

Studies in Computational Intelligence, Volume 139

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. 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
*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 Radosław 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 Canicuous Abeynayake (Eds.)
Computational Intelligence Paradigms, 2008
ISBN 978-3-540-79473-8

Vol. 138. Bruno Apolloni, Witold Pedrycz, Simone Bassi
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

Jan Drugowitsch

Design and Analysis of Learning Classifier Systems

A Probabilistic Approach



Springer

Dr. Jan Drugowitsch
Department of Brain & Cognitive Sciences
Meliora Hall
University of Rochester
Rochester, NY 14627
USA
Email: jdrugowitsch@bcs.rochester.edu

ISBN 978-3-540-79865-1

e-ISBN 978-3-540-79866-8

DOI 10.1007/978-3-540-79866-8

Studies in Computational Intelligence

ISSN 1860949X

Library of Congress Control Number: 2008926082

© 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

Foreword

This book is probably best summarized as providing a principled foundation for Learning Classifier Systems. Something is happening in LCS, and particularly XCS and its variants that clearly often produces good results. Jan Drugowitsch wishes to understand this from a broader machine learning perspective and thereby perhaps to improve the systems. His approach centers on choosing a statistical definition – derived from machine learning – of “a good set of classifiers”, based on a model according to which such a set represents the data. For an illustration of this approach, he designs the model to be close to XCS, and tests it by evolving a set of classifiers using that definition as a fitness criterion, seeing if the set provides a good solution to two different function approximation problems. It appears to, meaning that in some sense his definition of “good set of classifiers” (also, in his terms, a good model structure) captures the essence, in machine learning terms, of what XCS is doing.

In the process of designing the model, the author describes its components and their training in clear detail and links it to currently used LCS, giving rise to recommendations for how those LCS can directly gain from the design of the model and its probabilistic formulation. The seeming complexity of evaluating the quality of a set of classifiers is alleviated by giving an algorithmic description of how to do it, which is carried out via a simple Pittsburgh-style LCS. A final chapter on sequential decision tasks round out the model-centered formulation that has until then focused on function approximation and classification, by providing criteria for method stability and insights into new developments.

The link provided between LCS on a theoretical level and machine learning work in general is important, especially since the latter has a more developed theory and may in part stand off from LCS because of LCS’s relative lack thereof (I stress “relative”). Also the problem addressed is important because out of greater theoretical understanding may result better classifier systems, as already demonstrated in this work by the improvements suggested for current LCS.

A particularly appealing feature of Drugowitsch’s novel approach is its universal applicability to any kind of LCS that seeks to perform function approximation, classification, or handle sequential decision tasks by means of dynamic

programming or reinforcement learning. Its close relation to XCS in this book results from the authors commitment to an LCS model structure that relates to XCS, but nothing speaks against applying the same approach to greatly different model types, resulting in different, potentially novel, LCS.

While its connection to Pittsburgh-style LCS is straightforward and clearly established in this work, using the same approach for the design of Michigan-style LCS remains a significant future challenge. Also, it will be interesting to see how the author's theoretical basis for reinforcement learning is built upon in future LCS, and how LCS designed by the model-based approach perform in comparison to currently existing LCS.

Overall, the work is elegant and approaches LCS from a refreshingly different perspective. It's also stylistically pretty novel for LCS work - but that's surely healthy!

Concord, MA, USA
March, 2008

Stewart W. Wilson

Preface

I entered the world of Learning Classifier Systems (LCS) through their introduction by Will Browne as part of a lecture series on “Advanced Artificial Intelligence” at the University of Reading, UK. Their immediate appeal as a flexible architecture that combines the power of evolutionary computation with machine learning by splitting larger problems into tractable sub-problems made me decide to pursue them further, for which I got the opportunity during my Ph.D., supervised by Alwyn Barry, at the University of Bath.

Modest dissatisfaction followed my initial euphoria when I had to discover that their theoretical basis that I planned to rest my work upon did not live up to my initial expectation. Indeed, despite being generally referred to as Genetic-based Machine Learning, their formal development had little in common with machine learning itself. Their loose definition, ad-hoc design, complex structure of interwoven sub-components, and yet surprisingly competitive performance made me comprehend why David Goldberg referred to them as “a glorious, wondrous, and inventing quagmire, but a quagmire nonetheless.”

