

Mikhail Ju. Moshkov, Marcin Piliszczyk, and Beata Zielosko

Partial Covers, Reducts and Decision Rules in Rough Sets

Studies in Computational Intelligence, Volume 145

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. 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 (NICO 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

Vol. 139. Jan Drugowitsch
Design and Analysis of Learning Classifier Systems, 2008
ISBN 978-3-540-79865-1

Vol. 140. Nadia Magnenat-Thalmann, Lakhmi C. Jain and N. Ichalkaranje (Eds.)
New Advances in Virtual Humans, 2008
ISBN 978-3-540-79867-5

Vol. 141. Christa Sommerer, Lakhmi C. Jain and Laurent Mignonneau (Eds.)
The Art and Science of Interface and Interaction Design, 2008
ISBN 978-3-540-79869-9

Vol. 142. George A. Tsihrintzis, Maria Virvou, Robert J. Howlett and Lakhmi C. Jain (Eds.)
New Directions in Intelligent Interactive Multimedia, 2008
ISBN 978-3-540-68126-7

Vol. 143. Uday K. Chakraborty (Ed.)
Advances in Differential Evolution, 2008
ISBN 978-3-540-68827-3

Vol. 144. Andreas Fink and Franz Rothlauf (Eds.)
Advances in Computational Intelligence in Transport, Logistics, and Supply Chain Management, 2008
ISBN 978-3-540-69024-5

Vol. 145. Mikhail Ju. Moshkov, Marcin Piliszczyk and Beata Zielosko
Partial Covers, Reducts and Decision Rules in Rough Sets, 2008
ISBN 978-3-540-69027-6

Mikhail Ju. Moshkov
Marcin Piliszczyk
Beata Zielosko

Partial Covers, Reducts and Decision Rules in Rough Sets

Theory and Applications



Springer

Mikhail Ju. Moshkov
Institute of Computer Science
University of Silesia
39, Będzińska St.
Sosnowiec, 41-200
Poland
Email: moshkov@us.edu.pl

Marcin Piliszczuk
ING Bank Śląski S.A.
34, Sokolska St.
Katowice, 40-086
Poland
Email: marcin.piliszczuk@ingbank.pl

Beata Zielosko
Institute of Computer Science
University of Silesia
39, Będzińska St.
Sosnowiec, 41-200
Poland
Email: zielosko@us.edu.pl

ISBN 978-3-540-69027-6

e-ISBN 978-3-540-69029-0

DOI 10.1007/978-3-540-69029-0

Studies in Computational Intelligence

ISSN 1860949X

Library of Congress Control Number: 2008927877

© 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 our families

Preface

This monograph is devoted to theoretical and experimental study of partial reducts and partial decision rules on the basis of the study of partial covers. The use of partial (approximate) reducts and decision rules instead of exact ones allows us to obtain more compact description of knowledge contained in decision tables, and to design more precise classifiers.

We consider algorithms for construction of partial reducts and partial decision rules, bounds on minimal complexity of partial reducts and decision rules, and algorithms for construction of the set of all partial reducts and the set of all irreducible partial decision rules. We discuss results of numerous experiments with randomly generated and real-life decision tables. These results show that partial reducts and decision rules can be used in data mining and knowledge discovery both for knowledge representation and for prediction.

The results obtained in the monograph can be useful for researchers in such areas as machine learning, data mining and knowledge discovery, especially for those who are working in rough set theory, test theory and LAD (Logical Analysis of Data).

The monograph can be used under the creation of courses for graduate students and for Ph.D. studies.

An essential part of software used in experiments will be accessible soon in RSES – Rough Set Exploration System (Institute of Mathematics, Warsaw University, head of project – Professor Andrzej Skowron).

We are greatly indebted to Professor Andrzej Skowron for stimulated discussions and various support of our work. We are grateful to Professor Janusz Kacprzyk for helpful suggestions.

