

# Contents

<b>Preface</b>	<b>7</b>
<b>Personnel</b>	<b>9</b>
<b>Awards and activities</b>	<b>13</b>
<b>Courses</b>	<b>29</b>
<b>Doctoral dissertations</b>	<b>35</b>
<b>Theses</b>	<b>49</b>
<b>I–Adaptive Informatics Research Centre: Research Projects</b>	
<b>1 Introduction</b>	<b>55</b>
<i>Algorithms and Methods</i>	
<b>2 Bayesian learning of latent variable models</b>	<b>59</b>
<i>Juha Karhunen, Antti Honkela, Tapani Raiko, Markus Harva, Alexander Ilin, Matti Tornio, Harri Valpola . . . . .</i>	
2.1 Bayesian modeling and variational learning: introduction . . . . .	60
2.2 Natural conjugate gradient in variational inference . . . . .	62
2.3 Building blocks for variational Bayesian learning . . . . .	64
2.4 Nonlinear BSS and ICA . . . . .	65
2.5 Nonlinear state-space models . . . . .	66
2.6 Non-negative blind source separation . . . . .	70
2.7 PCA in the presence of missing values . . . . .	71
2.8 Predictive uncertainty . . . . .	72
2.9 Relational models . . . . .	73
2.10 Applications to astronomy . . . . .	75
<b>3 Independent component analysis and blind source separation</b>	<b>79</b>
<i>Erkki Oja, Juha Karhunen, Alexander Ilin, Antti Honkela, Karthikesh Raju, Tomas Ukkonen, Zhirong Yang, Zhijian Yuan . . . . .</i>	
3.1 Introduction . . . . .	80
3.2 Convergence and finite-sample behaviour of the FastICA algorithm . . . . .	81
3.3 Independent subspaces with decoupled dynamics . . . . .	84
3.4 Extending ICA for two related data sets . . . . .	85
3.5 ICA in CDMA communications . . . . .	86
3.6 Non-negative projections . . . . .	88

3.7	Climate data analysis with DSS . . . . .	90
-----	--	----

#### 4 Modeling of relevance 93

*Samuel Kaski, Jaakko Peltonen, Kai Puolamäki, Janne Sinkkonen, Jarkko Venna, Arto Klami, Jarkko Salojärvi, Eerika Savia . . . . .*

4.1	Introduction . . . . .	94
4.2	Relevance through data fusion . . . . .	95
4.3	Relevant subtask learning . . . . .	98
4.4	Discriminative generative modeling . . . . .	100
4.5	Visualization methods . . . . .	101
4.6	Networks . . . . .	103

### *Bioinformatics and Neuroinformatics*

#### 5 Bioinformatics 107

*Samuel Kaski, Janne Nikkilä, Merja Oja, Jaakko Peltonen, Jarkko Venna, Antti Ajanki, Andrey Ermolov, Ilkka Huopaniemi, Arto Klami, Leo Lahti, Jarkko Salojärvi, Abhishek Tripathi . . . . .*

5.1	Introduction . . . . .	108
5.2	Translational medicine on metabolical level . . . . .	109
5.3	Visualizing gene expression and interaction data . . . . .	111
5.4	Fusion of gene expression and other biological data sets . . . . .	113
5.5	Human endogenous retroviruses . . . . .	115

#### 6 Neuroinformatics 117

*Ricardo Vigário, Jaakko Särelä, Sergey Borisov, Astrid Pietilä, Jan-Hendrik Schleimer, Jarkko Ylipaavalniemi, Alexander Ilin, Samuel Kaski, Eerika Savia, Erkki Oja . . . . .*

6.1	Introduction . . . . .	118
6.2	Reliable ICA and subspaces . . . . .	120
6.3	Towards brain correlates of natural stimuli . . . . .	122
6.4	Synchrony exploration . . . . .	123
6.5	Overview of other topics . . . . .	125

### *Multimodal interfaces*

#### 7 Content-based information retrieval and analysis 131

*Erkki Oja, Jorma Laaksonen, Markus Koskela, Ville Viitaniemi, Zhirong Yang, Mats Sjöberg, Hannes Muurinen . . . . .*

7.1	Introduction . . . . .	132
7.2	Benchmark tasks of natural image content analysis . . . . .	132
7.3	Interactive facial image retrieval . . . . .	134
7.4	Content analysis and change detection in earth observation images . . . . .	135
7.5	Multimodal hierarchical objects in video retrieval . . . . .	136
7.6	Semantic concept detection . . . . .	137
7.7	Shot boundary detection . . . . .	138
7.8	Video summarization . . . . .	139