The work presented in this book is an attempt to “clean up” on LCS and lay the foundations for a principled approach to their design by pragmatically following the road of machine learning, in order to bridge the gap between LCS and machine learning. Their design is approached from first principles, based on the question “What is a classifier system supposed to learn?”. As presented here, the work is intended for researchers in LCS, genetic-based machine learning, and machine learning, but also for anyone else who is interested in LCS. The content is in most parts based on work performed during my Ph.D., but also includes extensions to it, most notably a complete formulation for classification tasks rather than only regression tasks. The content of this book is not to be understood as the development of a new LCS, but rather as the groundwork for a new approach to their design that I and hopefully others will build upon.

Numerous people have supported me in performing this work, and I am grateful for their constant encouragement. Most notably, I would not have been able to fully focus on my work without the generous financial support of my parents, Elsbeth and Knut Drugowitsch, during my Ph.D. time. Also, my Ph.D.

supervisor, Alwyn Barry, helped me to stay focused on the main questions, and his guidance, his constructive comments, and his initiative were essential to the completion of this work. Many people in and around Bath, UK, have helped me with comments, discussions, or equally valuable moral support: Dan Richardson, Marelee Hurn, Hagen Lehmann, Tristan Caulfield, Mark Price, Jonty Needham, Joanna Bryson, and especially Will Lowe for emphasising the model behind each method. Various researchers in LCS and machine learning have offered their support through constructive discussions at conferences or per e-mail: Pier Luca Lanzi, Daniele Loiacono, Martin Butz, Stewart Wilson, Will Browne, Tim Kovacs, Gavin Brown, James Marshall, Lashon Booker, Xavier Llorà, Gavin Brown, Christopher Bishop, Markus Svensén, Matthew Beal, Tommi Jaakkola, Lei Xu, Peter Grünwald, Arta Doci, and Michael Littman. Special thanks go to Larry Bull for not giving me a too hard time at my Ph.D. viva, and for encouraging me to publish my work as a book, therefore taking full responsibility for it. Last, but certainly not least, I am deeply grateful for the moral support and patience of Odali Sanhueza throughout the years that I was working on what resulted in this book.

Rochester, NY, USA
March, 2008

Jan Drugowitsch

Contents

1	Introduction	1
1.1	Machine Learning	1
1.1.1	Common Machine Learning Tasks	2
1.1.2	Designing an Unsupervised Learning Algorithm	3
1.2	Learning Classifier Systems	5
1.2.1	A Brief Overview	5
1.2.2	Applications and Current Issues	6
1.3	About the Model-Centred Approach to LCS	7
1.3.1	The Initial Approach	7
1.3.2	Taking a Model-Centred View	7
1.3.3	Summarising the Approach	8
1.3.4	Novelties	9
1.4	How to Read This Book	9
1.4.1	Chapter Overview	9
2	Background	13
2.1	A General Problem Description	14
2.2	Early Learning Classifier Systems	16
2.2.1	Initial Idea	16
2.2.2	The General Framework	17
2.2.3	Interacting Subsystems	19
2.2.4	The Genetic Algorithm in LCS	19
2.2.5	The Problems of Early LCS	20
2.3	The LCS Renaissance	21
2.3.1	Computing the Prediction	21
2.3.2	Localisation and Representation	22
2.3.3	Classifiers as Localised Maps from Input to Output	22
2.3.4	Recovering the Global Prediction	23
2.3.5	Michigan-Style vs. Pittsburgh-Style LCS	24
2.4	Existing Theory	24
2.4.1	The Holistic View	24

