ICANN 2011 Programme

June 14-17th, 2011

Foreword

The International Conference on Artificial Neural Networks (ICANN) is the annual flagship conference of the European Neural Network Society (ENNS). The idea of ICANN is to bring together researchers from two worlds: information sciences and neurosciences. The scope is wide, ranging from machine learning algorithms to models of real nervous systems. The aim is to facilitate discussions and interactions in the effort towards developing more intelligent artificial systems and increasing our understanding of neural and cognitive processes in the brain.

In 2011, ICANN returns to its roots after 20 years (for more information, see the ENNS web page www.e-nns.org). The very first ICANN in 1991 was organized on the premises of Helsinki University of Technology on its beautiful campus in Espoo, Finland. For ICANN 2011, we invited all neural network researchers worldwide to join us in celebrating this 20th anniversary of ICANN and to see the latest advancements in our fast progressing field.

ICANN 2011 has two basic tracks: Brain-inspired computing and Machine learning research, with program committee chairs from both worlds and a thorough reviewing system. The conference structure was built around plenary talks given by renowned scientists described briefly in the following.

- Prof. Thomas Griffiths (University of California, Berkeley, USA) is the Director of the Computational Cognitive Science Lab and the Institute of Cognitive and Brain Sciences at the University of California, Berkeley. His group develops, for instance, mathematical models of higher level cognition, including probabilistic reasoning, learning causal relationships, acquiring and using language, and inferring the structure of categories.
- Prof. Riitta Hari (Aalto University, Finland) is an internationally recognized and respected neuroscientist. She was newly appointed as Academician of Science, a title that can be held by only twelve scientists at a time in Finland. She has developed methods and applications of human brain imaging and contributed decisively to the progress of this branch of science. Prof. Hari's current focus is on the brain basis of social interaction.
- Prof. Geoffrey Hinton (University of Toronto, Canada), the first winner of the David E. Rumelhart Prize, has provided many influential contributions to the area of artificial neural networks and adaptive systems. A non-exhaustive list of the areas where he has contributed substantial inventions includes back-propagation algorithm, Boltzmann machines, distributed representations, time-delay neural networks, and mixtures of experts. Prof. Hinton was the founding director of the Gatsby Computational Neuroscience Unit at University College London.

- Prof. Aapo Hyvärinen (University of Helsinki, Finland) is widely known for his contributions for the theory and applications of independent component analysis. His recent work also includes research on natural image statistics. He has published in the major journals in the areas of neural networks and machine learning and his books have been translated into Japanese and Chinese.
- Prof. John Shawe-Taylor (University College London, UK) is the Director of the centre for computational statistics and machine learning at University College London, and the scientific coordinator of the Network of Excellence in Pattern Analysis, Statistical Modelling and Computational Learning (PASCAL). His main research area is statistical learning theory, but his contributions range from neural networks and machine learning to graph theory. He is the co-author of two very successful books on the theory of support vector machines and kernel methods.
- Prof. Joshua Tenenbaum (Massachusetts Institute of Technology, USA) is a prominent researcher in the area of computational cognitive science. With his research group, he explores topics such as learning concepts, judging similarity, inferring causal connections, forming perceptual representations, and inferring mental states of other people.

A special plenary talk, shared with the co-located WSOM 2011, Workshop on Self-Organizing Maps, is given by Prof. Emer. Teuvo Kohonen, Academician of Science. He has introduced several new concepts to neural computing including theories of distributed associative memory and optimal associative mappings, the learning subspace method, the self-organizing maps (SOM), the learning vector quantization (LVQ) and the adaptive-subspace SOM (ASSOM) in which invariant-feature filters emerge. Academician Teuvo Kohonen was the initiator and chair of the first ICANN conference in 1991.

The technical program of ICANN 2011 consists of 106 oral or poster presentations that highlight key advances in the areas of neural networks and statistical machine learning research. The overall quality of the contributions can be considered high, also due to the high rejection rate. Approximately only every fourth submission was accepted to be presented orally in the conference. In addition to the regular conference sessions, one day is devoted to five workshops on topics related to theory and applications of brain-inspired computing and statistical machine learning. Two of the workshops are related to special challenges. A mind reading competition on MEG data is sponsored by the PASCAL network of excellence. META-NET network of excellence sponsored a workshop on the use of context in machine translation.

The organizers have a chance to welcome the participants to the new Aalto University School of Science. Namely from the beginning of 2010, the one hundred years old Helsinki University of Technology changed its name and form. It merged with Helsinki School of Economics and University of Art and Design Helsinki into Aalto University which became the second largest university in Finland. The conference is organized at

Aalto University School of Science and the nearby Dipoli Congress Center. Both are centrally located in Otaniemi, Espoo, 15 minutes west of Helsinki. Otaniemi features a unique mix of world-class research organizations, academic institutions and over 800 companies from start-ups to multinational corporations operating around a compact campus. Otaniemi has been twice selected by the EU as one of the most innovative regions in Europe. It is a community of over 32,000 people with 16,000 students and 16,000 technology professionals.

We warmly thank all the authors of the contributed papers, workshop organizers and presenters. We also gratefully acknowledge the contribution of the plenary speakers whose presentations form the backbone of the conference. We express our gratitude to the highly respected international Area Chairs and members of the Program Committee whose role have been instrumental for the success of the conference. The Area Chairs, Program Committee members and the reviewers have ensured a timely and thorough evaluation of the papers.

We are grateful to the members of the Executive Committee whose contributions were essential in ensuring successful organization of the conference. Prof. Erkki Oja as the General Chair lead the conference organizations with his great experience. Dr. Amaury Lendasse, the Local Chair, kept all details of the organization under control. Ms. Mari-Sanna Paukkeri committed a lot of work to compile the proceedings. Mr. Ilari Nieminen took care of numerous details in the arrangements, especially related to the review process and compilation of the proceedings. Ms. Laura Kainulainen efficiently handled the matters related to registrations in collaboration with the Aalto University staff. Dr. Jaakko Peltonen, the Publicity Chair, made sure that the conference was announced in all major forums. Dr. Alexander Ilin took good care of the workshop organizations. Dr. Francesco Corona as the Finance Chair ensured that the budget stayed in balance. Dr. Yoan Miche contributed in several practical areas in the arrangements including the web. Dr. Ricardo Vigário was responsible for the social program and, in particular, the arrangements of the conference dinner. Mr. Tommi Vatanen organized the activities of the conference assistants and helped in preparing the evening program.

We are grateful to Microsoft Research whose representatives provided us with free access to their Conference Management Tool and helped in setting up the system. Last but not least, we would like to thank Springer for their co-operation in publishing the proceedings in the prestigious Lecture Notes in Computer Science series.

May 2011

Wlodzislaw Duch
Mark Girolami
Timo Honkela
Samuel Kaski

Program Co-chairs

Programme

WEDNESDAY, 15th of June

08:45 **Opening**

Heikki Mannila, Vice-Rector of Aalto University

09:00 **Plenary session**

Riitta Hari: *Towards two-person neuroscience*

10:00 **Coffee**

10:20 **Generative Models (p. 16)**

- Transformation Equivariant Boltzmann Machines
- Improved Learning of Gaussian-Bernoulli Restricted Boltzmann Machines
- A Hierarchical Generative Model of Recurrent Object-Based Attention in the Visual Cortex

11:20 **Brain-Computer Interfaces (p. 17)**

- L1-penalized Linear Mixed-Effects Models for BCI
- Slow Feature Analysis - A Tool for Extraction of Discriminating Event-Related Potentials in Brain-Computer Interfaces

12:00 **Lunch**

13:00 **Plenary session**

Geoffrey Hinton: *Learning structural descriptions of objects using equivariant capsules*

14:00 **Neural and Hybrid Architectures (p. 18)**

- Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction
- Error-backpropagation in Networks of Fractionally Predictive Spiking Neurons
- ESN intrinsic plasticity versus reservoir stability
- Adaptive Routing Strategies for Large Scale Spiking Neural Network Hardware Implementations

Self-organization (p. 20)

- Self-Organizing Map for the Multi-Goal Path Planning with Polygonal Goals
- Unlearning in the BCM learning rule for plastic self-organization in a multi-modal architecture
- Neuronal Projections Can Be Sharpened by a Biologically Plausible Learning Mechanism
- Explicit Class Structure by Weighted Cooperative Learning

15:20 **Coffee**

15:50-17:00 **Poster Spotlight Session (p. 34)**

17:30-21:00 **ICANN Reception and Poster Session (p. 34)**

THURSDAY, 16th of June

09:00 **Plenary session**

Aapo Hyvärinen: *Brain Imaging at rest: the ultimate neuroscience data set?*

10:00 **Coffee**

10:20 **Kernel Methods (p. 22)**

- Unsupervized data-driven partitioning of multiclass problems
- Bounds for Approximate Solutions of Fredholm Integral Equations using Kernel Networks
- An Improved Training Algorithm for the Linear Ranking Support Vector Machine
- Extending Tree Kernels with Topological Information
- Accelerating Kernel Neural Gas

Recurrent Networks and Temporal Processing (p. 24)

- State Prediction: A Constructive Method to Program Recurrent Neural Networks
- Cluster Self-organization of Known and Unknown Environmental Sounds using Recurrent Neural Network
- Time-Dependent Series Variance Estimation via Recurrent Neural Networks
- Historical Consistent Complex Valued Recurrent Neural Network

12:00 **Lunch**

13:00 **Plenary session**

John Shawe-Taylor: *Leveraging the data generating distribution for learning*

14:00 **Bayesian Learning (p. 25)**

- Sparse Spatio-Temporal Gaussian Processes with General Likelihoods
- Learning Curves for Gaussian Processes via Numerical Cubature Integration
- Cross-Species Translation of Multi-Way Biomarkers

Pattern recognition 1 (p. 26)

- An Evaluation of the Image Recognition Method Using Pulse Coupled Neural Network
- Using the Leader Algorithm with Support Vector Machines for Large Data Sets
- Automatic Seizure Detection Incorporating Structural Information

15:00 **Coffee**

15:30 **Topic models and matrix factorization (p. 27)**

- The Grouped Author-Topic Model for Unsupervised Entity Resolution
- Kullback-Leibler Divergence for Nonnegative Matrix Factorization

Dynamical Models (p. 28)

- Distributed deterministic temporal information propagated by feedforward neural networks
- Chaotic Complex-valued Multidirectional Associative Memory with Variable Scaling Factor

16:10-17:00 **ENNS General Assembly**

18:30 **Bus to Conference Dinner**

FRIDAY, 17th of June

09:00 **Plenary session**

Joshua Tenenbaum: *How to grow a mind: Statistics, structure and abstraction*

10:00 **Coffee**

10:20 **Cognitive Processes (p. 29)**

- Predicting Reaction Times in Word Recognition by Unsupervised Learning of Morphology
- An Examination of the Dynamic Interaction within Metaphor Understanding using a Model Simulation
- Visual Pathways for Shape Abstraction
- Improving Articulatory Feature and Phoneme Recognition using Multitask Learning
- OrBEAGLE: Integrating Orthography into a Holographic Model of the Lexicon

Feature Extraction and Complex Networks (p. 31)

- On the Problem of Finding the Least Number of Features by L1-Norm Minimisation
- Extracting Coactivated Features from Multiple Data Sets
- Single Layer Complex Valued Neural Network with Entropic Cost Function

12:00 **Lunch**

13:00 **Plenary session**

Thomas Griffiths: *Discovering human inductive biases*

14:00 **Panel Discussion**

15:00 **Non-Linear Projection (p. 32)**

- Batch Intrinsic Plasticity for Extreme Learning Machines
- An Empirical Study on the Performance of Spectral Manifold Learning Techniques
- Semi-supervised Learning for WLAN Positioning

Pattern Recognition 2 (p. 33)

- Theory of On-Line Ensemble-Teacher Learning through a Perceptron Rule with a Margin
- Topic-dependent Document Ranking: Citation Network Analysis by Analogy to Memory Retrieval in the Brain
- PADDLE: Proximal Algorithm for Dual Dictionaries LEarning

16:00 **Closing**

Organization

ICANN 2011 was held from 14th to 17th of June 2011, organized by the Department of Information and Computer Science, Aalto University School of Science. The head of the department is Professor Olli Simula, who is a long term member of the ENNS Executive Committee.

