BIENNIAL REPORT

2002 – 2003

Laboratory of Computer and Information Science

Neural Networks Research Centre

Helsinki University of Technology

P.O. Box 5400

FI-02015 HUT, Finland

K. Puolamäki and L. Koivisto, editors

Otaniemi, February 2004

ISBN 951-22-6973-2
ISBN 951-22-6974-0 (electronic version)

Yliopistopaino
Helsinki 2004

Contents

Preface **7**

Personnel **9**

Awards and activities **13**

Courses **25**

Doctoral dissertations **31**

Theses **43**

I Neural Networks Research Centre: Research Projects

1 Introduction **49**

2 Independent component analysis and blind source separation **53**

Erkki Oja, Juha Karhunen, Ella Bingham, Maria Funaro, Johan Himberg, Antti Honkela, Aapo Hyvärinen, Alexander Ilin, Karthikesh Raju, Tapani Ristaniemi, Jaakko Särelä, Harri Valpola, Ricardo Vigário . . . .

2.1 Introduction . . . 54
2.2 Theoretical advances . . . 55
2.3 Comparison studies on blind separation of post-nonlinear mixtures . . . 58
2.4 Text mining . . . 60
2.5 ICA for astronomical data . . . 61
2.6 ICA in CDMA communications . . . 62
2.7 Explorative investigation of the reliability of independent component estimates . . . 65
2.8 The European joint project BLISS . . . 67

3 Variational Bayesian learning of generative models **69**

Harri Valpola, Antti Honkela, Alexander Ilin, Tapani Raiko, Markus Harva, Tomas Östman, Juha Karhunen, Erkki Oja . . . .

3.1 Bayesian modeling and variational learning . . . 70
3.2 Theoretical improvements . . . 73
3.3 Building blocks for variational Bayesian learning . . . 75
3.4 Nonlinear static and dynamic blind source separation . . . 76
3.5 Hierarchical modeling of variances . . . 80
3.6 Applications . . . 82

4 Computational neuroscience **87**
Aapo Hyvärinen, Patrik Hoyer, Jarmo Hurri, Mika Inki . . . . . .
4.1 The statistical structure of natural images and visual representation 88

5 Analysis of independent components in biomedical signals **93**
Ricardo Vigário, Jaakko Särelä, Elina Karp, Jarkko Ylipaavalniemi . . . . .
5.1 Biomedical data analysis 94

6 Image analysis applications **99**
Erkki Oja, Jorma Laaksonen, Jukka Iivarinen, Markus Koskela, Ramūnas Girdziušas, Jussi Pakkanen, Ville Viitaniemi, Mika Rummukainen, Mats Sjöberg . . . . . .
6.1 Content-based image retrieval by self-organizing maps 100
6.2 Content-based retrieval of defect images 105
6.3 Extended fluid-based image registration 108

7 On-line recognition of handwritten characters **109**
Vuokko Vuori, Matti Aksela, Jorma Laaksonen, Erkki Oja . . . . . .
7.1 Introduction 110
7.2 Adaptive prototype-based character classifiers 111
7.3 Adaptive committee techniques 112

8 Self-organizing map **113**
Teuvo Kohonen, Samuel Kaski, Panu Somervuo, Krista Lagus, Merja Oja, Vesa Paatero . . . . . .
8.1 Self-organizing maps: introduction 114
8.2 5384 works on SOM 115
8.3 Median self-organizing map of human endogenous retroviruses 116
8.4 Self-organization of very large document collections 118

9 Adaptive cognitive systems **123**
Timo Honkela, Aapo Hyvärinen, Krista Lagus, Ville Könönen, Kevin I. Hynnä, Juha Winter, Jaakko Väyrynen . . . . . .
9.1 Introduction 124
9.2 Unsupervised learning for agent communication 125
9.3 Reinforcement learning in multiagent systems 127
9.4 Emergence of linguistic features using Independent Component Analysis . . 130

