

Lecture Notes in Statistics

112

Edited by P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg,
I. Olkin, N. Wermuth, S. Zeger

Springer

New York

Berlin

Heidelberg

Barcelona

Budapest

Hong Kong

London

Milan

Paris

Santa Clara

Singapore

Tokyo

Doug Fisher
Hans-J. Lenz
(Editors)

Learning from Data

Artificial Intelligence and Statistics V



Springer

Doug Fisher
Vanderbilt University
Department of Computer Science
Box 1679, Station B
Nashville, Tennessee 37235

Hans-J. Lenz
Free University of Berlin
Department of Economics
Institute of Statistics and
Econometrics
14185 Berlin, Garystre 21
Germany

CIP data available.

Printed on acid-free paper.

© 1996 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Camera ready copy provided by the author.

9 8 7 6 5 4 3 2 1

ISBN-13:978-0-387-94736-5 e-ISBN-13:978-1-4612-2404-4
DOI: 10.1007/978-1-4612-2404-4

Preface

Ten years ago Bill Gale of AT&T Bell Laboratories was primary organizer of the first *Workshop on Artificial Intelligence and Statistics*. In the early days of the Workshop series it seemed clear that researchers in AI and statistics had common interests, though with different emphases, goals, and vocabularies. In learning and model selection, for example, a historical goal of AI to build autonomous agents probably contributed to a focus on parameter-free learning systems, which relied little on an external analyst's assumptions about the data. This seemed at odds with statistical strategy, which stemmed from a view that model selection methods were tools to augment, not replace, the abilities of a human analyst. Thus, statisticians have traditionally spent considerably more time exploiting prior information of the environment to model data and exploratory data analysis methods tailored to their assumptions. In statistics, special emphasis is placed on model checking, making extensive use of residual analysis, because all models are 'wrong', but some are better than others. It is increasingly recognized that AI researchers and/or AI programs can exploit the same kind of statistical strategies to good effect.

Often AI researchers and statisticians emphasized different aspects of what in retrospect we might now regard as the same overriding tasks. For example, those in AI exploited computer processing to a great extent and explored a wide variety of search control strategies, while many in statistics focused on principled measures for evaluating the merits of varying models and states, often in computationally-limited circumstances. There were numerous other differences in the traditional research biases of AI and statistics, such as interests in deductive versus inductive inference. These differences in goals, methodology, and vocabulary fostered a healthy tension that fueled debate, and contributed to excitement and exchange of ideas among participants of the early AI and Statistics Workshops.

This volume contains a subset of papers presented at the *Fifth International Workshop on Artificial Intelligence and Statistics*, which was held at Ft. Lauderdale, Florida in January 1995. The primary theme of the Fifth Workshop was 'Learning from Data', but a variety of areas at the interface of AI and statistics were represented at the Workshop and in this volume. The Fifth workshop saw some of the same kind of tension and excitement that characterized earlier workshops, with one participant confiding that this was the first time that he/she had felt 'intimidated' in presenting a paper. However, there has also been noticeable growth, with researchers from AI and statistics becoming aware of and exploiting research in the other field. In some cases, such as Causality and Graphical Models, the papers of this volume suggest that a dichotomy between AI and statistics is artificial. In cases such as Learning, researchers might still be reasonably described as falling into the statistics or AI camps, but there is heightened awareness of commonalities across the disciplines.

As we have noted, a variety of research areas are represented in this volume. This variety contributed to a minor dilemma for the editors, since many contributions fell into more than one subject category. In some cases, such as Natural Language Processing (NLP), a categorization was fairly easy. While all of the NLP papers could be viewed as learning from data, the application of statistical methods in NLP is currently attracting the enthusiastic attention of researchers in both AI and statistics – a trend that we want to

highlight. In other cases, we have compromised in our classification, but the process of classifying the contributions, to say nothing of the paper selection process, led to some insights about the difference in meaning and emphases that probably remain in AI and statistics. We suspect, for example, that those in AI probably ascribe a more general meaning to the term ‘Exploratory Data Analysis’.