Executive Committee

General Chair:	Erkki Oja (Aalto University, Finland)
Program Chairs:	Wlodzislaw Duch (Nicolaus Copernicus University, Poland)
	Mark Girolami (University of Glasgow, UK)
	Timo Honkela (Aalto University, Finland)
	Samuel Kaski (Aalto University, Finland)
Local Chair:	Amaury Lendasse (Aalto University, Finland)
Publicity Chair:	Jaakko Peltonen (Aalto University, Finland)

Area Chairs

Peter Auer (Austria)
Christian Bauckhage (Germany)
Wray Buntine (Australia)
Vince Calhoun (USA)
Antonius Coolen (UK)
Fernando Morgado Dias (Portugal)
Barbara Hammer (Germany)
Giulio Jacucci (Finland)
Kristian Kersting (Germany)
Neil Lawrence (UK)
Te-Won Lee (USA)
Mikko Kurimo (Finland)
Hiroshi Mamitsuka (Japan)
Klaus-Robert Müller (Germany)
Klaus Obermayer (Germany)
Cheng Soon Ong (Switzerland)
Jan Peters (Germany)
Marios Polycarpou (Cyprus)
Jose Principe (USA)
Volker Roth (Switzerland)
Craig Saunders (UK)
Masashi Sugiyama (Japan)
Alan Stocker (USA)
Ron Sun (USA)
Alfred Ultsch (Germany)
Peter Tino (UK)
Koen Van Leemput (USA)
Michel Verleysen (Belgium)
Jean-Philippe Vert (France)
Ole Winther (Denmark)
Chang Yoo (Korea)

Program Committee and Referees

Luis Alexandre (Portugal)
Paul Aljabar (UK)
Tanel Alumae (Estonia)
Mauricio Álvarez (UK)
Elisabeth Andre (Germany)
Alessia Annibale (UK)
Ana Antunes (Portugal)
Bruno Apolloni (Italy)
Ebru Arisoy (Turkey)
Beghdadi Azeddine (France)
Vineeth Balasubramaniam (USA)
Peter Battaglia (USA)
Christian Bauckhage (Germany)
Luba Benuskova (New Zealand)
Ulysses Bernardet (Spain)
Monica Bianchini (Italy)
Ginestra Bianconi (USA)
Michael Biehl (Netherlands)
Benjamin Blankertz (Germany)
Jamie Blundel (UK)
Antoine Bordes (Canada)
Abdeslam Boularias (Canada)
Mikio Braun (Germany)
Sebastien Bubeck (France)
Samuel Bulo (Italy)
Guido Caldarelli (Italy)
Cristina Campi (Finland)
Tonatiuh Peña Centeno (UK)
Michal Cernansky (Slovakia)
Jung-Hsien Chiang (Taiwan)
Jens Christian Claussen (Germany)
James Cussens (UK)
Jesse David (Belgium)
Lieven De Lathauwer (Belgium)
Andrea De Martino (Italy)
Yannick Deville (France)
Dotan Di Castro (Israel)
Jan Drugowitsch (France)
Lan Du (Australia)
Carl Ek (Sweden)
Gernot Fink (Germany)
Benoit Frenay (Belgium)
Artu Garcez (UK)
Phil Garner (Switzerland)
Shuzhi Sam Ge (Singapore)
Mohammad Ghavamzadeh (France)
Fausto Giunchiglia (Italy)
Tobias Glasmachers (Switzerland)
Geoff Goodhill (Australia)
Dan Grollman (Switzerland)
Moritz Grosse-Wentrup (Germany)
Hirotaka Hachiya (Japan)
David Hardoon (UK)
Kohei Hatano (Japan)
Sebastien Helie (US)
Ricardo Henao (Denmark)
James Hensman (UK)
Tom Heskes (Netherlands)
Antti Honkela (Finland)
Naira Hovakimyan (USA)
Matthew Howard (UK)
Guang-Bin Huang (Singapore)
Zakria Hussain (UK)
Alexander Ilin (Finland)
Robert Jacob (USA)
Matjaz Jogan (USA)
Shivaram Kalyanakrishnan (USA)
Kittipat Kampa (USA)
Bert Kappen (Netherlands)
Juha Karhunen (Finland)
Matthias Kaschube (USA)
Dietrich Klakow (Germany)
Stefanos Kollias (Greece)
Vera Kurkova (Hungary)
Ville Kyrki (Finland)
Jorma Laaksonen (Finland)
Harri Lähdesmäki (Finland)
Stéphane Lallich (France)
Georg Langs (France)
Jan Larsen (Denmark)

Deniz Erdogmus (USA)
Igor Farkas (Slovakia)
Jason Farquhar (Netherlands)
Santo Fortunato (Italy)
Jyrki Lötjönen (Finland)
Mantas Lukosevicius (Germany)
Zhiyuan Luo (UK)
Mathew Magimai Doss (Switzerland)
Hiroshi Mamitsuka (Japan)
Amir Madany Mamlouk (Germany)
Danilo Mandic (UK)
Naoki Masuda (Japan)
Bjoern Menze (USA)
Alessio Micheli (Italy)
Kazushi Mimura (Japan)
Tetsuro Morimura (Japan)
Morten Mørup (Denmark)
Alexander Mozeika (UK)
Christian Müller (Germany)
Shinichi Nakajima (Japan)
Sriraam Natarajan (USA)
Tim Nattkemper (Germany)
Yizhao Ni (UK)
Matthias Nickles (UK)
Nikolay Nikolaev (UK)
Ann Nowe (Belgium)
Cheng Soon Ong (Switzerland)
Ronald Ortner (Austria)
Erhan Oztop (Japan)
Sebastian Pannasch (Germany)
Ulrich Paquet (UK)
Il Park (USA)
Emilio Parrado (Spain)
Elzbieta Pekalska (UK)
Jaakko Peltonen (Finland)
Andrew Philippides (UK)
Justus Piater (Austria)
Alex Pouget (USA)
Novi Quadrianto (Australia)
Sabrina Rabello (USA)
Achim Rettinger (USA)
Alexis Roche (Switzerland)
Te-Won Lee (USA)
Robert Leeb (Switzerland)
Amaury Lendasse (Finland)
Weifeng Liu (USA)
Mikkel Schmidt (Denmark)
Gerwin Schalk (USA)
Markus Schläpfer (Switzerland)
Benoit Scherrer (USA)
Mikkel Schmidt (Denmark)
Gabriele Schweikert (Germany)
Sambu Seo (Germany)
Sohan Seth (USA)
Vesa Siivola (Finland)
Sören Sonnenburg (Germany)
Emilio Soria-Olivas (Spain)
Alessandro Sperduti (Italy)
Jochen Steil (Germany)
Marc Strickert (Germany)
Masashi Sugiyama (Japan)
Taiji Suzuki (Japan)
Csaba Szepesvari (Canada)
Ichiro Takeuchi (Japan)
Michael Tangermann (Germany)
Franck Thollard (France)
Christian Thurau (Germany)
Michalis Titsias (UK)
Jussi Tohka (Finland)
Ryota Tomioka (Japan)
Antonios Tsourdos (UK)
Koji Tsuda (Japan)
Laurens van der Maaten (Netherlands)
Carmen Vidaurre (Germany)
Ricardo Vigário (Finland)
Silvia Villa (Italy)
Nathalie Villa-Vialaneix (France)
Draguna Vrabie (USA)
Shinji Watanabe (Japan)
Markus Weimer (Germany)
Stefan Wermter (Germany)
Heiko Wersing (Germany)
Daan Wierstra (Switzerland)
Nick Wilson (UK)

Michael Rohs (Germany)
Fabrice Rossi (France)
Volker Roth (Switzerland)
Mert Sabuncu (USA)
Jun Sakuma (Japan)
Scott Sanner (Australia)
Ignacio Santamaria (Spain)
Murat Saraclar (Turkey)
Jagannathan Sarangapani (USA)
Zhao Xu (Germany)
Zenglin Xu (Germany)
Makoto Yamada (Japan)
Yoshihiro Yamanishi (Japan)
Zhirong Yang (Finland)
Chang Yoo (Korea)
Massimo Zancanaro (Italy)
Xinhua Zhang (Canada)
Shanfeng Zhu (China)

Sponsors

ICANN 2011 is supported by

The European Neural Network Society (ENNS)

The Pattern Recognition Society of Finland

The Finnish Artificial Intelligence Society

Workshops

On Tuesday, the following workshops are organized:

- Computational Intelligence for Quality of Life Environmental Information Services
- Validation of Computational Models in Social and Economic Sciences
- Beyond Correlations: Developments in Supervised Learning Algorithms for Spiking Neural Networks
- Challenge Workshop: Mind Reading Competition on MEG Data
- META-NET Workshop: Context in Machine Translation

See the ICANN website for more information.

Generative Models

Transformation Equivariant Boltzmann Machines

Jyri Kivinen and Christopher Williams

We develop a novel modeling framework for Boltzmann machines, augmenting each hidden unit with a latent transformation assignment variable which describes the selection of the transformed view of the canonical connection weights associated with the unit. This enables the inferences of the model to transform in response to transformed input data in a stable and predictable way, and avoids learning multiple features differing only with respect to the set of transformations. Extending prior work on translation equivariant (convolutional) models, we develop translation and rotation equivariant restricted Boltzmann machines (RBMs) and deep belief nets (DBNs), and demonstrate their effectiveness in learning frequently occurring statistical structure from artificial and natural images.

Improved Learning of Gaussian-Bernoulli Restricted Boltzmann

KyungHyun Cho, Alexander Ilin and Tapani Raiko

We propose a few remedies to improve training of Gaussian-Bernoulli restricted Boltzmann machines (GBRBM), which is known to be difficult. Firstly, we use a different parameterization of the energy function, which allows for more intuitive interpretation of the parameters and facilitates learning. Secondly, we propose parallel tempering learning for GBRBM. Lastly, we use an adaptive learning rate which is selected automatically in order to stabilize training. Our extensive experiments show that the proposed improvements indeed remove most of the difficulties encountered when training GBRBMs using conventional methods.

A Hierarchical Generative Model of Recurrent Object-Based Attention in the Visual Cortex

David Reichert, Peggy Series and Amos Storkey

In line with recent work exploring Deep Boltzmann Machines (DBMs) as models of cortical processing, we demonstrate the potential of DBMs as models of object-based attention, combining generative principles with attentional ones. We show: (1) How inference in DBMs can be related qualitatively to theories of attentional recurrent processing in the visual cortex; (2) that deepness and topographic receptive fields are important for realizing the attentional state; (3) how more explicit attentional suppressive mechanisms can be implemented, depending crucially on sparse representations being formed during learning.

Brain-Computer Interfaces

ℓ_1-penalized Linear Mixed-Effects Models for BCI

Siamac Fazli, Márton Danóczy, Jürg Schelldorfer and Klaus-Robert Müller

A recently proposed novel statistical model estimates population effects and individual variability between subgroups simultaneously, by extending Lasso methods. We apply this ℓ_1-penalized linear regression mixed-effects model to a large scale real world problem: by exploiting a large set of brain computer interface data we are able to obtain a subject-independent classifier that compares favorably with prior zero-training algorithms. This unifying model inherently compensates shifts in the input space attributed to the individuality of a subject. In particular we are now able to differentiate *within-subject* and *between-subject variability*. A deeper understanding both of the underlying statistical and physiological structure of the data is gained.

