

BIENNIAL REPORT

2004 – 2005

Laboratory of Computer and Information Science

Neural Networks Research Centre

Helsinki University of Technology

P.O. Box 5400

FIN-02015 HUT, Finland

V. Könönen, K. Raivio, R. Vigário, and L. Koivisto, editors

Otaniemi, April 2006

ISSN 1795-5092
ISSN 1796-4121 (electronic version)

Yliopistopaino
Helsinki 2006

Contents

Preface	7
Personnel	9
Awards and activities	13
Courses	27
Doctoral dissertations	33
Theses	45
<i>I Neural Networks Research Centre: Research Projects</i>	
1 Introduction	51
2 Independent component analysis and blind source separation	55
<i>Erkki Oja, Juha Karhunen, Harri Valpola, Jaakko Särelä, Mika Inki, Antti Honkela, Alexander Ilin, Karthikesh Raju, Tapani Ristaniemi, Ella Bingham</i>	
2.1 Introduction	56
2.2 Finite sample behaviour of the FastICA algorithm	57
2.3 Nonlinear ICA and BSS	59
2.4 Denoising source separation	61
2.5 Climate data analysis with DSS	63
2.6 ICA and denoising source separation in CDMA communications	66
2.7 ICA for image representations	68
2.8 Analyzing 0-1 data	69
3 Neuroinformatics	71
<i>Ricardo Vigário, Jaakko Särelä, Sergey Borisov, Jarkko Ylipaavalniemi, Antti Honkela, Alexander Ilin, Elina Karp, Erkki Oja</i>	
3.1 Setup of the group	72
3.2 Source localization of low- and high-amplitude alpha activity	74
3.3 DSS extraction of the cardiac subspace in MEG	76
4 Variational Bayesian learning of generative models	79
<i>Juha Karhunen, Antti Honkela, Alexander Ilin, Tapani Raiko, Markus Harva, Harri Valpola, Erkki Oja</i>	
4.1 Bayesian modeling and variational learning: introduction	80
4.2 Theoretical improvements	82

4.3	Building blocks for variational Bayesian learning	85
4.4	Nonlinear and non-negative blind source separation	87
4.5	Dynamic modelling using nonlinear state-space models	89
4.6	Relational models	92
4.7	Applications to astronomy	93
5	Bioinformatics	97
	<i>Samuel Kaski, Janne Nikkilä, Merja Oja, Leo Lahti, Jarkko Venna, Eerika Savia, Arto Klami</i>	
5.1	Introduction	98
5.2	Yeast systems biology	99
5.3	Comparative functional genomics	102
5.4	Gene expression atlas	103
5.5	Genomics of human endogenous retroviruses	104
6	Dependency exploration and learning metrics	105
	<i>Samuel Kaski, Jaakko Peltonen, Janne Nikkilä, Jarkko Venna, Jarkko Salojärvi, Arto Klami, Leo Lahti, Janne Sinkkonen</i>	
6.1	Introduction	106
6.2	Supervised unsupervised learning	107
6.3	Dependency exploration	112
6.4	Discriminative learning	115
6.5	Visualization methods	117
7	Image retrieval and analysis	119
	<i>Erkki Oja, Jorma Laaksonen, Jukka Iivarinen, Markus Koskela, Ramūnas Girdziušas, Jussi Pakkanen, Ville Viitaniemi, Zhirong Yang, Rami Rautkorpi, Mats Sjöberg, Hannes Muurinen</i>	
7.1	Content-based image retrieval by self-organizing maps	120
7.2	Content-based retrieval of defect images	127
7.3	Alterable volume flow in the use of input deformations for a massive Gaussian process smoothing	130
7.4	Stopping criteria for nonlinear diffusion filters	131
8	Adaptive cognitive systems	133
	<i>Timo Honkela, Krista Lagus, Ville Könönen, Ann Russell, Mikaela Klami, Tiina Lindh-Knuutila, Matti Pöllä, Jaakko Väyrynen, Kevin Hynnä</i>	
8.1	Introduction	134
8.2	Emergence of cognitive and conceptual representations	135
8.3	Reinforcement learning in multiagent systems	139
9	Speech recognition	143
	<i>Mikko Kurimo, Panu Somervuo, Kalle Palomäki, Vesa Siivola, Teemu Hirsimäki, Janne Pylkkönen, Simo Broman, Ville Turunen, Sami Virpioja</i>	
9.1	The speech recognition tasks and systems	144
9.2	Acoustic modeling	145
9.3	Language modeling	148
9.4	Large vocabulary decoder	151
9.5	Spoken document retrieval	152

