

Larry Bull, Ester Bernadó-Mansilla and John Holmes (Eds.)

Learning Classifier Systems in Data Mining

Studies in Computational Intelligence, Volume 125

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. 103. Devendra K. Chaturvedi
Soft Computing Techniques and its Applications in Electrical Engineering, 2008
ISBN 978-3-540-77480-8
- Vol. 104. Maria Virvou and Lakhmi C. Jain (Eds.)
Intelligent Interactive Systems in Knowledge-Based Environment, 2008
ISBN 978-3-540-77470-9
- Vol. 105. Wolfgang Guenther
Enhancing Cognitive Assistance Systems with Inertial Measurement Units, 2008
ISBN 978-3-540-76996-5
- Vol. 106. Jacqueline Jarvis, Dennis Jarvis, Ralph Rönquist and Lakhmi C. Jain (Eds.)
Holonic Execution: A BDI Approach, 2008
ISBN 978-3-540-77478-5
- Vol. 107. Margarita Sordo, Sachin Vaidya and Lakhmi C. Jain (Eds.)
Advanced Computational Intelligence Paradigms in Healthcare - 3, 2008
ISBN 978-3-540-77661-1
- Vol. 108. Vito Trianni
Evolutionary Swarm Robotics, 2008
ISBN 978-3-540-77611-6
- Vol. 109. Panagiotis Chountas, Ilias Petrounias and Janusz Kacprzyk (Eds.)
Intelligent Techniques and Tools for Novel System Architectures, 2008
ISBN 978-3-540-77621-5
- Vol. 110. Makoto Yokoo, Takayuki Ito, Minjie Zhang, Juhnyoung Lee and Tokuro Matsuo (Eds.)
Electronic Commerce, 2008
ISBN 978-3-540-77808-0
- Vol. 111. David Elmakias (Ed.)
New Computational Methods in Power System Reliability, 2008
ISBN 978-3-540-77810-3
- Vol. 112. Edgar N. Sanchez, Alma Y. Alanís and Alexander G. Loukianov
Discrete-Time High Order Neural Control: Trained with Kalman Filtering, 2008
ISBN 978-3-540-78288-9
- Vol. 113. Gemma Bel-Enguix, M. Dolores Jimenez-Lopez and Carlos Martín-Vide (Eds.)
New Developments in Formal Languages and Applications, 2008
ISBN 978-3-540-78290-2

- Vol. 114. Christian Blum, Maria José Blesa Aguilera, Andrea Roli and Michael Sampels (Eds.)
Hybrid Metaheuristics, 2008
ISBN 978-3-540-78294-0
- Vol. 115. John Fulcher and Lakhmi C. Jain (Eds.)
Computational Intelligence: A Compendium, 2008
ISBN 978-3-540-78292-6
- Vol. 116. Ying Liu, Aixun Sun, Han Tong Loh, Wen Feng Lu and Ee-Peng Lim (Eds.)
Advances of Computational Intelligence in Industrial Systems, 2008
ISBN 978-3-540-78296-4
- 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 Liao (Eds.)
Data Mining: Foundations and Practice, 2008
ISBN 978-3-540-78487-6
- Vol. 119. Sławomir 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 (Ed.)
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

Larry Bull
Ester Bernadó-Mansilla
John Holmes
(Eds.)

Learning Classifier Systems in Data Mining

With 65 Figures and 53 Tables



Springer

Larry Bull
School of Computer Science
University of the West of England
Bristol BS16 1QY
U.K.
Larry.Bull@uwe.ac.uk

John Holmes
Centre for Clinical Epidemiology
and Biostatistics
University of Pennsylvania
Philadelphia, PA 19104
U.S.A.
jholmes@cceb.med.upenn.edu

Ester Bernadó-Mansilla
Enginyeria i Arquitectura La Salle
Universitat Ramon Llull
08022 Barcelona
Spain
esterb@salleURL.edu

ISBN 978-3-540-78978-9

e-ISBN 978-3-540-78979-6

Studies in Computational Intelligence ISSN 1860-949X

Library of Congress Control Number: 2008924365

© 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-Verlag. 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.