Slow Feature Analysis - A Tool for Extraction of Discriminating Event-Related Potentials in Brain-Computer Interfaces

Sven Dähne, Johannes Höhne, Martijn Schreuder and Michael Tangermann

The unsupervised signal decomposition method Slow Feature Analysis (SFA) is applied as a preprocessing tool in the context of EEG based Brain-Computer Interfaces (BCI). Classification results based on a SFA decomposition are compared to classification results obtained on Principal Component Analysis (PCA) decomposition and to those obtained on raw EEG channels. Both PCA and SFA improve classification to a large extend compared to using no signal decomposition and require between one third and half of the maximal number of components to do so. The two methods extract different information from the raw data and therefore lead to different classification results. Choosing between PCA and SFA based on classification of calibration data leads to a larger improvement in classification performance compared to using one of the two methods alone. Results are based on a large data set ($n = 31$ subjects) of two studies using auditory Event Related Potentials for spelling applications.

Neural and Hybrid Architectures

Transforming Auto-encoders

Geoffrey Hinton, Alex Krizhevsky and Sida Wang

The artificial neural networks that are used to recognize shapes typically use one or more layers of learned feature detectors that produce scalar outputs. By contrast, the computer vision community uses complicated, hand-engineered features, like SIFT, that produce a whole vector of outputs including an explicit representation of the pose of the feature. We show how neural networks can be used to learn features that output a whole vector of instantiation parameters and we argue that this is a much more promising way of dealing with variations in position, orientation, scale and lighting than the methods currently employed in the neural networks community. It is also more promising than the hand-engineered features currently used in computer vision because it provides an efficient way of adapting the features to the domain.

Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction

Jonathan Masci, Ueli Meier, Dan Ciresan and Jürgen Schmidhuber

We present a novel convolutional auto-encoder (CAE) for unsupervised feature learning. A stack of CAEs forms a convolutional neural network (CNN). Each CAE is trained using conventional on-line gradient descent without additional regularization terms. A max-pooling layer is essential to learn biologically plausible features consistent with those found by previous approaches. Initializing a CNN with filters of a trained CAE stack yields superior performance on a digit (MNIST) and an object recognition (CIFAR10) benchmark.

Error-backpropagation in Networks of Fractionally Predictive Spiking Neurons

Sander Bohte

We develop a learning rule for networks of spiking neurons where signals are encoded using fractionally predictive spike-coding. In this paradigm, neural output signals are encoded as a sum of shifted power-law kernels. Simple greedy thresholding can compute this encoding, and spike-trains are then exactly the signal's fractional derivative. Fractionally predictive spike-coding exploits natural statistics and is consistent with observed spike-rate adaptation in real neurons; its multiple-timescale properties also reconciles notions of spike-time coding and spike-rate coding. Previously, we argued that properly tuning the decoding kernel at receiving neurons can implement spectral filtering; the applicability to general temporal filtering was left open. Here, we present an error-backpropagation algorithm to learn these decoding filters, and we show that networks of fractionally predictive spiking neurons can then implement temporal filters such as delayed responses, delayed match-to-sampling, and temporal versions of the XOR problem.

ESN Intrinsic Plasticity versus Reservoir Stability

Petia Koprinkova-Hristova and Guenther Palm

The work presented in this paper was inspired by similarities between intrinsic plasticity (IP) pre-training of the ESN reservoir and the common RNN stability conditions derived from nonlinear control theory. The common theoretical stability conditions were applied to the ESN structure. It was proven that in fact IP training achieves a balance between maximization of entropy at the ESN output and the concentration of that output distribution around the pre-specified mean value. Thus the squeezing of the neuron nonlinearities is produced not only by nonzero biases and translation of the ESN equilibrium state but also by the chosen output distribution mean value. The numerical investigations of different random reservoirs showed that the IP improvement stabilizes even initially unstable reservoirs.

Adaptive Routing Strategies for Large Scale Spiking Neural Network Hardware Implementations

Snaider Carrillo, Jim Harkin, Liam McDaid, Sandeep Pande, Seamus Cawley and Fearghal Morgan

This paper presents an adaptive Network-on-Chip (NoC) router, which forms part of an embedded mixed signal Spiking Neural Network (SNN) architecture called EMBRACE (Emulating Biologically-inspiRed ArChitectures in hardware). The novel adaptive NoC router provides the inter-neuron connectivity for EMBRACE, maintaining router communication and avoiding dropped router packets by adapting to router traffic congestion. The router also adapts to NoC traffic congestion or broken NoC connections (faults) by reconfiguring the routing topology to select an alternative route. Performance, power and area analysis of the proposed adaptive router using Synopsys Design Compiler (for TSMC 90nm CMOS technology) indicates a router throughput of 3.2Gbps on each of 5 available router channels, low router power consumption (1.716mW) and small router area (0.056mm2). Router adaptive behaviour in the presence of applied real-time traffic congestion has been demonstrated on a Virtex II Pro Xilinx FPGA for a 4x2 router array. Results indicate the feasibility of using the proposed adaptive NoC router within a scalable EMBRACE hardware SNN architecture.

Self-organization

Self-Organizing Map for the Multi-Goal Path Planning with Polygonal Goals

Jan Faigl and Libor Preucil

This paper presents a self-organizing map approach for the multi-goal path planning problem with polygonal goals. The problem is to find a shortest closed collision free path for a mobile robot operating in a planar environment represented by a polygonal map. The requested path has to visit a given set of areas where the robot takes measurements in order to find an object of interest. Neurons' weights are considered as points and the solution is found as approximate shortest paths connecting the points (weights). The proposed self-organizing map has less number of parameters than a previous approach based on the self-organizing map for the traveling salesman problem. Moreover, the proposed algorithm provides better solutions within less computational time for problems with high number of polygonal goals.

Unlearning in the BCM Learning Rule for Plastic Self-Organization in a Multi-Modal Architecture

Mathieu Lefort, Yann Boniface and Bernard Girau

An agent moving in a real environment perceives it by numerous noisy sensors which provide some high dimensionality data with unknown topology. In order to interact in this complex and changing environment, according to the active perception theory, the agent needs to learn the correlations between its actions and the changes they induce in the environment. In the perspective of a bio-inspired architecture for the learning of multi-modal correlations, this article focuses on the ability to forget some previously learned selectivity in a model of perceptive map which spatially codes the sensor data. This perceptive map combines the Bienenstock Cooper Munro (BCM) learning rule, which raises a selectivity to a stimulus, with the neural field (NF) theory, which provides spatial constraints to self-organize the selectivities at the map level. The introduction of an unlearning term in the BCM learning rule (BCMu) improves the BCM-NF coupling by providing plasticity to the self-organization.

Neuronal Projections Can Be Sharpened by a Biologically Plausible Learning Mechanism

Matthew Cook, Florian Jug and Christoph Krautz

It is known that neurons can project topographically to their target area, and reciprocal projections back from the target area are typically aligned with the forward projection. However, the wide terminal arbors of individual axons limit the precision of such anatomical reciprocity. This leaves open the question of whether more precise reciprocal connectivity is obtainable through the adjustment of synaptic strengths. We have found that such a sharpening of projections can indeed result from a combination of biologically plausible mechanisms, namely Hebbian learning at synapses, continuous winner-take-all circuitry within areas, and homeostatic activity regulation within neurons. We show that this combination of mechanisms, which we refer to collectively as "sharp learning", is capable of sharpening inter-area projections in a variety of network architectures. Sharp learning offers an explanation for how precise topographic and reciprocal connections can emerge, even in early development.

Explicit Class Structure by Weighted Cooperative Learning

Ryotaro Kamimura

In this paper, we propose a new type of information-theoretic method called "weighted cooperative learning." In this method, two networks, namely, cooperative and uncooperative networks are prepared. The roles of these networks are controlled by the cooperation parameter α. As the parameter is increased, the role of cooperative networks becomes more important in learning. In addition, the importance of input units or variables is incorporated in the learning in terms of mutual information. We applied the method to the housing data from the machine learning database. Experimental results showed that weighted cooperative learning could be used to improve performance in terms of quantization and topographic errors. In addition, we could obtain much clearer class boundaries on the U-matrix by the weighted cooperative learning.

Tuesday

Wednesday

Thursday

Friday

Kernel Methods

Unsupervized Data-Driven Partitioning of Multiclass Problems

Hernan Ahumada, Guillermo Grinblat and Pablo Granitto

Many classification problems of high technological value are multiclass. In the last years, several improved solutions based on the combination of simple classifiers were introduced. An interesting kind of methods creates a hierarchy of sub-problems by clustering prototypes of each one of the classes, but the solution produced by the clustering stage is heavily influenced by the label's information. In this work we introduce a new strategy to solve multiclass problems that makes more use of spatial information than other methods. Based on our previous work on imbalanced problems, we construct a hierarchy of subproblems, but opposite to previous developments, based only on spatial information and not using class labels at any time. We consider different clustering methods (either agglomerative or divisive) for this task. We use an SVM for each sub-problem (if needed, because in several cases the clustering method directly gives a subset with samples of a single class). Using publicly available datasets we compare the new method with several previous approaches, finding promising results.

Bounds for Approximate Solutions of Fredholm Integral Equations using Kernel Networks

Giorgio Gnecco, Vera Kurková and Marcello Sanguineti

Approximation of solutions of integral equations by networks with kernel units is investigated theoretically. There are derived upper bounds on speed of decrease of errors in approximation of solutions of Fredholm integral equations by kernel networks with increasing numbers of units. The estimates are obtained for Gaussian and degenerate kernels.

An Improved Training Algorithm for the Linear Ranking Support Vector

Antti Airola, Tapio Pahikkala and Tapio Salakoski

We introduce an $O(ms + m\log(m))$ time complexity method for training the linear ranking support vector machine, where m is the number of training examples, and s the average number of non-zero features per example. The method generalizes the fastest previously known approach, which achieves the same efficiency only in restricted special cases. The excellent scalability of the proposed method is demonstrated experimentally.

Extending Tree Kernels with Topological Information

Fabio Aiolli, Giovanni Da San Martino and Alessandro Sperduti

The definition of appropriate kernel functions is crucial for the performance of a kernel method. In many of the state-of-the-art kernels for trees, matching substructures are considered independently from their position within the trees. However, when a match happens in similar positions, more strength could reasonably be given to it. Here, we give a systematic way to enrich a large class of tree kernels with this kind of information without affecting, in almost all cases, the worst case computational complexity. Experimental results show the effectiveness of the proposed approach.

Tuesday

Wednesday

Thursday

Friday

Accelerating Kernel Neural Gas

Frank-Michael Schleif, Andrej Gisbrecht and Barbara Hammer

Clustering approaches constitute important methods for unsupervised data analysis. Traditionally, many clustering models focus on spherical or ellipsoidal clusters in Euclidean space. Kernel methods extend these approaches to more complex cluster forms, and they have been recently integrated into several clustering techniques. While leading to very flexible representations, kernel clustering has the drawback of high memory and time complexity due to its dependency on the full Gram matrix and its implicit representation of clusters in terms of feature vectors. In this contribution, we accelerate the kernelized Neural Gas algorithm by incorporating a Nyström approximation scheme and active learning, and we arrive at sparse solutions by integration of a sparsity constraint. We provide experimental results which show that these accelerations do not lead to a deterioration in accuracy while improving time and memory complexity.

