Preliminary Papers

of the

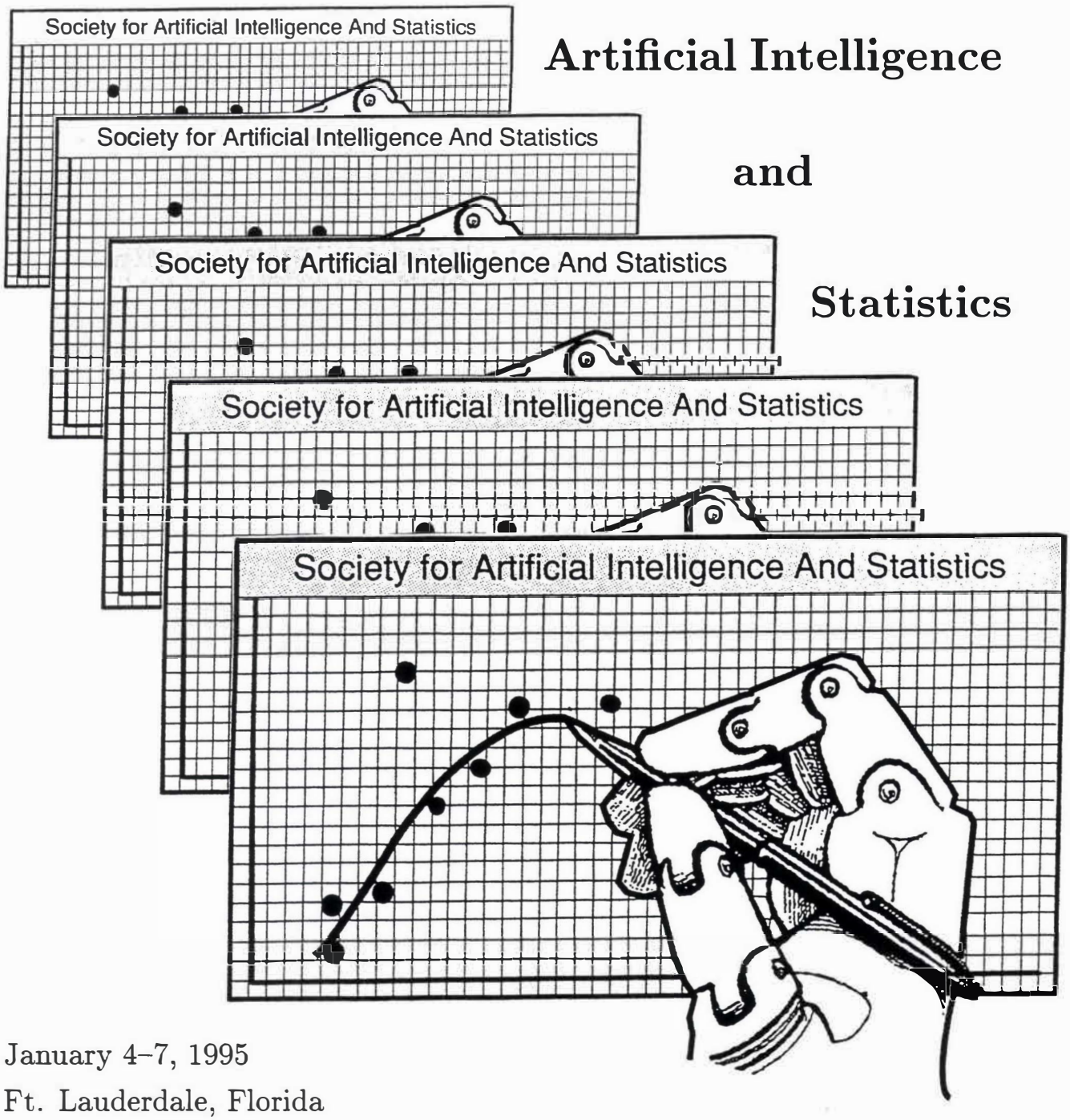
Fifth International Workshop

on

Artificial Intelligence

and

Statistics



January 4–7, 1995

Ft. Lauderdale, Florida

CONTENTS

PREFACE

Almost 10 years ago, in April 1985, Bill Gale and Daryl Pregibon organized the first *Workshop on Artificial Intelligence and Statistics* at Princeton, New Jersey. In his Preface to the edited volume Bill motivated the Workshop by noting that research in each field could benefit the other. The 15 papers presented at the first Workshop reflected a variety of interdisciplinary themes: task analysis of and expert advisory systems for statistical strategy, reasoning under uncertainty, supervised and unsupervised learning, heuristic search in pattern recognition, statistical tools for knowledge acquisition, and characterizing the behavior of 'intelligent' systems through statistical analysis. Papers in subsequent workshops have further developed some of these themes and/or branched into other areas at the interface of AI and statistics.