Cover design: Deblik, Berlin, Germany

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Contents

Learning Classifier Systems in Data Mining:

An Introduction

<i>Larry Bull, Ester Bernadó-Mansilla, and John Holmes</i>	1
1 Introduction	1
2 Holland's LCS	2
3 Wilson's ZCS	4
4 Wilson's XCS	6
5 Pittsburgh-Style LCS	7
6 Previous Research on LCS in Data Mining.....	8
7 Learning Classifier Systems in Data Mining: An Overview	10
8 Summary.....	12
References	12

Data Mining in Proteomics with Learning Classifier Systems

<i>Jaume Bacardit, Michael Stout, Jonathan D. Hirst, and Natalio Krasnogor</i>	17
1 Introduction	17
2 Problem Definition	19
2.1 Protein Structure and Coordination Number Prediction	19
2.2 HP Models.....	21
3 The GAssist Learning Classifier System	22
4 Experimental Framework	24
4.1 Experimental Battery I: Primary Sequence and Coordination Number	24
4.2 Experimental Battery II: Assessment of Input Information Sources and Class Partitions for Coordination Number	26
5 Results.....	29
5.1 Experimental Battery I.....	29
5.2 Experimental Battery II	31

6	Discussion	36
6.1	Battery of Experiments I	36
6.2	Battery of Experiments II	39
6.3	Brief Description of Other Related Work	40
7	Conclusions and Further Work	41
	References	42

Improving Evolutionary Computation Based Data-Mining for the Process Industry: The Importance of Abstraction

	<i>William N.L. Browne</i>	47
1	Introduction	47
2	Background	49
2.1	Description of the HSM and Downcoilers of Llanwern Works ..	49
2.2	LCS for Data Mining	49
3	Development of iLCS	50
4	Initial Data-Mining Results	52
4.1	Discovery Training Performance	53
4.2	Discovery Training Results	54
4.3	Advisory Performance	56
4.4	Summary of Industrial LCS	56
5	Abstraction	59
5.1	Abstraction Algorithm Overview	60
6	Results	62
6.1	Q-Learning and Standard XCS	63
6.2	Modified XCS With and Without Abstraction	63
6.3	Effect of Abstraction	64
6.4	Robustness of the Systems	65
7	Conclusions	66
	References	66

Distributed Learning Classifier Systems

	<i>Hai H. Dam, Pornthep Rojanavasu, Hussein A. Abbass, and Chris Lokan</i>	69
1	Introduction	69
2	Distributed Data Mining	72
2.1	Overview	72
2.2	Learning in Distributed Environments	73
3	Learning Classifier Systems	75
4	Distributed Learning Classifier Systems (DLCS)	76
4.1	The Framework of DLCS	76
4.2	Knowledge Combination at the Server	77
4.3	Traffic Load	78
5	Experimental Setup	79
6	Combining Local Predictions at the Server	79
6.1	Noise Free Environments	80
6.2	Noisy Environments	81

7	An Effect of the Number of Clients	83
7.1	Noise Free Environments	83
7.2	Noisy Environments	84
8	Investigation of Knowledge Passing in DLCS	85
8.1	Between Clients and Server	85
8.2	Between Clients	87
9	Conclusion	89
	References	90

Knowledge Discovery from Medical Data: An Empirical Study with XCS

	<i>Faten Kharbat, Mohammed Odeh, and Larry Bull</i>	93
1	Introduction	93
2	The Four-Phases Discovery Process	94
3	Phase 1: Understanding the Dataset	95
4	Phase 2: Data Preparation	96
4.1	Data Pre-Processing	96
4.2	Data Formatting and Decoding	97
4.3	The Imbalance Problem	97
4.4	Missing Data Problem	99
5	Phase 3: Data Mining and Knowledge Discovery	100
5.1	Well-Known Classification Techniques Results	100
5.2	XCS Results	101
6	Phase 4: Evaluation of the Discovered Knowledge	110
6.1	Analysis of C4.5 Results	111
6.2	Analysis of XCS Results	113
7	Conclusion	116
	References	116