Tuesday

Wednesday

Thursday

Friday

Recurrent Networks and Temporal Processing

State Prediction: A Constructive Method to Program Recurrent Neural Networks

R. Felix Reinhart and Jochen Steil

We introduce a novel technique to program desired state sequences into recurrent neural networks in one shot. The basic methodology and its scalability to large and input-driven networks is demonstrated by shaping attractor landscapes, transient dynamics and programming limit cycles. The approach unifies programming of transient and attractor dynamics in a generic framework.

Cluster Self-organization of Known and Unknown Environmental Sounds using Recurrent Neural Network

Yang Zhang, Shun Nishide, Toru Takahashi, Hiroshi Okuno and Tetsuya Ogata

Our goal is to develop a system that is able to learn and classify environmental sounds for robots working in the real world. In the real world, two main restrictions pertain in learning. First, the system has to learn using only a small amount of data in a limited time because of hardware restrictions. Second, it has to adapt to unknown data since it is virtually impossible to collect samples of all environmental sounds. We used a neuro-dynamical model to build a prediction and classification system which can self-organize sound classes into parameters by learning samples. The proposed system searches space of parameters for classifying. In the experiment, we evaluated the accuracy of classification for known and unknown sound classes.

Time-Dependent Series Variance Estimation via Recurrent Neural Networks

Nikolay Nikolaev, Peter Tino and Evgueni Smirnov

This paper presents a nonlinear model for computing the time-dependent evolution of the variance in time series of returns on assets. First, we design a recurrent network representation of the variance, which extends the typically linear models. Second, we derive temporal training equations with which the network weights are inferred so as to maximize the likelihood of the data. Experimental results show that this dynamic recurrent network model yields results with improved statistical characteristics and economic performance.

Historical Consistent Complex Valued Recurrent Neural Network

Hans-Georg Zimmermann, Alexey Minin and Victoria Kusherbaeva

Recurrent Neural Networks are in the scope of the community for many years. The are some obvious problems in constructing and training such networks. One of the natural extensions of recurrent neural networks is complex valued recurrent neural network, since modeling of complex valued dynamics is very interesting subject in electrical engineering and temporal dynamics analysis. Complex Valued Neural Networks are very difficult to train. In the current paper we will discuss the so called Historical Consistent Recurrent Neural Network and it's extension to the complex valued case. We will discuss complex valued back propagation and it's application to the recurrent neural network. We will show, that training of recurrent neural network of this type is robust and stable. As a result of the work done we will present the results of this network for the chaotic Lorenz system modeling (real valued dynamics) and for the electrical circuit modeling (complex valued dynamics). At the end we will discuss the advantages and disadvantages of the proposed algorithms and will give the outlook.

Bayesian Learning

Sparse Spatio-Temporal Gaussian Processes with General Likelihoods

Jouni Hartikainen, Jaakko Riihimäki and Simo Särkkä

In this paper, we consider learning of spatio-temporal processes by formulating a Gaussian process model as a solution to an evolution type stochastic partial differential equation. Our approach is based on converting the stochastic infinite-dimensional differential equation into a finite dimensional linear time invariant (LTI) stochastic differential equation (SDE) by discretizing the process spatially. The LTI SDE is time-discretized analytically, resulting in a state space model with linear-Gaussian dynamics. We use expectation propagation to perform approximate inference on non-Gaussian data, and show how to incorporate sparse approximations to further reduce the computational complexity. We briefly illustrate the proposed methodology with a simulation study and with a real world modelling problem.

Learning Curves for Gaussian Processes via Numerical Cubature Integration

Simo Särkka

This paper is concerned with estimation of learning curves for Gaussian process regression with multidimensional numerical integration. We propose an approach where the recursion equations for the generalization error are approximately solved using numerical cubature integration methods. The advantage of the approach is that the eigenfunction expansion of the covariance function does not need to be known. The accuracy of the proposed method is compared to eigenfunction expansion based approximations to the learning curve.

Cross-Species Translation of Multi-Way Biomarkers

Tommi Suvitaival, Ilkka Huopaniemi, Matej Oreši and Samuel Kaski

We present a Bayesian translational model for matching patterns in data sets which have neither co-occurring samples nor variables, but only a similar experiment design dividing the samples into two or more categories. The model estimates covariate effects related to this design and separates the factors that are shared across the data sets from those specific to one data set. The model is designed to find similarities in medical studies, where there is great need for methods for linking laboratory experiments with model organisms to studies of human diseases and new treatments.

Pattern Recognition 1

An Evaluation of the Image Recognition Method Using Pulse Coupled Neural Network

Masato Yonekawa and Hiroaki Kurokawa

A technique for the image recognition is major issue in the image processing and the image recognition method using pulse coupled neural network (PCNN) have been studied as one of the valid method. The most outstanding feature of the method using PCNN is that the method is valid for the rotation, magnification and shrinking of the image. Also, the good compatibility to the hardware implementation is significant feature of the PCNN. In our previous study, we proposed the GA based learning method for the PCNN parameters which enable the reliable results of image recognition. In this study, we evaluate the image recognition method using PCNN with our learning method. In the simulation results, we clarify the characteristics of recognition rate to the number of the images to be learned using our proposed learning method.

Using the Leader Algorithm with Support Vector Machines for Large Data Sets

Enrique Romero

One of the main drawbacks of Support Vector Machines (SVM) is their high computational cost for large data sets. We propose the use of the Leader algorithm as a preprocessing procedure for SVM with large data sets, so that the obtained leaders are used as the training set for the SVM. The result is an algorithm where the Leader algorithm allows to construct a sample of the data set whose granularity level and computational cost are controlled by the threshold parameter. Despite its apparent simplicity, the proposed model obtains similar accuracies to standard *LIBSVM* with fewer number of support vectors and less execution times.

Automatic Seizure Detection Incorporating Structural Information

Borbala Hunyadi, Maarten De Vos, Marco Signoretto, Johan Suykens, Wim Van Paesschen and Sabine Van Huffel

Traditional seizure detection algorithms act on single channels ignoring the synchronously recorded, inherently interdependent multichannel nature of EEG. However, the spatial distribution and evolution of the ictal pattern is a crucial characteristic of the seizure. Two different approaches aiming at including such structural information into the data representation are presented in this paper. Their performance is compared to the traditional approach both in a simulation study and a real-life example, showing that spatial and structural information facilitates precise classification.

Topic models and matrix factorization

The Grouped Author-Topic Model for Unsupervised Entity Resolution

Andrew Dai and Amos Storkey

This paper describes a generative approach for tackling the problem of identity resolution in a completely unsupervised context with no fixed assumption regarding the true number of identities. The problem of entity resolution involves associating different references to authors (in a paper's author list, for example) with real underlying identities. The references may be written in differing forms or may have errors, and identical references may refer to different real identities. The approach taken here uses a generative model of both the abstract of a document and its list of authors to resolve identities in a corpus of documents. In the model, authors and topics are associated with latent groups. For each document, an abstract and an author list are generated conditioned on a given group. Results are presented on real-world datasets, and outperform the best performing unsupervised methods.

Kullback-Leibler Divergence for Nonnegative Matrix Factorization

Zhirong Yang, He Zhang, Zhijian Yuan and Erkki Oja

The I-divergence or unnormalized generalization of Kullback-Leibler (KL) divergence is commonly used in Nonnegative Matrix Factorization (NMF). This divergence has the drawback that its gradients with respect to the factorizing matrices depend heavily on the scales of the matrices, and learning the scales in gradient-descent optimization may require many iterations. This is often handled by explicit normalization of one of the matrices, but this step may actually increase the I-divergence and is not included in the NMF monotonicity proof. A simple remedy that we study here is to normalize the input data. Such normalization allows the replacement of the I-divergence with the original KL-divergence for NMF and its variants. We show that using KL-divergence takes the normalization structure into account in a very natural way and brings improvements for nonnegative matrix factorizations: the gradients of the normalized KL-divergence are well-scaled and thus lead to a new projected gradient method for NMF which runs faster or yields better approximation than three other widely used NMF algorithms.

Dynamical Models

Distributed Deterministic Temporal Information Propagated by Feedforward Neural Networks

Yoshiyuki Asai and Alessandro Villa

A ten layers feedforward network characterized by diverging/converging patterns of projection between successive layers is activated by an external spatio-temporal input pattern fed to layer 1 in presence of stochastic background activities fed to all layers. We used three dynamical systems to derive the external input spike trains including the temporal information, and two types of neuron models for the network, i.e. either a simple spiking neuron (SSN) or a multiple-timescale adaptive threshold neuron (MAT). We observed an unimodal integration effect as a function of the order of the layers and confirmed that the MAT model is likely to be more efficient in integrating and transmitting the temporal structure embedded in the external input.

Chaotic Complex-valued Multidirectional Associative Memory with Variable Scaling Factor

Akio Yoshida and Yuko Osana

In this paper, we propose a Chaotic Complex-valued Multidirectional Associative Memory (CCMAM) with variable scale factor which can realize one-to-many associations of M-tuple multi-valued patterns. The proposed model is based on the Multidirectional Associative Memory, and is composed of complex-valued neurons and chaotic complex-valued neurons. In the proposed model, associations of multi-valued patterns are realized by using complex-valued neurons, and one-to-many associations are realized by using chaotic complex-valued neurons. Moreover, in the proposed model, the appropriate parameters of chaotic complex-valued neurons can be determined easily than in the original Chaotic Complex-valued Multidirectional Associative Memory. We carried out a series of computer experiments and confirmed that the proposed model has superior one-to-many association ability than that of the conventional model.

Cognitive Processes

Predicting Reaction Times in Word Recognition by Unsupervised Learning of Morphology

Sami Virpioja, Minna Lehtonen, Annika Hultén, Riitta Salmelin and Krista Lagus

A central question in the study of the mental lexicon is how morphologically complex words are processed. We consider this question from the viewpoint of statistical models of morphology. As an indicator of the mental processing cost in the brain, we use reaction times to words in a visual lexical decision task on Finnish nouns. Statistical correlation between a model and reaction times is employed as a goodness measure of the model. In particular, we study Morfessor, an unsupervised method for learning concatenative morphology. The results for a set of inflected and monomorphemic Finnish nouns reveal that the probabilities given by Morfessor, especially the Categories-MAP version, show considerably higher correlations to the reaction times than simple word statistics such as frequency, morphological family size, or length. These correlations are also higher than when any individual test subject is viewed as a model.

An Examination of the Dynamic Interaction within Metaphor Understanding using a Model Simulation

Asuka Terai, Saori Hirose, Naoko Kuriyama and Masanori Nakagawa

The purpose of this study is to examine the understanding mechanism for metaphors represented in the form of "A is like B" (simile) using model simulation. In a previous experimental study, the priming effects of visual images on metaphor processing were demonstrated in a psychological experiment. In that experiment, the presentation of a picture of a vehicle ("B") interfered with metaphor comprehension, even when the picture did not directly inhibit features that relate to metaphor interpretation. The previous research has suggested that priming effects arise from interaction among features. In this research, in order to elucidate the dynamic interaction among features within metaphor understanding, the priming effect is examined by simulating computational models that do and do not incorporate detailed processes of dynamic interaction. Furthermore, the strengths of the dynamic interactions among features are estimated as the parameters of the model.

Visual Pathways for Shape Abstraction

Konstantinos Raftopoulos and Stefanos Kollias

The Medial Axis Transform (MAT) (or skeleton transform) is one of the most studied shape representation techniques with established advantages for general 2D shape recognition. Embedding local boundary information in the skeleton, in particular, has been shown to improve 2D shape recognition capability to state of the art levels. In this paper we present a visual pathway for extracting an analogous to the MAT skeleton abstraction of shape that also contains local boundary curvature information. We refer to this structure with the term *curvature-skeleton*. The proposed architecture is inspired by the biological findings regarding the cortical neurons of the visual cortex and their special purpose Receptive Fields (RFs). Points of high curvature are initially identified and subsequently combined by means of a visual pathway that achieves an analogous to the MAT abstraction of shape but also embeds in the skeleton local curvature information of the shape's boundary. We present experimental results illustrating that such an abstraction can improve the recognition capability of multi layered neural network classifiers.

