data quality, and collaborative clustering as a strategy to generate consistent classification rules.

Contents

The General Framework

1 Granule Formation Around Data 3
1.1 The Roots of the Statistical Inference 5
1.1.1 From Granules to Distributions 8
1.1.2 Learning Distribution Parameters 9
1.2 The Borders of the Unawareness 15
1.3 From Data to Functions 19
1.4 Functions of Granules 22
1.5 The Benchmarks of This Book 25
1.5.1 Synthetic Datasets 25
1.5.2 Leukemia 25
1.5.3 Cox 26
1.5.4 Guinea Pig 26
1.5.5 SMSA 27
1.5.6 Swiss 28
1.5.7 Iris 28
1.5.8 MNIST Digits Data Base 28
1.5.9 Vote 29
1.5.10 Sonar Signals 29
1.6 Exercises 30

Further Reading 33

References 37

Part I: Algorithmic Inference

Introduction 43

2 Modeling Samples 45
2.1 Samples and Properties about Them 47
2.1.1 Samples 47
2.1.2 Properties about Samples 52
2.2 Organizing Granules into Populations 55
2.3 Blowing up Items into Populations 59
2.4 Conclusions 62
2.5 Exercises 63

3 Inferring from Samples 65
3.1 Learning from Samples 65
3.1.1 Statistic Identification 69
3.1.2 Parameter Distribution Law 71
3.1.3 Drawing Decisions 78
3.2 Learning from Examples 84
3.2.1 Confidence Regions 85
3.2.2 Numerical Examples 90
3.2.3 PAC Learning Theory Revisited 96
3.2.4 A General Error Distribution Law 103
3.3 Conclusions 113
3.4 Exercises 114

Further Reading 117

References 121

Part II: The Development of Fuzzy Sets

Introduction 127

4 Construction of Information Granules: From Data and Perceptions to Information Granules 129
4.1 Semantics of Fuzzy Sets: Some Insights 130
4.2 Transferring Meaning into Fuzzy Sets 131
4.2.1 Fuzzy Set as a Descriptor of Feasible Solutions 131
4.2.2 Fuzzy Set as a Descriptor of the Notion of *Typicality* 133
4.2.3 Membership Functions for Visualizing Solution Preferences 135
4.2.4 Nonlinear Transformation of Templates 136
4.2.5 Shadowed Sets as a Three-Valued Logic Characterization of Fuzzy Sets 138

4.2.6 Interval-Valued Fuzzy Sets 143
4.3 Fuzzy Sets of Higher Order 144
4.3.1 Second Order Fuzzy Sets 144
4.3.2 Linguistic Approximation 146
4.3.3 Rough Fuzzy Sets and Fuzzy Rough Sets 147
4.4 Type-2 Fuzzy Sets 149
4.5 Conclusions 152
4.6 Exercises 152

5 Estimating Fuzzy Sets 155
5.1 Vertical and Horizontal Schemes of Membership Estimation 156
5.2 Maximum Likelihood Estimate 160
5.2.1 In Search of the Maximum Compliance between Pivotal Granules and a Model 162
5.3 Saaty's Priority Method of Pairwise Membership Function Estimation 165
5.4 Fuzzy Sets as Granular Representatives of Numeric Data 168
5.5 Fuzzy Sets That Reduce the Descriptional Length of a Formula 175
5.6 Fuzzy Equalization 179
5.7 Conclusions 181
5.8 Exercises 182

Further Reading 185

References 187

Part III: Expanding Granula into Boolean Functions

Introduction 191

6 The Clustering Problem 193
6.1 The Roots of Cluster Analysis 193
6.2 Genotype Value of Metrics 201
6.3 Clustering Algorithms 204
6.3.1 A Template Procedure 205
6.3.2 Initial Numbers and Values of Centroids 210
6.3.3 Aggregation vs. Agglomeration 212
6.3.4 Implementation Mode and Stopping Rule 219
6.4 The Weighted Sum Option 220