Mining Imbalanced Data with Learning Classifier Systems

	<i>Albert Orriols-Puig and Ester Bernadó-Mansilla</i>	123
1	Introduction	123
2	Description of XCS	125
2.1	Knowledge Representation	125
2.2	Performance Component	125
2.3	Parameter's Update	126
2.4	Discovery Component	126
3	XCS and Class Imbalances	126
3.1	The Imbalanced Multiplexer	126
3.2	XCS on the Imbalanced Multiplexer	127
4	Modeling Parameter's Bounds	128
4.1	Imbalance Bound	128
4.2	Theoretical and Experimental Bounds: Analysis of the Deviation	130
4.3	Learning Rate and Error Estimates	130

4.4	Occurrence-Based Reproduction	132
4.5	Guidelines for Parameters Configuration	134
5	Online Configuration of XCS to Handle Imbalanced Problems	136
5.1	Online Adaptation Algorithm	136
5.2	Results	138
6	LCS for Mining Imbalanced Datasets	138
6.1	Methodology	139
6.2	Results	139
7	Summary and Conclusions	142
	References	143

XCS for Fusing Multi-Spectral Data in Automatic Target Recognition

Avinash Gandhe, Ssu-Hsin Yu, Raman Mehra, and Robert E. Smith 147

1	Introduction	147
2	Description of Datasets	148
2.1	Datasets Examined	148
2.2	RIT WASP Sensor Data	148
3	Proposed Classification Scheme	150
3.1	Registration of RIT WASP Data	151
3.2	The XCS Classifier	152
4	XCS as a Learning Classifier System	153
5	Feature Representation	154
6	XCS Based Classification Algorithm	155
7	Detection	156
7.1	Spatial Processing	157
7.2	Spectral Processing	158
7.3	Classification	161
8	Final Comments and Future Directions	166
	References	166

Foreign Exchange Trading Using a Learning Classifier System

Christopher Stone and Larry Bull 169

1	Introduction	169
2	Foreign Exchange Trading	171
2.1	Technical Analysis	171
2.2	Foreign Exchange Transactions	171
2.3	Transaction Costs and Slippage	172
2.4	Genetic Programming Approach	172
3	Learning Classifier System Approach	173
3.1	Representation	173
3.2	Excess Return	176
3.3	Transaction Costs	177
4	Experiments	177
4.1	Experimental Setup	177

4.2	Validation Period from 1981 to 1995	180
4.3	Validation Period from 1975 to 1995	182
5	Conclusions	186
	References	187

Towards Clustering with Learning Classifier Systems

	<i>Kreangsak Tamee, Larry Bull, and Owen Pinngern</i>	191
1	Introduction	191
2	A Simple LCS for Clustering	192
3	Initial Performance	194
4	Rule Compaction	195
5	Modifying XCS for Clustering	197
6	Local Search	198
7	Adaptive Threshold Parameter	199
8	Increased Complexity	201
9	Conclusions	201
	References	203

A Comparative Study of Several Genetic-Based Supervised Learning Systems

	<i>Albert Orriols-Puig, Jorge Casillas, and Ester Bernadó-Mansilla</i>	205
1	Introduction	205
2	Genetic-Based Machine Learning	206
3	Description the GBML Systems Used in the Comparison	208
	3.1 UCS	208
	3.2 GAssist	209
	3.3 SLAVE	210
	3.4 Fuzzy AdaBoost	212
	3.5 Fuzzy LogitBoost	213
4	Experimentation	214
	4.1 Methodology	215
	4.2 Results	218
5	Summary and Conclusions	227
	References	227