Improving Articulatory Feature and Phoneme Recognition using Multitask Learning

Ramya Rasipuram and Mathew Magimai Doss

Speech sounds can be characterized by articulatory features. Articulatory features are typically estimated using a set of multilayer perceptrons (MLPs), i.e., a separate MLP is trained for each articulatory feature. In this paper, we investigate multitask learning (MTL) approach for joint estimation of articulatory features with and without phoneme classification as subtask. Our studies show that MTL MLP can estimate articulatory features compactly and efficiently by learning the inter-feature dependencies through a common hidden layer representation. Furthermore, adding phoneme as subtask while estimating articulatory features improves both articulatory feature estimation and phoneme recognition. On TIMIT phoneme recognition task, articulatory feature posterior probabilities obtained by MTL MLP achieve a phoneme recognition accuracy of 73.2%, while the phoneme posterior probabilities achieve an accuracy of 74.0%.

OrBEAGLE: Integrating Orthography into a Holographic Model of the Lexicon

George Kachergis, Gregory Cox and Michael Jones

Many measures of human verbal behavior deal primarily with semantics (e.g., associative priming, semantic priming). Other measures are tied more closely to orthography (e.g., lexical decision time, visual word-form priming). Semantics and orthography are thus often studied and modeled separately. However, given that concepts must be built upon a foundation of percepts, it seems desirable that models of the human lexicon should mirror this structure. Using a holographic, distributed representation of visual word-forms in BEAGLE, a corpus-trained model of semantics and word order, we show that free association data is better explained with the addition of orthographic information. However, we find that orthography plays a minor role in accounting for cue-target strengths in free association data. Thus, it seems that free association is primarily conceptual, relying more on semantic context and word order than word form information.

Feature Extraction and Complex Networks

On the Problem of Finding the Least Number of Features by L1-Norm minimisation

Sascha Klement and Thomas Martinetz

Recently, the so-called Support Feature Machine (SFM) was proposed as a novel approach to feature selection for classification. It relies on approximating the zero-norm minimising weight vector of a separating hyperplane by optimising for its one-norm. In contrast to the L1-SVM it uses an additional constraint based on the average of data points. In experiments on artificial datasets we observe that the SFM is highly superior in returning a lower number of features and a larger percentage of truly relevant features. Here, we derive a necessary condition that the zero-norm and 1-norm solution coincide. Based on this condition the superiority can be made plausible.

Extracting Coactivated Features from Multiple Data Sets

Michael Gutmann and Aapo Hyvärinen

We present a nonlinear generalization of Canonical Correlation Analysis (CCA) to find related structure in multiple data sets. The new method allows to analyze an arbitrary number of data sets, and the extracted features capture higher-order statistical dependencies. The features are independent components that are coupled across the data sets. The coupling takes the form of coactivation (dependencies of variances). We validate the new method on artificial data, and apply it to natural images and brain imaging data.

Single Layer Complex Valued Neural Network with Entropic Cost Function

Luís Alexandre

This paper presents the adaptation of a single layer complex valued neural network (NN) to use entropy in the cost function instead of the usual mean squared error (MSE). This network has the good property of having only one layer so that there is no need to search for the number of hidden layer neurons: the topology is completely determined by the problem. We extend the existing stochastic MSE based learning algorithm to a batch MSE version first and then to a batch minimum error entropy (MEE). We present experiments showing the the proposed algorithms are competitive with other learning machines.

Non-Linear Projection

Batch Intrinsic Plasticity for Extreme Learning Machines

Klaus Neumann and Jochen Steil

Extreme learning machines are single-hidden layer feed-forward neural networks, where the training is restricted to the output weights in order to achieve fast learning with good performance. The success of learning strongly depends on the random parameter initialization. To overcome the problem of unsuited initialization ranges, a novel and efficient pretraining method to adapt extreme learning machines task-specific is presented. The pretraining aims at desired output distributions of the hidden neurons. It leads to better performance and less dependence on the size of the hidden layer.

An Empirical Study on the Performance of Spectral Manifold Learning Techniques

Peter Mysling, Søren Hauberg and Kim Pedersen

In recent years, there has been a surge of interest in spectral manifold learning techniques. Despite the interest, only little work has focused on the empirical behavior of these techniques. We construct synthetic data of variable complexity and observe the performance of the techniques as they are subjected to increasingly difficult problems. We evaluate performance in terms of both a classification and a regression task. Our study includes Isomap, LLE, Laplacian eigenmaps, and diffusion maps. Among others, our results indicate that the techniques are highly dependent on data density, sensitive to scaling, and greatly influenced by intrinsic dimensionality.

Semi-supervised Learning for WLAN Positioning

Teemu Pulkkinen, Teemu Roos and Petri Myllymäki

Currently the most accurate WLAN positioning systems are based on the fingerprinting approach, where a "radio map" is constructed by modeling how the signal strength measurements vary according to the location. However, collecting a sufficient amount of location-tagged training data is a rather tedious and time consuming task, especially in indoor scenarios — the main application area of WLAN positioning — where GPS coverage is unavailable. To alleviate this problem, we present a semi-supervised manifold learning technique for building accurate radio maps from partially labeled data, where only a small portion of the signal strength measurements need to be tagged with the corresponding coordinates. The basic idea is to construct a non-linear projection that maps high-dimensional signal fingerprints onto a two-dimensional manifold, thereby dramatically reducing the need of location-tagged data. Our results from a deployment in a real-world experiment demonstrate the practical utility of the method.

Pattern Recognition 2

Ensemble-Teacher Learning through a Perceptron Rule with a Margin

Kazuyuki Hara and Seiji Miyoshi

In ensemble-teacher learning, a student learns from a quasi-optimal-teacher selected randomly from a pool of many quasi-optimal-teachers, and the student performs better than the quasi-optimal teachers after the learning. The student performance is improved by using many quasi-optimal-teachers when a Hebbian rule is used. However, a perceptron rule cannot improve the student performance. We previously proposed a novel ensemble-teacher learning using a perceptron rule with a margin. A perceptron rule with a margin is mid-way between a Hebbian rule and a perceptron rule. We have found through computer simulation that a perceptron rule with a margin can improve student performance. In this paper, we provide theoretical support to the proposed method by using statistical mechanics methods.

Topic-dependent Document Ranking: Citation Network Analysis by Analogy to Memory Retrieval in the Brain

Hiroshi Okamoto

We propose a method of citation analysis for evaluating the topic-dependent importance of individual scientific papers. This method assumes spreading activation in citation networks with a multi-hysteretic input/output relationship for each node (paper). The multi-hysteretic property renders the steady state of spreading activation continuously dependent on the initial state. Given a topic represented by the initial state, the importance of individual papers can be defined by the activities they have in the steady state. We have devised this method inspired by memory retrieval in the brain, where the multi-hysteretic property of single cells or neuronal networks is considered to play an essential role for cue-dependent retrieval of memory. Quantitative evaluation using a restoration problem has revealed that the performance of the proposed method is considerably higher than that of the benchmark method. We demonstrate the practical usefulness of the proposed method by applying it to a citation network of neuroscience papers.

PADDLE: Proximal Algorithm for Dual Dictionaries LEarning

Curzio Basso, Matteo Santoro, Alessandro Verri and Silvia Villa

Recently, considerable research efforts have been devoted to the design of methods to learn from data overcomplete dictionaries for sparse coding. However, learned dictionaries require the solution of a sparse approximation problem for coding new data. In order to overcome this drawback, we propose an algorithm aimed at learning both a dictionary and its dual: a linear mapping directly performing the coding. Our algorithm is based on proximal methods and jointly minimizes the reconstruction error of the dictionary and the coding error of its dual; the sparsity of the representation is induced by an ℓ_1-based penalty on its coefficients. Experimental results show that the algorithm is capable of recovering the expected dictionaries. Furthermore, on a benchmark dataset the image features obtained from the dual matrix yield state-of-the-art classification performance while being much less computational intensive.

Posters

A Markov Random Field Approach to Neural Encoding and Decoding

Marcel van Gerven, Eric Maris and Tom Heskes

We introduce a new approach to neural encoding and decoding which makes use of sparse regression and Markov random fields. We show that interesting response functions were estimated from neuroimaging data acquired while a subject was watching checkerboard patterns and geometrical figures. Furthermore, we demonstrate that reconstructions of the original stimuli can be generated by loopy belief propagation in a Markov random field.

Weakly Supervised Learning of Foreground-Background Segmentation using Masked RBMs

Nicolas Heess, Nicolas Le Roux and John Winn

We propose an extension of the Restricted Boltzmann Machine (RBM) that allows the joint shape and appearance of foreground objects in cluttered images to be modeled independently of the background. We present a learning scheme that learns this representation directly from cluttered images with only very weak supervision. The model generates plausible samples and performs foreground-background segmentation. We demonstrate that representing foreground objects independently of the background can be beneficial in recognition tasks.

Recursive Multi-Way PLS for Adaptive Calibration of Brain Computer Interface System

Andrey Eliseyev, Alim-Louis Benabid and Tatiana Aksenova

In the present article a Recursive Multi-Way PLS algorithm for adaptive calibration of a BCI system is proposed. It combines the NPLS tensors decomposition with a scheme of recursive calculation. This Recursive algorithm allows treating data arrays of huge dimension. In addition, adaptive calibration provides a fast adjustment of the BCI system to mild changes of the signal. The proposed algorithm was validated on artificial and real data sets. In comparison to generic Multi-Way PLS, the recursive algorithm demonstrates good performance and robustness.

Transformation of Edge Weights in a Graph Bipartitioning Problem

Iakov Karandashev and Boris Kryzhanovsky

In this paper we consider the problem of partitioning a graph into two parts of equal sizes with minimal sum of edge weights between them. It is known that this problem is NP-complete and can be reduced to the minimization of a quadratic binary functional with constraints. In previous work it was shown that raising the matrix of couplings to some power leads to a significant increase of the basin of attraction of the deepest functional minima. This means that such transformation possesses great optimizing abilities. In this paper we show that in spite of the constraints present in the graph bipartitioning problem, the proposed matrix transformation approach works very well with this problem.

A Distributed Behavioral Model using Neural Fields

Mohamed Oubbati, Josef Frick and Günther Palm

We investigate the use of neural fields for building a distributed behavioral model enabling several agents to move in a flock. No leader is required, and each agent is implemented as an independent element that follows its own behavioral model which is composed of four steering behaviors: *separation*, *cohesion*, *alignment* and *obstacle avoidance*. The synchronized motion of the flock emerges from combination of those behaviors. The control design will be discussed in theoretical terms, supported by simulation results.

A Hypothetical Free Synaptic Energy Function and Related States of Synchrony

Karim El-Laithy and Martin Bogdan

A simple hypothetical energy function is proposed for a dynamic synaptic model. It is an approach based on the theoretical thermodynamic principles that are conceptually similar to the Hopfield ones. We show that using this approach a synapse exposes stable operating points in terms of its excitatory postsynaptic potential (EPSP) as a function of its synaptic strength. We postulate that synapses in a network operating at these stable points can drive this network to an internal state of synchronous firing. The presented analysis is related to the widely investigated temporal coherent activities (cell assemblies) over a certain range of time scales (binding-by-synchrony). The results illustrate that a synaptic dynamical model has more than one stable operating point regarding the postsynaptic energy transfer. This proposes a novel explanation of the observed synchronous activities within networks regarding the synaptic (coupling) functionality.