We were both pleased to have helped organize the Fifth Workshop on Artificial Intelligence and Statistics; the excitement that we felt from the exchange of ideas made it worth the effort. We also extend our thanks to a several groups. Foremost, we thank the authors – those represented in this volume and the Workshop proceedings of preliminary papers. The program committee, which reviewed initial submissions to the Workshop, contributed mightily to the quality and vibrancy of the Workshop. We gratefully acknowledge the sponsorship of the *Society for Artificial Intelligence and Statistics* for covering initial costs, and the *International Association for Statistical Computing* for promulgating information about the Workshop. Finally, on this, the ten-year anniversary of the AI and Statistics Workshop series, we thank Bill Gale and his colleagues for their foresight in organizing the first workshop.

Doug Fisher, Vanderbilt University, USA
Hans-J. Lenz, Free University of Berlin, Germany

Previous workshop volumes:

Gale, W. A., ed. (1986) *Artificial Intelligence and Statistics*, Reading, MA: Addison-Wesley.

Hand, D. J., ed. (1990) *Artificial Intelligence and Statistics II. Annals of Mathematics and Artificial Intelligence*, 2.

Hand, D. J., ed. (1993) *Artificial Intelligence Frontiers in Statistics: AI and Statistics III*, London, UK: Chapman & Hall.

Cheeseman, P., & Oldford, R. W., eds. (1994) *Selecting Models from Data: Artificial Intelligence and Statistics IV*, New York, NY: Springer Verlag.

1995 Organizing and Program Committees:

General Chair:	D. Fisher	Vanderbilt U., USA
Program Chair:	H.-J. Lenz	Free U., Berlin, Germany

Program Committee Members:

W. Buntine	NASA (Ames), USA
J. Catlett	AT&T Bell Labs, USA
P. Cheeseman	NASA (Ames), USA
P. Cohen	U. of Mass., USA
D. Draper	U. of Bath, UK
Wm. Dumouchel	Columbia U., USA
A. Gammerman	U. of London, UK
D. J. Hand	Open U., UK
P. Hietala	U. Tampere, Finland
R. Kruse	TU Braunschweig, Germany
S. Lauritzen	Aalborg U., Denmark
W. Oldford	U. of Waterloo, Canada
J. Pearl	UCLA, USA
D. Pregibon	AT&T Bell Labs, USA
E. Roedel	Humboldt U., Germany
G. Shafer	Rutgers U., USA
P. Smyth	JPL, USA

Tutorial Chair:	P. Shenoy	U. Kansas, USA
Local Arrangements:	D. Pregibon	AT&T Bell Labs, USA

Sponsors: Society for Artificial Intelligence and Statistics
International Association for Statistical Computing

Contents

Preface	v
I Causality	1
1 Two Algorithms for Inducing Structural Equation Models from Data <i>Paul. R. Cohen, Dawn E. Gregory, Lisa Ballesteros and Robert St. Amant</i>	3
2 Using Causal Knowledge to Learn More Useful Decision Rules from Data <i>Louis Anthony Cox, Jr.</i>	13
3 A Causal Calculus for Statistical Research <i>Judea Pearl</i>	23
4 Likelihood-based Causal Inference <i>Qing Yao and David Tritchler</i>	35
II Inference and Decision Making	45
5 Ploxoma: Testbed for Uncertain Inference <i>H. Blau</i>	47
6 Solving Influence Diagrams Using Gibbs Sampling <i>Ali Jenzarli</i>	59
7 Modeling and Monitoring Dynamic Systems by Chain Graphs <i>Alberto Lekuona, Beatriz Lacruz and Pilar Lasala</i>	69
8 Propagation of Gaussian Belief Functions <i>Liping Liu</i>	79
9 On Test Selection Strategies for Belief Networks <i>David Madigan and Russell G. Almond</i>	89
10 Representing and Solving Asymmetric Decision Problems Using Valuation Networks <i>Prakash P. Shenoy</i>	99
11 A Hill-Climbing Approach for Optimizing Classification Trees <i>Xiaorong Sun, Steve Y. Chiu and Louis Anthony Cox</i>	109

