Studies in Fuzziness and Soft Computing, Volume 229

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

Vol. 214. Irina Georgescu
Fuzzy Choice Functions, 2007
ISBN 978-3-540-68997-3

Vol. 215. Paul P. Wang, Da Ruan,
Etienne E. Kerre (Eds.)
Fuzzy Logic, 2007
ISBN 978-3-540-71257-2
Vol. 216. Rudolf Seising
The Fuzzification of Systems, 2007
ISBN 978-3-540-71794-2

Vol. 217. Masoud Nikravesh, Janusz Kacprzyk,
Lofti A. Zadeh (Eds.)
Forging New Frontiers: Fuzzy Pioneers I, 2007
ISBN 978-3-540-73181-8

Vol. 218. Masoud Nikravesh, Janusz Kacprzyk,
Lofti A. Zadeh (Eds.)
Forging New Frontiers: Fuzzy Pioneers II, 2007
ISBN 978-3-540-73184-9

Vol. 219. Roland R. Yager, Liping Liu (Eds.)
Classic Works of the Dempster-Shafer Theory of Belief Functions, 2007
ISBN 978-3-540-25381-5

Vol. 220. Humberto Bustince,
Francisco Herrera, Javier Montero (Eds.)
Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, 2007
ISBN 978-3-540-73722-3

Vol. 221. Gleb Beliakov, Tomasa Calvo,
Ana Pradera
Aggregation Functions: A Guide for Practitioners, 2007
ISBN 978-3-540-73720-9

Vol. 222. James J. Buckley,
Leonard J. Jowers
Monte Carlo Methods in Fuzzy Optimization, 2008
ISBN 978-3-540-76289-8

Vol. 223. Oscar Castillo, Patricia Melin
Type-2 Fuzzy Logic: Theory and Applications, 2008
ISBN 978-3-540-76283-6

Vol. 224. Rafael Bello, Rafael Falcón,
Witold Pedrycz, Janusz Kacprzyk (Eds.)
Contributions to Fuzzy and Rough Sets Theories and Their Applications, 2008
ISBN 978-3-540-76972-9

Vol. 225. Terry D. Clark, Jennifer M. Larson,
John N. Mordeson, Joshua D. Potter,
Mark J. Wierman
Applying Fuzzy Mathematics to Formal Models in Comparative Politics, 2008
ISBN 978-3-540-77460-0

Vol. 226. Bhanu Prasad (Ed.)
Soft Computing Applications in Industry, 2008
ISBN 978-3-540-77464-8

Vol. 227. Eugene Roventa, Tiberiu Spircu
Management of Knowledge Imperfection in Building Intelligent Systems, 2008
ISBN 978-3-540-77462-4

Vol. 228. Adam Kasperski
Discrete Optimization with Interval Data, 2008
ISBN 978-3-540-78483-8

Vol. 229. Sadaaki Miyamoto,
Hidetomo Ichihashi, Katsuhiro Honda
Algorithms for Fuzzy Clustering, 2008
ISBN 978-3-540-78736-5

Sadaaki Miyamoto,
Hidetomo Ichihashi, Katsuhiro Honda

Algorithms for Fuzzy Clustering

Methods in c-Means Clustering with Applications

Springer

Authors

Dr. Sadaaki Miyamoto
University of Tsukuba
Inst. Information Sciences
and Electronics
Ibaraki
305-8573 Japan
Email: miyamoto@risk.tsukuba.ac.jp

Dr. Katsuhiro Honda
Osaka Prefecture University
Graduate School of Engineering
1-1 Gakuen-cho
Sakai
Osaka, 599-8531 Japan

Dr. Hidetomo Ichihashi
Osaka Prefecture University
Graduate School of Engineering
1-1 Gakuen-cho
Sakai
Osaka, 599-8531 Japan

ISBN 978-3-540-78736-5 e-ISBN 978-3-540-78737-2

DOI 10.1007/978-3-540-78737-2

Studies in Fuzziness and Soft Computing ISSN 1434-9922

Library of Congress Control Number: 2008922722

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

Preface

Recently many researchers are working on cluster analysis as a main tool for exploratory data analysis and data mining. A notable feature is that specialists in different fields of sciences are considering the tool of data clustering to be useful. A major reason is that clustering algorithms and software are flexible in the sense that different mathematical frameworks are employed in the algorithms and a user can select a suitable method according to his application. Moreover clustering algorithms have different outputs ranging from the old dendrograms of agglomerative clustering to more recent self-organizing maps. Thus, a researcher or user can choose an appropriate output suited to his purpose, which is another flexibility of the methods of clustering.

An old and still most popular method is the K-means which use K cluster centers. A group of data is gathered around a cluster center and thus forms a cluster. The main subject of this book is the fuzzy c-means proposed by Dunn and Bezdek and their variations including recent studies. A main reason why we concentrate on fuzzy c-means is that most methodology and application studies in fuzzy clustering use fuzzy c-means, and fuzzy c-means should be considered to be a major technique of clustering in general, regardless whether one is interested in fuzzy methods or not. Moreover recent advances in clustering techniques are rapid and we require a new textbook that includes recent algorithms. We should also note that several books have recently been published but the contents do not include some methods studied herein.

Unlike most studies in fuzzy c-means, what we emphasize in this book is a family of algorithms using entropy or entropy-regularized methods which are less known, but we consider the entropy-based method to be another useful method of fuzzy c-means. For this reason we call the method of fuzzy c-means by Dunn and Bezdek as the standard method to distinguish it from the entropy-based method. Throughout this book one of our intentions is to uncover theoretical and methodological differences between the standard method and the entropy-based method. We do note claim that the entropy-based method is better than the standard method, but we believe that the methods of fuzzy c-means become *complete* by adding the entropy-based method to the standard one by Dunn

and Bezdek, since we can observe natures of the both methods more deeply by contrasting these two methods.

Readers will observe that the entropy-based method is similar to the statistical model of Gaussian mixture distribution since both of them are using the error functions, while the standard method is very different from a statistical model. For this reason the standard method is purely fuzzy while the entropy-based method connects a statistical model and a fuzzy model.

The whole text is divided into two parts: The first part that consists of Chapters 1∼5 is theoretical and discusses basic algorithms and variations. This part has been written by Sadaaki Miyamoto.

The second part is application-oriented. Chapter 6 which has been written by Hidetomo Ichihashi studies classifier design; Katsuhiro Honda has written Chapters 7∼9 where clustering algorithms are applied to a variety of methods in multivariate analysis.

The authors are grateful to Prof. Janusz Kacprzyk, the editor, for his encouragement to contribute this volume to this series and helpful suggestions throughout the publication process. We also thank Dr. Mika Sato-Ilic and Dr. Yasunori Endo for their valuable comments to our works.

We finally note that studies related to this book have partly been supported by the Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science, No.16300065.

January 2008

Sadaaki Miyamoto
Hidetomo Ichihashi
Katsuhiro Honda

Contents