Observational Learning Based on Models of Overlapping Pathways

Emmanouil Hourdakis and Panos Trahanias

Brain imaging studies in macaque monkeys have recently shown that the observation and execution of specific types of grasp actions activate the same regions in the parietal, primary motor and somatosensory lobes. In the present paper we consider how learning via observation can be implemented in an artificial agent based on the above overlapping pathway of activations. We demonstrate that the circuitry developed for action execution can be activated during observation, if the agent is able to perform action association, i.e. relate its own actions with the ones of the demonstrator. In addition, by designing the model to activate the same neural codes during execution and observation, we show how the agent can accomplish observational learning of novel objects.

On the Capacity of Transient Internal States in Liquid-State Machines

Karim El-Laithy and Martin Bogdan

Liquid-state machines (LSM) represent a class of neural networks that are able to introduce multitasking by implicit representation of input information over the entire network components. How exactly the input information can be represented and how the computations are accomplished, stay however unresolved. In order to tackle this issue, we demonstrate how LSM can process different input information as a varying set of transiently stable states of collective activity. This is performed by adopting a relatively complex dynamic synaptic model. Some light is shed on the relevance of the usage of the developed framework to mimic complex cortical functions, e.g. content-addressable memory.

Hybrid Parallel Classifiers for Semantic Subspace Learning

Nandita Tripathi, Michael Oakes and Stefan Wermter

Subspace learning is very important in today's world of information overload. Distinguishing between categories within a subset of a large data repository such as the web and the ability to do so in real time is critical for a successful search technique. The characteristics of data belonging to different domains are also varying widely. This merits the need for an architecture which caters to the differing characteristics of different data domains. In this paper we present a novel hybrid parallel architecture using different types of classifiers trained on different subspaces. We further compare the performance of our hybrid architecture with a single classifier and show that it outperforms the single classifier system by a large margin when tested with a variety of hybrid combinations. Our results show that subspace classification accuracy is boosted and learning time reduced significantly with this new hybrid architecture.

Temperature Prediction in Electric Arc Furnace with Neural Network Tree

Miroslaw Kordos, Marcin Blachnik and Tadeusz Wieczorek

This paper presents a neural network tree regression system with dynamic optimization of input variable transformations and post-training optimization. The decision tree consists of MLP neural networks, which optimize the split points and at the leaf level predict final outputs. The system is designed for regression problems of big and complex datasets. It was applied to the problem of steel temperature prediction in the electric arc furnace in order to decrease the process duration at one of the steelworks.

Optimizing Linear Discriminant Error Correcting Output Codes Using Particle Swarm Optimization

Dimitrios Bouzas, Nikolaos Arvanitopoulos and Anastasios Tefas

Error Correcting Output Codes reveal an efficient strategy in dealing with multi-class classification problems. According to this technique, a multi-class problem is decomposed into several binary ones. On these created sub-problems we apply binary classifiers and then, by combining the acquired solutions, we are able to solve the initial multi-class problem. In this paper we consider the optimization of the Linear Discriminant Error Correcting Output Codes framework using Particle Swarm Optimization. In particular, we apply the Particle Swarm Optimization algorithm in order to optimally select the free parameters that control the split of the initial problem's classes into sub-classes. Moreover, by using the Support Vector Machine as classifier we can additionally apply the Particle Swarm Optimization algorithm to tune its free parameters. Our experimental results show that by applying Particle Swarm Optimization on the Sub-class Linear Discriminant Error Correcting Output Codes framework we get a significant improvement in the classification performance.

SOS-HMM: Self-Organizing Structure of Hidden Markov Model

Rakia Jaziri, Mustapha Lebbah, Younès Bennani and Jean-Hugues Chenot

We propose in this paper a novel approach which makes self-organizing maps (SOM) and the Hidden Markov Models (HMMs) cooperate. Our approach (SOS-HMM: Self Organizing Structure of HMM) allows to learn the Hidden Markov Models topology. The main contribution for the proposed approach is to automatically extract the structure of a hidden Markov model without any prior knowledge of the application domain. This model can be represented as a graph of macro-states, where each state represents a micro model. Experimental results illustrate the advantages of the proposed approach compared to a fixed structure approach.

Image Receptive Fields Neural Networks for Object Recognition

Paméla Daum, Jean-Luc Buessler and Jean-Philippe Urban

This paper extends a recent and very appealing approach of computational learning to the field of image analysis. Recent works have demonstrated that the implementation of Artificial Neural Networks (ANN) could be simplified by using a large amount of neurons with random weights. Only the output weights are adapted, with a single linear regression. Supervised learning is very fast and efficient. To adapt this approach to image analysis, the novelty is to initialize weights, not as independent random variables, but as Gaussian functions with only a few random parameters. This creates smooth random receptive fields in the image space. These *Image Receptive Fields - Neural Networks* (IRF-NN) show remarkable performances for recognition applications, with extremely fast learning, and can be applied directly to images without pre-processing.

A Comparison of the Electric Potential through the Membranes of Ganglion Neurons and Neuroblastoma Cells

Thiago Pinto, Roseli Wedemann and Célia Cortez

We have modeled the electric potential profile, across the membranes of the ganglion neuron and neuroblastoma cells. We considered the resting and action potential states, and analyzed the influence of fixed charges of the membrane on the electric potential of the surface of the membranes of these cells, based on experimental values of membrane properties. The ganglion neuron portrays a healthy neuron, and the neuroblastoma cell, which is tumorous, represents a pathologic neuron. We numerically solved the non-linear Poisson-Boltzmann equation, by considering the densities of charges dissolved in an electrolytic solution and fixed on both glycocalyx and cytoplasmic proteins. We found important differences among the potential profiles of the two cells.

SNPboost: Interaction Analysis and Risk Prediction on GWA Data

Ingrid Brænne, Jeanette Erdmann and Amir Madany Mamlouk

Genome-wide association (GWA) studies, which typically aim to identify single nucleotide polymorphisms (SNPs) associated with a disease, yield large amounts of high-dimensional data. GWA studies have been successful in identifying single SNPs associated with complex diseases. However, so far, most of the identified associations do only have a limited impact on risk prediction. Recent studies applying SVMs have been successful in improving the risk prediction for Type I and II diabetes, however, a drawback is the poor interpretability of the classifier. Training the SVM only on a subset of SNPs would imply a preselection, typically by the p-values. Especially for complex diseases, this might not be the optimal selection strategy. In this work, we propose an extension of Adaboost for GWA data, the so-called SNPboost. In order to improve classification, SNPboost successively selects a subset of SNPs. On real GWA data (German MI family study II), SNPboost outperformed linear SVM and further improved the performance of a non-linear SVM when used as a preselector. Finally, we motivate that the selected SNPs can be put into a biological context.

Binary Patterns Identification by Vector Neural Network with Measure of Proximity Between Neuron States

Vladimir Kryzhanovsky

I describe a new vector neural network, in which a priori information about the distribution of noise is easily and naturally embedded. Taking into ac-count the noise distribution allows to essentially increase the system noise immunity. A measure of proximity between neuron states is embedded for the first time. It makes possible to use the prior information. On binary identification problem the one order increase of storage capacity is shown.

Emerging Bayesian Priors in a Self-Organizing Recurrent Network

Andreea Lazar, Gordon Pipa and Jochen Triesch

We explore the role of local plasticity rules in learning statistical priors in a self-organizing recurrent neural network (SORN). The network receives input sequences composed of different symbols and learns the structure embedded in these sequences via a simple spike-timing-dependent plasticity rule, while synaptic normalization and intrinsic plasticity maintain a low level of activity. After learning, the network exhibits spontaneous activity that matches the stimulus-evoked activity during training and thus can be interpreted as samples from the network's prior probability distribution over evoked activity states. Further, we show how learning the frequency and spatio-temporal characteristics of the input sequences influences network performance in several classification tasks. These results suggest a novel connection between low level learning mechanisms in recurrent networks and high level concepts of statistical inference.

Momentum Acceleration of Least–Squares Support Vector Machines

Jorge López, Álvaro Barbero and José Dorronsoro

Least–Squares Support Vector Machines (LS–SVMs) have been a successful alternative model for classification and regression Support Vector Machines (SVMs), and used in a wide range of applications. In spite of this, only a limited effort has been realized to design efficient algorithms for the training of this class of models, in clear contrast to the vast amount of contributions of this kind in the field of classic SVMs. In this work we propose to combine the popular Sequential Minimal Optimization (SMO) method with a momentum strategy that manages to reduce the number of iterations required for convergence, while requiring little additional computational effort per iteration, especially in those situations where the standard SMO algorithm for LS–SVMs fails to obtain fast solutions.

Fast Support Vector Training by Newton's Method

Shigeo Abe

We discuss a fast training method of support vector machines using Newton's method combined with fixed-size chunking. To speed up training, we limit the number of upper or lower bounded variables in the working set to two so that the corrections of the variables do not violate the bounding conditions. If similar working sets occur alternately, we merge these two working sets into one, and if similar working sets occur consecutively, we use incremental Cholesky factorization in calculating corrections. By computer experiments, we show that the proposed method is comparable to or faster than SMO (Sequential minimum optimization) using the second order information.

Linear Operators and Stochastic Partial Differential Equations in Gaussian Process Regression

Simo Särkkä

In this paper we shall discuss an extension to Gaussian process (GP) regression models, where the measurements are modeled as linear functionals of the underlying GP and the estimation objective is a general linear operator of the process. We shall show how this framework can be used for modeling physical processes involved in measurement of the GP and for encoding physical prior information into regression models in form of stochastic partial differential equations (SPDE). We shall also illustrate the practical applicability of the theory in a simulated application.

Learning from Multiple Annotators with Gaussian Processes

Perry Groot, Adriana Birlutiu and Tom Heskes

In many supervised learning tasks it can be costly or infeasible to obtain objective, reliable labels. We may, however, be able to obtain a large number of subjective, possibly noisy, labels from multiple annotators. Typically, annotators have different levels of expertise (i.e., novice, expert) and there is considerable diagreement among annotators. We present a Gaussian process (GP) approach to regression with multiple labels but no absolute gold standard. The GP framework provides a principled non-parametric framework that can automatically estimate the reliability of individual annotators from data without the need of prior knowledge. Experimental results show that the proposed GP multi-annotator model outperforms models that either average the training data or weigh individually learned single-annotator models.

Estimation of the Number of Clusters Using Heterogeneous Multiple Classifier System

Omar Ayad, Moamar Sayed-Mouchaweh and Patrice Billaudel

Assessing the number of clusters of statistical populations is a challenging problem in unsupervised learning. In this paper, we propose to overcome this problem by estimating the number of clusters using a novel clustering ensemble scheme. This one combines clustering and classification methods in order to increase the clustering performances. In the first time, the proposed approach divides the patterns into stable and ambiguous sets. The stable set gathers the patterns belonging to one cluster while the ambiguous set corresponds to ambiguous patterns located between different clusters. To detect the appropriate number of clusters, the proposed approach ignores ambiguous patterns and preserves the stable set as good "prototypes". Then, the different partitions obtained from the stable set are evaluated by several cluster validation criteria. Finally, the patterns of the unstable set are assigned to the obtained clusters by supervised classifier.

A Distributed Self-adaptive Nonparametric Change-Detection Test for Sensor/Actuator Networks

Cesare Alippi, Giacomo Boracchi and Manuel Roveri

The prompt detection of faults and, more in general, changes in stationarity in networked systems such as sensor/actuator networks is a key issue to guarantee robustness and adaptability in applications working in real-life environments. Traditional change-detection methods aiming at assessing the stationary of a data generating process would require a centralized availability of all observations, solution clearly unacceptable when large scale networks are considered and data have local interest. Differently, distributed solutions based on decentralized change-detection tests exploiting information at the unit and cluster level would be a solution. This work suggests a novel distributed change-detection test which operates at two-levels: the first, running on the unit, is particularly reactive in detecting small changes in the process generating the data, whereas the second exploits distributed information at the cluster-level to reduce false positives. Results can be immediately integrated in the machine learning community where adaptive solutions are envisaged to address changes in stationarity of the considered application. A large experimental campaign shows the effectiveness of the approach both on synthetic and real data applications.