10 Bioinformatics **133**
Samuel Kaski, Janne Nikkilä, Merja Oja, Leo Lahti, Jarkko Venna, Eerika Savia, Janne Sinkkonen, Jaakko Peltonen . . . . . .
10.1 Introduction 134
10.2 Exploratory analysis of gene expression 135
10.3 Exploratory analysis of dependencies between functional genomics data sets 137

11 Learning metrics **141**
Samuel Kaski, Janne Sinkkonen, Jaakko Peltonen, Jarkko Venna, Arto Klami, Jarkko Salojärvi . . . . . .
11.1 Introduction 142
11.2 Learning metrics for information visualization 143
11.3 Discriminative Clustering (DC) 145

11.4 Discriminative components . . . 148
11.5 Visualization of posterior distributions . . . 149

12 Natural language processing **153**
Krista Lagus, Mathias Creutz, Mikko Kurimo, Krister Lindén . . .
12.1 Unsupervised segmentation of words into morphs . . . 154
12.2 Word sense disambiguation using document maps . . . 157
12.3 Topically focusing language model . . . 159
12.4 Semantic analysis of Finnish words and sentences . . . 162

13 Speech recognition **165**
Mikko Kurimo, Panu Somervuo, Vesa Siivola, Teemu Hirsimäki . . .
13.1 Acoustic modeling . . . 166
13.2 Language modeling . . . 168
13.3 Large vocabulary decoder . . . 170

14 SOM in data mining **171**
Esa Alhoniemi, Johan Himberg, Jaakko Hollmén, Sampsa Laine, Pasi Lehtimäki, Kimmo Raivio, Timo Similä, Olli Simula, Miki Sirola, Mika Sulkava, Jarkko Tikka, Juha Vesanto . . .
14.1 Introduction . . . 172
14.2 Clustering of the SOM . . . 173
14.3 Use of operator maps to analyze mobile access network . . . 174
14.4 Use of LogSig-scaling to incorporate expert knowledge to SOM-based visualization of GSM-network data . . . 175
14.5 Analysis of mobile access network . . . 176
14.6 Impact of R&D on growth quantiles in manufacturing firms . . . 177

15 Intelligent data engineering **179**
Esa Alhoniemi, Jaakko Hollmen, Johan Himberg, Sampsa Laine, Golan Lampi, Pasi Lehtimäki, Teppo Marin, Jukka Parviainen, Kimmo Raivio, Timo Similä, Olli Simula, Miki Sirola, Mika Sulkava, Jarkko Tikka, Juha Vesanto . . .
15.1 Spatio-temporal analysis of forest nutrition data . . . 180
15.2 Using visualization, variable selection and feature extraction to learn from industrial data . . . 181
15.3 Decision models for computerized decision support . . . 182
15.4 Context awareness . . . 183
15.5 Dependency trees from industrial time-series data . . . 185

16 Other projects **187**
16.1 PRIMA—Proactive information retrieval by adaptive models of users' attention and interests . . . 188
16.2 Data analysis using a tree-shaped neural network . . . 190
16.3 Computational model of visual attention . . . 191

Publications of the Neural Networks Research Centre **197**

II From Data to Knowledge Research Unit: Research Projects under the CIS Laboratory

17 From Data to Knowledge Research Unit **215**

Heikki Mannila, Jaakko Hollmén, Ella Bingham, Johan Himberg, Mikko Koivisto, Anne Patrikainen, Salla Ruosaari, Jouni K. Seppänen, Mikko Katajamaa, Heli Juntunen, Nikolaj Tatti, Antti Rasinen, Kalle Korpiaho, Jaripekka Juhala, Antti Savolainen, Mikko Korpela, Janne Toivonen . . . .

17.1 Data mining 216

17.2 Latent topics in 0-1 data 218

17.3 Applications in bioinformatics 220

Publications of the From Data to Knowledge Research Unit **223**