It is difficult to know the influence of the AI and Statistics Workshops on research in the two communities. Suffice it to say that I have used these Workshops to guide, gauge, and 'correct' my own work. I have heard several volunteer it as their favorite forum for discussion and debate, and in any case, it seems generally acknowledged as a workshop of very high quality. Finally, in contrast to 15 papers presented at the first Workshop in 1985, there are 62 papers, selected from a total of 111 submissions, being presented at the *Fifth International Workshop on AI and Statistics* in 1995. Apparently, many concur with Bill's motivation for the original Workshop.

There are many people that helped bring this Workshop to fruition. Hans Lenz, the Program Chair, was responsible for selecting the program committee, assigning papers, collecting reviews, and deciding the technical program. He also iterated with me in scheduling the presentations. Despite occasional logistical problems in the Nashville–Berlin link, it was a pleasure working with Hans. Hans and I thank those that provided submissions for review – unfortunately, not all quality submissions could be accommodated. We thank the program committee for reviews and advise on paper selection. Daryl Pregibon handled finances, local arrangements, and provided input on scheduling – if there was a General CoChair position, then the title would belong to Daryl. The Tutorial Chair, Prakash Shenoy, put together a wonderful tutorial program. Our tutorial presenters, David Aha, Trevor Hastie, Steffen Lauritzen, and Glenn Shafer deserve special thanks – the time investment necessary for a good tutorial is very high, and we appreciate their efforts. My wife, Pat Fisher, and Jing Lin, Julio Ortega, and Doug Talbert of Vanderbilt University helped put together the proceedings. Robyn Landers of the University of Waterloo has been of great service by maintaining the AI and Statistics electronic mailing list, over which considerable Workshop information has been promulgated.

Finally, I greatly appreciate the sage advice of past organizers, particularly their input on program committee selection, but on a variety of other issues as well. They include Peter Cheeseman, Bill DuMouchel, David Hand, Wayne Oldford, and Daryl Pregibon. I am optimistic that this year's meeting will follow in the tradition of high quality that they helped hone.

Doug Fisher, General Chair

ORGANIZATION

General Chair:	D. Fisher	Vanderbilt U., USA
Program Chair:	H. 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

TUTORIALS

Wednesday, January 4

Machine Learning ..... 9:00 AM to 12:15 PM
Dr. David Aha, Naval Research Lab

Statistical Methods for Inducing Models from Data ..... 9:00 AM to 12:15 PM
Prof. Steffen Lauritzen, Aalborg University

Probabilistic Models of Causality ..... 2:00 PM to 5:15 PM
Prof. Glenn Shafer, Rutgers University

Statistical Models for Function Estimation and Classification .... 2:00 PM to 5:15 PM
Prof. Trevor Hastie, Stanford University

SCHEDULE

Thursday, January 5

8:45	Opening Comments
9:00	A causal calculus for statistical research, with applications to observational and experimental studies – *J. Pearl* 430
9:30	Causal discovery from data in the presence of selection bias –*G. F. Cooper* ... 140
10:00	Likelihood-based causal inference – *Q. Yao, D. Tritchler* 520
10:30	Break
11:00	Missing data models as meta-data – *R. G. Almond, J. Schimert* 23
11:30	Learning in hybrid noise environments using statistical queries – *S. E. Decatur* 175
12:00	Lunch
1:30	Heuristic search for model structure – *J. F. Elder* 199
2:00	Tree structured interpretable regression – *D. Lubinsky* 331
2:30	Break
3:00	Poster Summaries
5:00	Break
6:00	Dinner
6:30	Poster Session 1 (ends at 8:10)
8:20	Poster Session 2 (ends at 10:00)

Friday, January 6

8:30	Learning Bayesian networks: Search methods and experimental results – *D. M. Chickering, D. Geiger, D. Heckerman* 112
9:00	Derivation DAGs for inferring interaction models – *F. M. Malvestuto* 354
9:30	Learning possibilistic networks from data – *J. Gebhardt, R. Kruse* 233
10:00	Break
10:30	Software for data analysis with graphical models – *W. L. Buntine, H. S. Roy* .. 76
11:00	Preliminary system design for an EDA assistant – *R. St Amant, P. R. Cohen* . 502
11:30	Framework for a generic knowledge discovery toolkit – *P. Riddle, R. Fresnedo, D. Newman* 457
12:00	Lunch
1:30	A generalization of the Tetrad representation theorem – *G. Shafer, A. Kogan, P. Spirtes* 476
2:00	Two algorithms for inducing structural equation models from data – *P. R. Cohen, D. E. Gregory, L. Ballesteros, R. St Amant* 129
2:30	Break
3:00	Panel/Debate (topic TBA)
8:00	Informal Discussions (on topics TBA)