Weighted Mutual Information for Feature Selection

Erik Schaffernicht and Horst-Michael Gross

In this paper, we apply weighted Mutual Information for effective feature selection. The presented hybrid filter wrapper approach resembles the well known AdaBoost algorithm by focusing on those samples that are not classified or approximated correctly using the selected features. Redundancies and bias of the employed learning machine are handled implicitly by our approach. In experiments, we compare the weighted Mutual Information algorithm with other basic approaches for feature subset selection that use similar selection criteria. The efficiency and effectiveness of our method are demonstrated by the obtained results.

Face Prediction from fMRI Data During Movie Stimulus: Strategies for Feature Selection

Jukka-Pekka Kauppi, Heikki Huttunen, Heikki Korkala, Iiro Jääskeläinen, Mikko Sams and Jussi Tohka

We investigate the suitability of the multi-voxel pattern analysis approach to analyze diverse movie stimulus functional magnetic resonance imaging (fMRI) data. We focus on predicting the presence of faces in the drama movie based on the fMRI measurements of 12 subjects watching the movie. We pose the prediction as a regression problem where regression coefficients estimated from the training data are used to estimate the presence of faces in the stimulus for the test data. Because the number of features (voxels) exceeds the number of training samples, an emphasis is placed on the feature selection. We compare four automatic feature selection approaches. The best results were achieved by sparse regression models. The correlations between the face presence time-course predicted from fMRI data and manual face annotations were in the range from 0.43 to 0.62 depending on the subject and pre-processing options, i.e., the prediction was successful. This suggests that proposed methods are useful in testing novel research hypotheses with natural stimulus fMRI data.

The Authentication System for Multi-Modal Behavior Biometrics using Concurrent Pareto Learning SOM

Hiroshi Dozono, Shinsuke Ito and Masanori Nakakuni

We have proposed the integration of behavior biometrics using Supervised Pareto learning SOM to improve the accuracy of authentication. For small systems such as mobile devices, this method may be heavy, because of the memory usage or computational power. In this paper, we propose the application of Concurrent Pareto learning SOM, which uses a small map for each user. The performance of this method is confirmed by authentication experiments using behavior biometrics of keystroke timings and key typing sounds.

Hermite Polynomials and Measures of Non-Gaussianity

Jouni Puuronen and Aapo Hyvärinen

We first review some rigorous properties of the Hermite polynomials, and demonstrate their usefulness in estimating probability distributions as series from data samples. We then proceed to explain how these series can be used to obtain precise and robust measures of non-Gaussianity. Our measures of non-Gaussianity detect all kinds of deviations from Gaussianity, and thus provide reliable objective functions for ICA. With a linear computational complexity with respect to the sample size, our method is also suitable for large data sets.

Complex-valued Independent Component Analysis of Natural Images

Valero Laparra, Michael Gutmann, Jesús Malo and Aapo Hyvärinen

Linear independent component analysis (ICA) learns simple cell receptive fields from natural images. Here, we show that linear complex-valued ICA learns complex cell properties from Fourier-transformed natural images, i.e. two Gabor-like filters with quadrature-phase relationship. Conventional methods for complex-valued ICA assume that the phases of the output signals have uniform distribution. We show here that for natural images the phase distributions are, however, often far from uniform. We thus relax the uniformity assumption and model also the phase of the sources in complex-valued ICA. Compared to the original complex ICA model, the new model provides a better fit to the data, and leads to Gabor filters of qualitatively different shape.

Improving Gaussian Process Value Function Approximation in Policy Gradient Algorithms

Hunor Jakab and Lehel Csató

The use of value-function approximation in reinforcement learning (RL) problems is widely studied, the most common application of it being the extension of value-based RL methods to continuous domains. We combine it with Williams' episodic REINFORCE algorithm to reduce the variance of the gradient estimates. A significant computational overload of the algorithm is caused by the need to completely re-estimate the value-function after each gradient update step. To overcome this problem we propose a measure composed of a KL distance–based score and a time dependent factor to exchange obsolete basis vectors with newly acquired measurements. This method leads to a more stable estimation of the action value-function and also reduces gradient variance.

Application of Nonlinear Neural Network Model for Self Sensing Characteristic in an Ionic Polymer Metal Composite (IPMC) Actuator

Ngoc Chi Nam Doan, Kyoung Kwan Ahn, Quang Truong Dinh and Jong Il Yoon

This paper focuses on a self sensing characteristic of an Ionic Polymer Metal Composite (IPMC) bases on a novel accurate nonlinear black-box model (NBBM) to estimate the IPMC tip displacement. The NBBM is formed by a recurrent multi-layer perceptron neural network (RMLPNN) and a self-adjustable learning mechanism (SALM). The model parameters are optimized by using a set of training data. The ability of NBBM model is evaluated by a comparison of the estimated and real IPMC bending characteristic.

Optimal Control Using Functional Type SIRMs Fuzzy Reasoning Method

Takashi Mitsuishi and Yasunari Shidama

A mathematical framework for studying a fuzzy optimal control using functional type SIRMs reasoning method is discussed in this paper. The existence of SIRMs which minimize the cost function of fuzzy control system is proved with continuity of approximate reasoning and topological property of the set of membership functions in SIRMs.

High-Dimensional Surveillance

Saylisse Dávila, George Runger and Eugene Tuv

Many systems (manufacturing, environmental, health, etc.) generate counts (or rates) of events that are monitored to detect changes. Modern data complements event counts with many additional measurements (such as geographic, demographic, and others) that comprise high-dimensional attributes. This leads to an important challenge to detect a change that only occurs within a region, initially unspecified, defined by these attributes and current methods to handle the attribute information are challenged by high-dimensional data. Our approach transforms the problem to supervised learning, so that properties of an appropriate learner can be described. Rather than error rates, we generate a signal (of a system change) from an appropriate feature selection algorithm. A measure of statistical significance is included to control false alarms. Results on simulated examples are provided.

A One-layer Dual Recurrent Neural Network with a Heaviside Step Activation Function for Linear Programming with Its Linear Assignment Application

Qingshan Liu and Jun Wang

This paper presents a one-layer recurrent neural network for solving linear programming problems. The proposed neural network is guaranteed to be globally convergent in finite time to the optimal solutions under a mild condition on a derived lower bound of a single gain parameter. The number of neurons in the neural network is the same as the number of decision variables of the dual optimization problem. Compared with the existing neural networks for linear programming, the proposed neural network has salient features such as finite-time convergence and lower model complexity. Specifically, the proposed neural network is tailored for solving the linear assignment problem with simulation results to demonstrate the effectiveness and characteristics of the proposed neural network.

Neural Network Solution of Optimal Control Problem with Control and State Constraints

Tibor Kmet

A neural network based optimal control synthesis is presented for solving optimal control problems with control and state constraints. The optimal control problem is transcribed into nonlinear programming problem which is implemented with adaptive critic neural network. The proposed simulation methods is illustrated by the optimal control problem of feeding adaptation of filter feeders of Daphnia. Results show that adaptive critic based systematic approach holds promise for obtaining the optimal control with control and state constraints.

Singular Perturbation Approach with Matsuoka Oscillator and Synchronization Phenomena

Yasuomi Sato, Kazuki Nakada and Kiyotoshi Matsuoka

We study the singular perturbation approach in a pair of Matsuoka nonlinear neural oscillators, which consist of membrane potential (v) and recovery (u) dynamics with a relaxation rate (P). This shows that the u-coupled system of the Matsuoka oscillators would be valid for the modeling of neural firings. The coupled integrate-and-fire model of the improved type with an impulse-like interval results from the u-coupled system, under taking the limit of $P \to \infty$, without loss of any coupling properties. We simulate systematically synchronization of both the v-coupled and u-coupled systems. We also discuss potential capabilities of the u-coupled system of Matsuoka oscillators.

A RANSAC-based ISOMAP for Filiform Manifolds in Nonlinear Dynamical Systems

Hiromichi Suetani and Shotaro Akaho

Trajectories generated from a chaotic dynamical system are lying on a nonlinear manifold in the state space. Even if the dimensionality of such a manifold is much lower than that of the full state space, we need many state variables to trace a motion on it as far as we remain to employ the original coordinate, so the resulting expression of the dynamics becomes redundant. In the present study, we employ one of the manifold learning algorithms, ISOMAP, to construct a new nonlinear coordinate that globally covers the manifold, which enables us to describe the dynamics on it as a low-dimensional dynamical system. Here, in order to improve the conventional ISOMAP, we propose an approach based on a combination with RANSAC for pruning the misconnected edges in the neighboring graph. We show that a clear deterministic relationship is extracted from time series of a mass-spring model for the chaotic dripping faucet using the proposed method.

Manifold Learning for Visualization of Vibrational States of a Rotating Machine

Ignacio Díaz, Abel Cuadrado, Alberto Diez and Manuel Domínguez

This paper describes a procedure based on the use of manifold learning algorithms to visualize periodic –or nearly periodic– time series produced by processes with different underlying dynamics. The proposed approach is done in two steps: a feature extraction stage, where a set of descriptors in the frequency domain is extracted, and a manifold learning stage that finds low dimensional structures in the feature space and obtains projections on a low dimensional space for visualization. This approach is applied on vibration data of an electromechanical rotating machine to visualize different vibration conditions under two kinds of asymmetries, using four state-of-the-art manifold learning algorithms for comparison purposes. In all cases, the methods yield consistent results and produce insightful visualizations, suggesting future developments and application in engineering problems.

Bias of Importance Measures for Multi-valued Attributes and Solutions

Houtao Deng, George Runger and Eugene Tuv

Attribute importance measures for supervised learning are important for improving both learning accuracy and interpretability. However, it is well-known there could be bias when the predictor attributes have different numbers of values. We propose two methods to solve the bias problem. One uses an out-of-bag sampling method called OOBForest and one, based on the new concept of a partial permutation test, is called pForest. The existing research has considered the bias problem only among irrelevant attributes and equally informative attributes, while we compare to existing methods in a situation where unequally informative attributes (with or without interactions) and irrelevant attributes co-exist. We observe that the existing methods are not always reliable for multi-valued predictors, while the proposed methods compare favorably in our experiments.

A Computationally Efficient Information Estimator for Weighted Data

Hideitsu Hino and Noboru Murata

The Shannon information content is a fundamental quantity and it is of great importance to estimate it from observed dataset in the field of statistics, information theory, and machine learning. In this study, an estimator for the information content using a given set of weighted data is proposed. The empirical data distribution varies depending on the weight. The notable features of the proposed estimator are its computational efficiency and its ability to deal with weighted data. The proposed estimator is extended in order to estimate cross entropy, entropy and KL divergence with weighted data. Then, the estimators are applied to classification with one-class samples, and distribution preserving data compression problems.

Top-down Induction of Reduced Ordered Decision Diagrams from Neural Networks

Jan Chorowski and Jacek Zurada

Neural networks offer good generalization performance, noise robustness, and model complexity control. However, neural network mappings are expressed in terms of complicated mathematical functions that are inherently hard to understand. To overcome this limitation rule extraction methods have been proposed. This paper presents a novel method of rule extraction which recursively, in a top-down manner, builds a Reduced Ordered Decision Diagram. The diagram structure allows sharing of nodes, which partially overcomes two problems present in Decision Tree-based rule extraction – the problem of subtree replication and of training set fragmentation. A method for reducing the rule search space by identifying regions in which the network shows similar behavior is presented. Preliminary results of the method performance are reported.