2.4.2	Approaches from the Genetic Algorithm Side	25
2.4.3	Approaches from the Function Approximation Side	26
2.4.4	Approaches from the Reinforcement Learning Side	27
2.5	Discussion and Conclusion	27
3	A Learning Classifier Systems Model	29
3.1	Task Definitions	30
3.1.1	Expected Risk vs. Empirical Risk	30
3.1.2	Regression	33
3.1.3	Classification	34
3.1.4	Sequential Decision	34
3.1.5	Batch vs. Incremental Learning	35
3.2	LCS as Parametric Models	38
3.2.1	Parametric Models	39
3.2.2	An LCS Model	39
3.2.3	Classifiers as Localised Models	40
3.2.4	Recovering the Global Model	41
3.2.5	Finding a Good Model Structure	41
3.2.6	Considerations for Model Structure Search	42
3.2.7	Relation to the Initial LCS Idea	42
3.3	Summary and Outlook	43
4	A Probabilistic Model for LCS	45
4.1	The Mixtures-of-Experts Model	46
4.1.1	Likelihood for Known Gating	46
4.1.2	Parametric Gating Network	47
4.1.3	Training by Expectation-Maximisation	48
4.1.4	Localisation by Interaction	50
4.1.5	Training Issues	51
4.2	Expert Models	51
4.2.1	Experts for Linear Regression	51
4.2.2	Experts for Classification	52
4.3	Generalising the MoE Model	53
4.3.1	An Additional Layer of Forced Localisation	53
4.3.2	Updated Expectation-Maximisation Training	55
4.3.3	Implications on Localisation	55
4.3.4	Relation to Standard MoE Model	55
4.3.5	Relation to LCS	56
4.3.6	Training Issues	58
4.4	Independent Classifier Training	59
4.4.1	The Origin of Local Maxima	59
4.4.2	What Does a Classifier Model?	59
4.4.3	Introducing Independent Classifier Training	60
4.4.4	Training the Gating Network	61
4.4.5	Implications on Likelihood and Assumptions about the Data	61

4.5	A Brief Comparison to Linear LCS Models	62
4.6	Discussion and Summary	64
5	Training the Classifiers	65
5.1	Linear Classifier Models and Their Underlying Assumptions ...	66
5.1.1	Linear Models	66
5.1.2	Gaussian Noise	67
5.1.3	Maximum Likelihood and Least Squares	68
5.2	Batch Learning Approaches to Regression	68
5.2.1	The Weight Vector	69
5.2.2	The Noise Precision	69
5.3	Incremental Learning Approaches to Regression	70
5.3.1	The Principle of Orthogonality	71
5.3.2	Steepest Gradient Descent	72
5.3.3	Least Mean Squared	74
5.3.4	Normalised Least Mean Squared	76
5.3.5	Recursive Least Squares	77
5.3.6	The Kalman Filter	80
5.3.7	Incremental Noise Precision Estimation	85
5.3.8	Summarising Incremental Learning Approaches	88
5.4	Empirical Demonstration	89
5.4.1	Experimental Setup	90
5.4.2	Weight Vector Estimate	93
5.4.3	Noise Variance Estimate	93
5.5	Classification Models	94
5.5.1	A Quality Measure for Classification	95
5.5.2	Batch Approach for Classification	95
5.5.3	Incremental Learning for Classification	96
5.6	Discussion and Summary	97
6	Mixing Independently Trained Classifiers	101
6.1	Using the Generalised Softmax Function	103
6.1.1	Batch Learning by Iterative Reweighted Least Squares	103
6.1.2	Incremental Learning by Least Squares	105
6.2	Heuristic-Based Mixing Models	106
6.2.1	Properties of Weighted Averaging Mixing	107
6.2.2	Inverse Variance	110
6.2.3	Prediction Confidence	110
6.2.4	Maximum Prediction Confidence	111
6.2.5	XCS	111
6.3	Empirical Comparison	112
6.3.1	Experimental Design	113
6.3.2	Results	115
6.3.3	Discussion	118