6.4.1 The Fuzzy C-Means Algorithm 222
6.5 Discriminant Analysis 227
6.6 Conclusions 230
6.7 Exercises 231

7 Suitably Representing Data 233
7.1 Principal Components Analysis 235
7.2 Independent Components 237
7.3 Sparse Representations 248
7.3.1 Incremental Clustering 249
7.3.2 Analytical Solution 255
7.4 Nonlinear Mappings 256
7.5 Conclusions 261
7.6 Exercises 262

Further Reading 265

References 269

Part IV: Directing Populations

Introduction 275

8 The Main Paradigms of Social Computation 277
8.1 Extending Sampling Mechanisms to Generate Tracks 278
8.2 Knowledge Representation: From Phenotype to Genotype Space 281
8.3 Starting from an Artificial Population 284
8.3.1 A Very Schematic Genetic Algorithm 285
8.4 The Seeds of Unawareness 291
8.4.1 The Neural Networks Paradigm 292
8.4.2 Training a Boltzmann Machine 296
8.4.3 Multilayer Perceptron 300
8.4.4 From a Population of Genes to a Population of Experts 307
8.5 Conclusions 313
8.6 Exercises 314

9 If also Ants Are Able... 317
9.1 Swarm Intelligence 318
9.1.1 Ant Colony 319
9.1.2 Swarm Optimization 322
9.2 Immune Systems 325
9.3 Aging the Organisms 327

9.3.1 The Neuron's Life 328
9.4 Some Practical Design and Implementation 332
9.5 Conclusions 334

Further Reading 335

References 337

Part V: Granular Constructs

Introduction 343

10 Granular Constructs 345
10.1 The Architectural Blueprint of Fuzzy Models 345
10.2 The Rule System 348
10.2.1 Designing a Rule System Layout 349
10.2.2 Evaluating a Rule System 352
10.3 The Master Scheme 353
10.4 The Clustering Part 355
10.4.1 The Cluster-Based Representation of the Input-Output Mappings 356
10.4.2 The Context-Based Clustering in the Development of Granular Models 363
10.5 The Connectionist Part 371
10.5.1 Granular Neuron 371
10.6 The Neural Architecture 374
10.6.1 Defuzzification at the Last 375
10.6.2 Use Numbers as Soon as Possible 378
10.7 Conclusions 380
10.8 Exercises 381

11 Identifying Fuzzy Rules 385
11.1 Training of a Granular Neuron 385
11.2 Partially Supervised Identification of Fuzzy Clusters 388
11.2.1 Using Data Labels When They Are Known 389
11.2.2 The Development of Human-Centric Clusters 392
11.3 The Overall Training of a Neuro-fuzzy System 396
11.3.1 Learning Algorithm for Neuro-fuzzy Systems 398
11.4 Conclusions 403
11.5 Exercises 404

Further Reading 405

References 407

A Conceptual Synthesis

12 Knowledge Engineering 411
12.1 Separating Points through SVM 411
12.1.1 Binary Linear Classification 412
12.1.2 Regression 422
12.2 Distributed Architectures for Granular Computing 431
12.2.1 Collaborative Clustering 433
12.2.2 The General Flow of Collaborative Processing 435
12.2.3 Evaluation of the Quality of Collaboration 437
12.2.4 Experience-Consistent Fuzzy Models 439
12.3 Conclusions 446

References 449

Appendices 451

A Norms 453
A.1 Norms 453
A.2 General Norms 453
A.3 Fuzzy Norms 460

B Some Statistical Distribution 463
B.1 Some One-Dimensional Variables 463
B.2 Multidimensional Gaussian Variable 466

C Some Membership Functions 469

D Some Algebra 471
D.1 Eigenvalues and Eigenvectors 472
D.2 Kernels 472

E List of Symbols 477
E.1 Variables 477
E.2 Functions 479
E.3 Operations on Sets 479
E.4 Operators 480

Index 481