<b>8</b>	<b>Automatic speech recognition</b>	<b>143</b>
	<i>Mikko Kurimo, Kalle Palomäki, Vesa Siivola, Teemu Hirsimäki, Janne Pykkönen, Ville Turunen, Sami Virpioja, Matti Varjokallio, Ulpu Remes, Antti Puurula</i> . . . . .	
8.1	Introduction . . . . .	144
8.2	Acoustic modeling . . . . .	146
8.3	Language modeling . . . . .	150
8.4	Applications and tasks . . . . .	152
<b>9</b>	<b>Proactive information retrieval</b>	<b>157</b>
	<i>Samuel Kaski, Kai Puolamäki, Antti Ajanki, Jarkko Salojärvi</i> . . . . .	
9.1	Introduction . . . . .	158
9.2	Implicit queries from eye movements . . . . .	159
<b>10</b>	<b>Natural language processing</b>	<b>161</b>
	<i>Krista Lagus, Mikko Kurimo, Timo Honkela, Mathias Creutz, Jaakko J. Väyrynen, Sami Virpioja, Ville Turunen, Matti Varjokallio</i> . . . . .	
10.1	Unsupervised segmentation of words into morphs . . . . .	162
10.2	Morpho Challenge . . . . .	166
10.3	Emergence of linguistic features using independent component analysis . . . . .	167
 <i>Computational Cognitive Systems</i>		
<b>11</b>	<b>Emergence of linguistic and cognitive representations</b>	<b>173</b>
	<i>Timo Honkela, Krista Lagus, Tiina Lindh-Knuutila, Matti Pöllä, Juha Raitio, Sami Virpioja, Jaakko J. Väyrynen and Paul Wagner</i> . . . . .	
11.1	Introduction . . . . .	174
11.2	Research on emergence . . . . .	174
11.3	Events and projects . . . . .	175
<b>12</b>	<b>Learning social interactions between agents</b>	<b>177</b>
	<i>Ville Könönen, Timo Honkela, Tiina Lindh-Knuutila, Mari-Sanna Paukkeri</i>	
12.1	Introduction . . . . .	178
12.2	Applications of multiagent reinforcement learning . . . . .	179
12.3	Meaning negotiation using simulated naming games . . . . .	180
<b>13</b>	<b>Learning to translate</b>	<b>183</b>
	<i>Timo Honkela, Mathias Creutz, Tiina Lindh-Knuutila, Sami Virpioja, Jaakko J. Väyrynen</i> . . . . .	
13.1	Introduction . . . . .	184
13.2	Analyzing structural complexity of languages . . . . .	186
13.3	Morphology-Aware Statistical Machine Translation . . . . .	188
13.4	Self-Organizing Semantic Representations for Machine Translation . . . . .	191
<b>14</b>	<b>Knowledge translation and innovation using adaptive informatics</b>	<b>195</b>
	<i>Timo Honkela, Mikaela Klami, Matti Pöllä, Ilari Nieminen</i> . . . . .	
14.1	Introduction . . . . .	196
14.2	Statistical machine learning systems as traveling computational models . . . . .	197
14.3	Modeling and simulating practices . . . . .	199
14.4	Analysis of interdisciplinary Text Corpora . . . . .	201
14.5	Quality analysis of medical web content . . . . .	203

***Adaptive Informatics Applications*****15 Intelligent data engineering 207**

*Olli Simula, Jaakko Hollmén, Kimmo Raivio, Miki Sirola, Timo Similä, Mika Sulkava, Pasi Lehtimäki, Jarkko Tikka, Jukka Parviainen, Jaakko Talonen, Golan Lampi, Mikko Multanen, Tuomas Alhonnoro, Risto Hakala . . .*

- 15.1 Failure management with data analysis . . . . . 208
- 15.2 Cellular network performance analysis . . . . . 210
- 15.3 Predictive GSM network optimization . . . . . 211
- 15.4 Learning from environmental data . . . . . 213
- 15.5 Parsimonious signal representations in data analysis . . . . . 216

**16 Time series prediction 219**

*Amaury Lendasse, Francesco Corona, Antti Sorjamaa, Elia Liitiäinen, Tuomas Kärnä, Yu Qi, Emil Eirola, Yoan Miché, Yongnang Ji, Olli Simula . .*

- 16.1 Introduction . . . . . 220
- 16.2 European Symposium on Time Series Prediction . . . . . 221
- 16.3 Methodology for long-term prediction of time series . . . . . 222
- 16.4 Nonparametric noise estimation . . . . . 223
- 16.5 Chemoinformatics . . . . . 224

***Individual projects***

- A. Approximation of an input data item by a linear mixture of SOM models . . 229
- B. Independent variable group analysis . . . . . 232
- C. Analysis of discrete diffusion scale-spaces . . . . . 234
- D. Feature selection for steganalysis . . . . . 236
- E. Adaptive committee techniques . . . . . 238

**Publications of the Adaptive Informatics Research Centre 239****II—From Data to Knowledge Research Unit: Research Projects under the CIS Laboratory****17 From Data to Knowledge Research Unit 261**

*Heikki Mannila, Jaakko Hollmén, Kai Puolamäki, Gemma Garriga, Jouni Seppänen, Robert Gwadera, Sami Hanhijärvi, Hannes Heikinheimo, Samuel Myllykangas, Antti Ukkonen, Nikolaj Tatti, Jarkko Tikka . . . . .*

- 17.1 Finding and using patterns . . . . . 262
- 17.2 Data mining theory . . . . . 268
- 17.3 Analyzing ordered data . . . . . 273
- 17.4 Randomization methods in data analysis . . . . . 276
- 17.5 Applications . . . . . 278
- 17.6 Segmentation . . . . . 281

**Publications of the From Data to Knowledge Research Unit 283**