6.4	Relation to Previous Work and Alternatives	119
6.5	Summary and Outlook	120
7	The Optimal Set of Classifiers	123
7.1	What Is Optimal?	124
7.1.1	Current LCS Approaches	124
7.1.2	Model Selection	126
7.1.3	Bayesian Model Selection	126
7.1.4	Applying Bayesian Model Selection to Finding the Best Set of Classifiers	128
7.1.5	The Model Structure Prior $p(\mathcal{M})$	128
7.1.6	The Myth of No Prior Assumptions	129
7.2	A Fully Bayesian LCS for Regression	130
7.2.1	Data, Model Structure, and Likelihood	131
7.2.2	Multivariate Regression Classifiers	133
7.2.3	Priors on the Classifier Model Parameters	133
7.2.4	Mixing by the Generalised Softmax Function	135
7.2.5	Priors on the Mixing Model	135
7.2.6	Joint Distribution over Random Variables	136
7.3	Evaluating the Model Evidence	136
7.3.1	Variational Bayesian Inference	137
7.3.2	Classifier Model $q_{W,\tau}^*(W, \tau)$	138
7.3.3	Classifier Weight Priors $q_\alpha^*(\alpha)$	141
7.3.4	Mixing Model $q_V^*(V)$	142
7.3.5	Mixing Weight Priors $q_\beta^*(\beta)$	144
7.3.6	Latent Variables $q_Z^*(Z)$	144
7.3.7	Required Moments of the Variational Posterior	146
7.3.8	The Variational Bound $\mathcal{L}(q)$	148
7.3.9	Independent Classifier Training	152
7.3.10	How to Get $p(\mathcal{M} \mathcal{D})$ for Some \mathcal{M}	153
7.4	Predictive Distribution	154
7.4.1	Deriving $p(y' x', \mathcal{D})$	154
7.4.2	Mean and Variance	156
7.5	Model Modifications to Perform Classification	156
7.5.1	Local Classification Models and Their Priors	157
7.5.2	Variational Posteriors and Moments	158
7.5.3	Variational Bound	158
7.5.4	Independent Classifier Training	159
7.5.5	Predictive Density	159
7.6	Alternative Model Selection Methods	160
7.6.1	Minimum Description Length	160
7.6.2	Structural Risk Minimisation	161
7.6.3	Bayesian Ying-Yang	162
7.6.4	Training Data-Based Approaches	162
7.7	Discussion and Summary	162

8	An Algorithmic Description	165
8.1	Computing $p(\mathcal{M} \mathcal{D})$	166
8.1.1	Model Probability and Evidence	167
8.1.2	Training the Classifiers	168
8.1.3	Training the Mixing Model	170
8.1.4	The Variational Bound	175
8.1.5	Scaling Issues	177
8.2	Two Alternatives for Model Structure Search	178
8.2.1	Model Structure Search by a Genetic Algorithm	179
8.2.2	Model Structure Search by Markov Chain Monte Carlo	181
8.2.3	Building Blocks in Classifier Sets	183
8.3	Empirical Demonstration	184
8.3.1	Representations	184
8.3.2	Generated Function	187
8.3.3	Sparse, Noisy Data	190
8.3.4	Function with Variable Noise	192
8.3.5	A Slightly More Complex Function	193
8.4	Improving Model Structure Search	196
8.4.1	Using More Information	196
8.4.2	Incremental Implementations	198
8.5	Summary	200
9	Towards Reinforcement Learning with LCS	203
9.1	Problem Definition	205
9.1.1	Markov Decision Processes	205
9.1.2	The Value Function, the Action-Value Function and Bellman's Equation	206
9.1.3	Problem Types	208
9.1.4	Matrix Notation	208
9.2	Dynamic Programming and Reinforcement Learning	208
9.2.1	Dynamic Programming Operators	208
9.2.2	Value Iteration and Policy Iteration	209
9.2.3	Approximate Dynamic Programming	210
9.2.4	Temporal-Difference Learning	210
9.2.5	SARSA(λ)	211
9.2.6	Q-Learning	212
9.2.7	Approximate Reinforcement Learning	213
9.3	Reinforcement Learning with LCS	213
9.3.1	Approximating the Value Function	214
9.3.2	Bellman's Equation in the LCS Context	215
9.3.3	Asynchronous Value Iteration with LCS	216
9.3.4	Q-Learning by Least Mean Squares	217
9.3.5	Q-Learning by Recursive Least Squares	218
9.3.6	XCS with Gradient Descent	219

9.4	Stability of RL with LCS	220
9.4.1	Stability of Approximate Dynamic Programming	220
9.4.2	Stability on the Structure and the Parameter Learning Level	221
9.4.3	Non-expansion with Respect to $\ \cdot\ _\infty$	223
9.4.4	Non-expansion with Respect to $\ \cdot\ _D$	225
9.4.5	Consequences for XCS and XCSF	227
9.5	Further Issues	227
9.5.1	Long Path Learning	227
9.5.2	Exploration and Exploitation	231
9.6	Summary	234
10	Concluding Remarks	237
A	Notation	241
B	XCS and XCSF	247
B.1	Classifier Model and Mixing Model	247
B.2	Model Structure Search	248
	References	251
	Index	265