Sosnowiec, Poland
April 2008

Mikhail Ju. Moshkov
Marcin Piliszczuk
Beata Zielosko

Contents

Introduction	1
1 Partial Covers, Reducts and Decision Rules	7
1.1 Partial Covers	8
1.1.1 Main Notions	8
1.1.2 Known Results	9
1.1.3 Polynomial Approximate Algorithms	10
1.1.4 Bounds on $C_{\min}(\alpha)$ Based on Information about Greedy Algorithm Work	13
1.1.5 Upper Bound on $C_{\text{greedy}}(\alpha)$	17
1.1.6 Covers for the Most Part of Set Cover Problems	18
1.2 Partial Tests and Reducts	22
1.2.1 Main Notions	22
1.2.2 Relationships between Partial Covers and Partial Tests	23
1.2.3 Precision of Greedy Algorithm	24
1.2.4 Polynomial Approximate Algorithms	25
1.2.5 Bounds on $R_{\min}(\alpha)$ Based on Information about Greedy Algorithm Work	26
1.2.6 Upper Bound on $R_{\text{greedy}}(\alpha)$	28
1.2.7 Tests for the Most Part of Binary Decision Tables	29
1.3 Partial Decision Rules	35
1.3.1 Main Notions	36
1.3.2 Relationships between Partial Covers and Partial Decision Rules	37
1.3.3 Precision of Greedy Algorithm	38
1.3.4 Polynomial Approximate Algorithms	38
1.3.5 Bounds on $L_{\min}(\alpha)$ Based on Information about Greedy Algorithm Work	40

1.3.6	Upper Bound on $L_{\text{greedy}}(\alpha)$	43
1.3.7	Decision Rules for the Most Part of Binary Decision Tables	43
1.4	Conclusions	49
2	Partial Covers, Reducts and Decision Rules with Weights	51
2.1	Partial Covers with Weights	53
2.1.1	Main Notions	53
2.1.2	Some Known Results	54
2.1.3	Polynomial Approximate Algorithms	55
2.1.4	Comparison of Usual Greedy Algorithm and Greedy Algorithm with Two Thresholds	56
2.1.5	Two Modifications of Greedy Algorithm	59
2.1.6	Lower Bound on $C_{\min}(\alpha)$	61
2.1.7	Upper Bounds on $C_{\text{greedy}}^{\gamma}(\alpha)$	64
2.1.8	Results of Experiments for α -Covers	66
2.2	Partial Tests and Reducts with Weights	70
2.2.1	Main Notions	71
2.2.2	Relationships between Partial Covers and Partial Tests	72
2.2.3	Precision of Greedy Algorithm with Equal Thresholds	73
2.2.4	Polynomial Approximate Algorithms	74
2.2.5	Two Modifications of Greedy Algorithm	74
2.2.6	Bounds on $R_{\min}(\alpha)$ and $R_{\text{greedy}}^{\gamma}(\alpha)$	77
2.2.7	Results of Experiments for α -Tests and α -Reducts	78
2.3	Partial Decision Rules with Weights	83
2.3.1	Main Notions	83
2.3.2	Relationships between Partial Covers and Partial Decision Rules	84
2.3.3	Precision of Greedy Algorithm with Equal Thresholds	86
2.3.4	Polynomial Approximate Algorithms	86
2.3.5	Two Modifications of Greedy Algorithm	87
2.3.6	Bounds on $L_{\min}(\alpha)$ and $L_{\text{greedy}}^{\gamma}(\alpha)$	89
2.3.7	Results of Experiments for α -Decision Rules	91
2.4	Conclusions	96
3	Construction of All Irreducible Partial Covers, All Partial Reducts and All Irreducible Partial Decision Rules	97
3.1	Irreducible t -Covers	98
3.1.1	Cardinality of Irreducible t -Covers	98
3.1.2	Number of Irreducible t -Covers	100
3.1.3	Algorithms for Construction of All Irreducible t -Covers	101

3.2	Set $TAB_D(m, n)$ of Decision Tables	101
3.3	t -Reducts	102
3.3.1	Cardinality of t -Reducts	102
3.3.2	Number of t -Reducts	105
3.3.3	Algorithms for Construction of All t -Reducts	106
3.3.4	Results of Experiments	107
3.4	Irreducible t -Decision Rules	109
3.4.1	Length of Irreducible t -Decision Rules	109
3.4.2	Number of Irreducible t -Decision Rules	112
3.4.3	Algorithms for Construction of All Irreducible t -Decision Rules	113
3.4.4	Results of Experiments	114
3.5	Conclusions	116
4	Experiments with Real-Life Decision Tables	117
4.1	0.5-Hypothesis for Reducts	118
4.2	0.5-Hypothesis for Decision Rules	122
4.3	Classifiers Based on Partial Reducts	126
4.4	Classifiers Based on Partial Decision Rules	129
4.5	Conclusions	133
5	Universal Attribute Reduction Problem	135
5.1	From Data Table to Decision Table	136
5.2	Problem of Attribute Reduction	137
5.2.1	Definition of Problem	137
5.2.2	Examples	138
5.2.3	Maximally Discerning Reducts	139
5.3	Greedy Algorithm	139
5.3.1	Precision of Greedy Algorithm	140
5.3.2	Polynomial Approximate Algorithms	140
5.3.3	Lower Bound on $R_{\min}(\alpha)$	141
5.3.4	Upper Bound on $R_{\text{greedy}}(\alpha)$	141
5.4	Conclusions	142
	Final Remarks	143
	References	145
	Index	149