III Search Control in Model Hunting	119
12 Learning Bayesian Networks is NP-Complete <i>David Maxwell Chickering</i>	121
13 Heuristic Search for Model Structure: The Benefits of Restraining Greed <i>John F. Elder IV</i>	131
14 Learning Possibilistic Networks from Data <i>Jörg Gebhardt and Rudolf Kruse</i>	143
15 Detecting Imperfect Patterns in Event Streams Using Local Search <i>Adele E. Howe</i>	155
16 Structure Learning of Bayesian Networks by Hybrid Genetic Algorithms <i>Pedro Larrañaga, Roberto Murga, Mikel Poza and Cindy Kuijpers</i>	165
17 An Axiomatization of Loglinear Models with an Application to the Model-Search Problem <i>Francesco M. Malvestuto</i>	175
18 Detecting Complex Dependencies in Categorical Data <i>Tim Oates, Matthew D. Schmill, Dawn E. Gregory and Paul R. Cohen</i>	185
IV Classification	197
19 A Comparative Evaluation of Sequential Feature Selection Algorithms <i>David W. Aha and Richard L. Bankert</i>	199
20 Classification Using Bayes Averaging of Multiple, Relational Rule-Based Models <i>Kamal Ali and Michael Pazzani</i>	207
21 Picking the Best Expert from a Sequence <i>Ruth Bergman and Ronald L. Rivest</i>	219
22 Hierarchical Clustering of Composite Objects with a Variable Number of Components <i>A. Ketterlin, P. Gañçarski and J. J. Korczak</i>	229

23	Searching for Dependencies in Bayesian Classifiers	239
	<i>Michael J. Pazzani</i>	
V	General Learning Issues	249
24	Statistical Analysis fo Complex Systems in Biomedicine	251
	<i>Harry B. Burke</i>	
25	Learning in Hybrid Noise Environments Using Statistical Queries	259
	<i>Scott E. Decatur</i>	
26	On the Statistical Comparison of Inductive Learning Methods	271
	<i>A. Feelders and W. Verkooijen</i>	
27	Dynamical Selection of Learning Algorithms	281
	<i>Christopher J. Merz</i>	
28	Learning Bayesian Networks Using Feature Selection	291
	<i>Gregory M. Provan and Moninder Singh</i>	
29	Data Representations in Learning	301
	<i>Geetha Srikantan and Sargur N. Srihari</i>	
VI	EDA: Tools and Methods	311
30	Rule Induction as Exploratory Data Analysis	313
	<i>Jason Catlett</i>	
31	Non-Linear Dimensionality Reduction: A Comparative Performance Analysis	323
	<i>Olivier de Vel, Sofianto Li and Danny Coomans</i>	
32	Omega-Stat: An Environment for Implementing Intelligent Modeling Strategies	333
	<i>E. James Harner and Hanga C. Galfalvy</i>	
33	Framework for a Generic Knowledge Discovery Toolkit	343
	<i>Pat Riddle, Roman Fresnedo and David Newman</i>	
34	Control Representation in an EDA Assistant	353
	<i>Robert St. Amant and Paul R. Cohen</i>	

VII	Decision and Regression Tree Induction	363
35	A Further Comparison of Simplification Methods for Decision-Tree Induction <i>Donato Malerba, Floriana Esposito and Giovanni Semeraro</i>	365
36	Robust Linear Discriminant Trees <i>George H. John</i>	375
37	Tree Structured Interpretable Regression <i>David Lubinsky</i>	387
38	An Exact Probability Metric for Decision Tree Splitting <i>J. Kent Martin</i>	399
VIII	Natural Language Processing	411
39	Two Applications of Statistical Modelling to Natural Language Processing <i>William DuMouchel, Carol Friedman, George Hripcsak, Stephen B. Johnson and Paul D. Clayton</i>	413
40	A Model for Part-of-Speech Prediction <i>Alexander Franz</i>	423
41	Viewpoint-Based Measurement of Semantic Similarity Between Words <i>Kaname Kasahara, Kazumitsu Matsuzawa, Tsutomu Ishikawa and Tsukasa Kawaoka</i>	433
42	Part-of-Speech Tagging from “Small” Data Sets <i>Eric Neufeld and Greg Adams</i>	443