Saturday, January 7

9:00	On test selection strategies for belief networks – *D. Madigan, R. G. Almond*	342
9:30	Using causal knowledge to learn more useful decision rules from data – *L. A. Cox, Jr.*	151
10:00	Dynamic learning bias selection – *C. J. Merz*	386
10:30	Break	
11:00	Viewpoint-based measurement of semantic similarity between words – *K. Kasahara, K. Matsuzawa, T. Ishikawa, T. Kawaoka*	292
11:30	Discovering morphemic suffixes: A case study in MDL induction – *M. R. Brent, S. K. Murthy, A. Lundberg*	64
12:00	Closing Comments	
12:15	Break	
12:45	Business Meeting	

Poster Session 1
Thursday, January 5

A comparative evaluation of sequential feature selection algorithms – *D. W. Aha, R. L. Bankert*

Hypergraph grammars for knowledge-based model construction – *R. G. Almond*

Tools for empirically analyzing AI programs – *S. D. Anderson, D. M. Hart, D. L. Westbrook, P. R. Cohen, A. Carlson*

Decision-theoretic case-based reasoning – *J. S. Breese, D. Heckerman*

Tailoring rulesets to misclassification costs – *J. Catlett*

Analysis and Application of the Generalized Mean-Shift Process – *Y. Cheng*

A further study of pruning methods in decision tree induction – *F. Esposito, D. Malerba, G. Semeraro*

Classifying new words for robust parsing – *A. Franz*

Omega-Stat: An environment for implementing intelligent modeling strategies – *E. J. Harner, H. C. Galfalvy*

A decision-based view of causality – *D. Heckerman, R. Shachter*

Solving influence diagrams using Gibbs sampling – *A. Jenzarli*

Hierarchical clustering of composite objects with a variable number of components – *A. Ketterlin, P. Gançarski, J. J. Korczak*

Structure learning of Bayesian networks by hybrid genetic algorithms – *P. Larrañaga, R. H. Murga, M. Poza, C. M. H. Kuijpers*

An exact probability metric for decision tree splitting and stopping – *J. K. Martin*

Modeling life time data by neural networks – *Y. B. Moon, H-J. Kim*

Statistical preprocessing for decision tree induction – *S. K. Murthy*

Detecting complex dependencies in categorical data – *T. Oates, D. Gregory, P. R. Cohen*

Learning Bayesian networks using feature selection – *G. M. Provan, M. Singh*

Combining statistics and AI in the optimization of semiconductors for solar cells – *J. Risius, G. Seidelmann*

Data representations in learning – *G. Srikantan, S. N. Srihari*

Learning from data by guiding the analyst: On the representation, use, and creation of visual statistical strategies – *F. W. Young, D. J. Lubinsky*

Poster Session 2
Thursday, January 5

Learning multiple relational rule-based models – *K. Ali, C. Brunk, M. Pazzani*

Picking the best expert from a sequence – *R. Bergman, R. L. Rivest*

Ploxoma: Testbed for uncertain inference – *H. Blau*

Comparing the prediction accuracy of statistical models and artificial neural networks in breast cancer – *H. B. Burke, D. B. Rosen, P. H. Goodman*

Predicting stock returns with genetic programming: Do the short-term nonlinear regularities exist? – *S-H. Chen, C-H. Yeh*

Truncated Gaussians as tolerance sets – *F. Cozman, E. Krotkov*

Textual data mining – *S. J. Cunningham*

Non-Linear dimensionality reduction: A comparative performance analysis – *O. de Vel, S. Li, D. Coomans*

Two applications of statistical modelling to natural language processing – *W. DuMouchel, C. Friedman, G. Hripcsak, S. B. Johnson, P. D. Clayton*

Which method learns most from the data? Methodological issues in the analysis of comparative studies – *A. Feelders, W. Verkooijen*

Abductive reasoning in Bayesian belief networks using a genetic algorithm – *E. S. Gelsema*

Finding dependencies in event streams using local search – *A. E. Howe*

Robust linear discriminant trees – *G. H. John*

On graphical models for dynamic systems – *A. Lekuona, B. Lacruz, P. Lasala*

Propagation of Gaussian belief functions – *L. Liu*

MENTOR: A Bayesian model for prediction and intervention in mental retardation – *S. Mani, M. Valtorta, S. McDermott*

Evaluating and comparing classifiers: Complexity measures – *J. K. Martin*

Part-of-speech tagging from small data sets – *E. Neufeld, G. Adams, H. Choy, R. Orthner, T. Philip, A. Tawfik*

Searching for attribute dependencies in Bayesian classifiers – *M. Pazzani*

Representing and solving asymmetric decision problems using valuation networks – *P. Shenoy*

A hill-climbing approach to construct near optimal decision trees – *X. Sun, Y. Qiu, L. A. Cox, Jr.*