10 Natural language processing	155
<i>Krista Lagus, Timo Honkela, Mathias Creutz, Mikko Kurimo, Sami Virpioja, Jaakko Väyrynen, Oskar Kohonen, and Krister Lindén</i>	
10.1 Unsupervised segmentation of words into morphs	156
10.2 Word sense disambiguation using document maps	160
10.3 Topically focusing language model	162
10.4 Emergence of linguistic features using independent component analysis . . .	165
10.5 SOM-based analysis of words and sentences	167
11 Intelligent data engineering	169
<i>Olli Simula, Jaakko Hollmen, Sampsa Laine, Kimmo Raivio, Miki Sirola, Pasi Lehtimäki, Timo Similä, Mika Sulkava, Jarkko Tikka, Jukka Parviainen, Teppo Marin, Golan Lampi, Mikko Multanen, Risto Hakala, Petri Saarikko</i>	
11.1 A knowledge-based model for analysis of GSM network performance	170
11.2 Analysis of mobile radio access network	172
11.3 Using visualization, variable selection and feature extraction to learn from industrial data	173
11.4 SOM in decision support	175
11.5 Interpreting dependencies in data using the Self-Organizing Map	176
11.6 Analysis of forest nutrition data	178
11.7 Parsimonious signal representations in data analysis	180
12 Time series prediction	183
<i>Amaury Lendasse, Yongnan Ji, Nima Reyhani, Jin Hao, Antti Sorjamaa</i>	
12.1 Introduction	184
12.2 The CATS benchmark	185
12.3 Methodology for long-term prediction of time series	186
12.4 Input selection strategies	187
12.5 Chemometry	188
13 Proactive information retrieval	191
<i>Samuel Kaski, Jarkko Salojärvi, Eerika Savia, and Kai Puolamäki</i>	
13.1 Introduction	192
13.2 Implicit relevance feedback from eye movements	193
13.3 Collaborative Filtering: Inferring user interests from other available sources	194
13.4 Application: Proactive information retrieval	195
14 Other projects	197
14.1 Adaptive committee techniques	198
14.2 Data analysis using the Evolving Tree	199
14.3 Independent variable group analysis	200
14.4 Worldwide research on and using the Self-Organizing Map	202
14.5 Self-Organizing Neural Projections	204
14.6 Applications of the Self-Organizing Map	206
Publications of the Neural Networks Research Centre	207
II From Data to Knowledge Research Unit: Research Projects under the CIS Laboratory	

15 From Data to Knowledge Research Unit	227
<i>Heikki Mannila, Jaakko Holm�en, Kai Puolam�aki, Ella Bingham, Johan Himberg, Robert Gwadera, Hannes Heikinheimo, Antti Ukkonen, Jouni K. Sepp�anen, Nikolaj Tatti, Heli Hiisil�, Antti Rasinen, Mikko Kor- pela, Janne Toivola</i>	
15.1 Data mining at the Pattern Discovery group	228
15.2 Discovering orderings	229
15.3 Theoretical aspects of data mining	231
15.4 Extending frequent itemsets: dense itemsets and tiles	233
15.5 Basis segmentation	235
15.6 Data mining in bioinformatics	236
Publications of the From Data to Knowledge Research Unit	239