A Framework for Application-Oriented Design of Large-Scale Neural Networks

David Bouchain, Florian Hauser and Günther Palm

Tools for simulations of neural networks exist aplenty. They range from simulators for detailed multi-compartment neurons, over packages for precise reconstruction of small biological networks, to simulators for large-scale networks with stochastic connectivity properties. However, no frameworks for constructing large-scale, dedicated networks exist. Based on the design principles used for our previous work, we introduce a C++ framework which is specifically tailored to simplify the construction of large networks with specific cognitive functionalities.

A Dynamic Field Model of Ordinal and Timing Properties of Sequential Events

Flora Ferreira, Wolfram Erlhagen and Estela Bicho

Recent evidence suggests that the neural mechanisms underlying memory for serial order and interval timing of sequential events are closely linked. We present a dynamic neural field model which exploits the existence and stability of multi-bump solutions with a gradient of activation to store serial order. The activation gradient is achieved by applying a state-dependent threshold accommodation process to the firing rate function. A field dynamics of lateral inhibition type is used in combination with a dynamics for the baseline activity to recall the sequence from memory. We show that depending on the time scale of the baseline dynamics the precise temporal structure of the original sequence may be retrieved or a proactive timing of events may be achieved.

Robot Trajectory Prediction and Recognition based on a Computational Mirror Neurons Model

Junpei Zhong, Cornelius Weber and Stefan Wermter

Model Mirror neurons are premotor neurons that are considered to play a role in goal-directed actions, action understanding and even social cognition. As one of the promising research areas in psychology, cognitive neuroscience and cognitive physiology, understanding mirror neurons in a social cognition context, whether with neural or computational models, is still an open issue. In this paper, we mainly focus on the action understanding aspect of mirror neurons, which can be regarded as a fundamental function of social cooperation and social cognition. Our proposed initial architecture is to learn a simulation of the walking pattern of a humanoid robot and to predict where the robot is heading on the basis of its previous walking trajectory.

A Sentence Generation Network that Learns Surface and Abstract Syntactic Structures

Martin Takac, Luba Benuskova and Alistair Knott

In this paper we present a connectionist model of sentence generation based on the novel idea that sentence meanings are represented in the brain as sequences of sensorimotor signals which are replayed during sentence generation. Our model can learn surface patterns in language as well as abstract word-ordering conventions. The former is achieved by a recurrent network module; the latter by a feed-forward network that learns to inhibit overt pronunciation of predicted words in certain phases of sensorimotor sequence rehearsal. Another novel element of the model is adaptive switching of control based on uncertainty (entropy) of predicted word distributions. Experiments with the model show that it can learn the syntax, morphology and semantics of a target language and generalize well to unseen meanings/sentences.

A Perceptual Memory System for Affordance Learning in Humanoid Robots

Marc Kammer, Marko Tscherepanow, Thomas Schack and Yukie Nagai

Memory constitutes an essential cognitive capability of humans and animals. It allows them to act in very complex, non-stationary environments. In this paper, we propose a perceptual memory system, which is intended to be applied on a humanoid robot learning affordances. According to the properties of biological memory systems, it has been designed in such a way as to enable life-long learning without catastrophic forgetting. Based on clustering sensory information, a symbolic representation is derived automatically. In contrast to alternative approaches, our memory system does not rely on pre-trained models and works completely unsupervised.

Probabilistic Proactive Timeline Browser

Antti Ajanki and Samuel Kaski

We have developed a browser suitable for finding events from timelines, in particular from life logs and other timelines containing a familiar narrative. The system infers the relevance of events based on the user's browsing behavior and increases the visual saliency of relevant items along the timeline. As recognized images are strong memory cues, the user can quickly determine if the salient images are relevant and, if they are, it is quick and easy to select them by clicking since they are salient. Even if the inferred relevance was not correct, the timeline will help: The user may remember if the sought event was before or after a saliently shown event which limits the search space. A user study shows that the browser helps in locating relevant images quicker, and augmenting explicit click feedback with implicit mouse movement patterns further improves the performance.

Person Tracking based on a Hybrid Neural Probabilistic Model

Wenjie Yan, Cornelius Weber and Stefan Wermter

This article presents a novel approach for a real-time person tracking system based on particle filters that use different visual streams. Due to the difficulty of detecting a person from a top view, a new architecture is presented that integrates different vision streams by means of a Sigma-Pi network. A short-term memory mechanism enhances the tracking robustness. Experimental results show that robust real-time person tracking can be achieved.

Gaze- and Speech-Enhanced Content-Based Image Retrieval in Image Tagging

He Zhang, Teemu Ruokolainen, Jorma Laaksonen, Christina Hochleitner and Rudolf Traunmüller

We describe a setup and experiments where users are checking and correcting image tags given by an automatic tagging system. We study how much the application of a content-based image retrieval (CBIR) method speeds up the process of finding and correcting the erroneously-tagged images. We also analyze the use of implicit relevance feedback from the user's gaze tracking patterns as a method for boosting up the CBIR performance. Finally, we use automatic speech recognition for giving the correct tags for those images that were wrongly tagged. The experiments show a large variance in the tagging task performance, which we believe is primarily caused by the users' subjectivity in image contents as well as their varying familiarity with the gaze tracking and speech recognition setups. The results suggest potentials for gaze and/or speech enhanced CBIR method in image tagging, at least for some users.

Modelling Hypothetical Wage Equation by Neural Networks

Jaakko Talonen and Miki Sirola

In this paper, a hypothetical wage equation is modelled using quarterly data from United Kingdom. Wage and price data have a great importance for the overall features of large-scale macro models and for example for the different policy actions. The modelled feature in this paper is the real wage, the differential of a nominal wage and a price index. In the variable selection phase, the stationary properties of the data were investigated by augmented Dickey-Fuller tests (ADF). The main idea in this paper is to present a neural network model, which has a better fit than conventional MLR model.

On the Designing of Spikes Band-Pass Filters for FPGA

Manuel Domínguez-Morales, Angel Jimenez-Fernandez, Elena Cerezuela-Escudero, Rafael Paz-Vicente, Alejandro Linares-Barranco and Gabriel Jimenez

In this paper we present two implementations of spike-based band-pass filters, which are able to reject out-of-band frequency components in the spike domain. First one is based on the use of previously designed spike-based low-pass filters. With this architecture the quality factor, Q, is lower than 0.5. The second implementation is inspired in the analog multi-feedback filters (MFB) topology, it provides a higher than 1 Q factor, and ideally tends to infinite. These filters have been written in VHLD, and synthesized for FPGA. Two spike-based band-pass filters presented take advantages of the spike rate coded representation to perform a massively parallel processing without complex hardware units, like floating point arithmetic units, or a large memory. These low requirements of hardware allow the integration of a high number of filters inside a FPGA, allowing to process several spike coded signals fully in parallel.

An Information Geometrical View of Stationary Subspace Analysis

Motoaki Kawanabe, Wojciech Samek, Paul von Bünau and Frank Meinecke

Stationary Subspace Analysis (SSA) is an unsupervised learning method that finds subspaces in which data distributions stay invariant over time. It has been shown to be very useful for studying non-stationarities in various applications . In this paper, we present the first SSA algorithm based on a full generative model of the data. This new derivation relates SSA to previous work on finding interesting subspaces from high-dimensional data in a similar way as the three easy routes to independent component analysis, and provides an information geometric view.

Forecasting Road Condition after Maintenance Works by Linear Methods

Konsta Sirvio and Jaakko Hollmén

Forecasting road condition after maintenance can help in better road maintenance planning. As road administrations annually collect and store road-related data, data-driven methods can be used in determining forecasting models that result in improved accuracy. In this paper, we compare the prediction models identified by experts and currently used in road administration with simple data-driven prediction models, and parsimonious models based on a input selection algorithm. Furthermore, non-linear prediction using radial basis function networks is performed. We estimate and validate the prediction models with a database containing data of over two million road segments.

Multistart Strategy using Delta Test for Variable Selection

Dušan Sovilj

Proper selection of variables is necessary when dealing with large number of input dimensions in regression problems. In the paper, we investigate the behaviour of landscape that is formed when using Delta test as the optimization criterion. We show that simple and greedy Forward-backward selection procedure with multiple restarts gives optimal results for data sets with large number of samples. An improvement to multistart Forward-backward selection is presented that uses information from previous iterations in the form of long-term memory.

Speech Recognition Based on the Processing Solutions of Auditory Cortex

Patrick May and Hannu Tiitinen

Speech recognition in the human brain depends on spectral analysis coupled with temporal integration of auditory information. In primates, these processes are mirrored as selective responsiveness of neurons to species-specific vocalizations. Here, we used computational modeling of cortical neural networks to investigate how they achieve selectivity to speech stimuli. Stimulus material comprised multiple pseudowords. We found that synaptic depression was crucial for the emergence of neurons sensitive to the temporal structure of the stimuli. Further, the subdivision of the network into several parallel processing streams was needed for stimulus selectivity to occur. In general, stimulus selectivity and temporal integration seems to be supported by networks with high values of small-world connectivity. The current results might serve as a preliminary pointer for developing speech recognition solutions based on the neuroanatomy and -physiology of auditory cortex.

A Geometric Bio-Inspired Model for Recognition of Low-Level Structures

E. Ulises Moya-Sánchez and Eduardo Vázquez-Santacruz

A new bio-inspired model is proposed in this paper. This model mimetizes the simple cells of the mammalian visual processing system in order to recognize low-level geometric structures such as oriented lines, edges and other constructed with these. It takes advantage of geometric algebra in order to represent structures and symmetric operators by estimating the relation between geometric entities and encoding it. This geometric model uses symmetric relations in which exist a invariance under some transformation according to biological models. It is based on a Quaternionic Atomic Function and its phase information to detect oriented lines, edges and geometric structures defined by lines. Also, it uses a geometric neural network to encode the transformation between lines and then for classifying of geometric structures.

View-tuned Approximate Partial Matching Kernel from Hierarchical Growing Neural Gases

Marco Kortkamp and Sven Wachsmuth

The problem of comparing images or image regions can be considered as the problem of matching unordered sets of high dimensional visual features. We show that an hierarchical Growing Neural Gas (GNG) can robustly be used to approximate the optimal partial matching cost between vector sets. Further, we extend the unordered set matching, such that the matching of local features pays attention to the structure of the object and the relative positions of the parts. This view-tuning is also realized with hierarchical GNGs and yields an efficient Mercer Kernel.

ANGE – Automatic Neural Generator

Leonardo Reis, Luis Aguiar, Dar ıo Baptista and Fernando Morgado Dias

Artificial Neural Networks have a wide application in terms of research areas but they have never really lived up to the promise they seemed to be in the beginning of the 80s. One of the reasons for this is the lack of hardware for their implementation in a straightforward and simple way. This paper presents a tool to respond to this need: An Automatic Neural Generator. The generator allows a user to specify the number of bits used in each part of the neural network and programs the selected FPGA with the network. To measure the accuracy of the implementation an automatically built neural network was inserted in a control loop and compared with Matlab.

Uncertainty Sampling-Based Active Selection of Datasetoids for Meta-Learning

Ricardo Prudêncio, Carlos Soares and Teresa Ludermir

Several meta-learning approaches have been developed for the problem of algorithm selection. In this context, it is of central importance to collect a sufficient number of datasets to be used as meta-examples in order to provide reliable results. Recently, some proposals to generate datasets have addressed this issue with successful results. These proposals include datasetoids, which is a simple manipulation method to obtain new datasets from existing ones. However, the increase in the number of datasets raises another issue: in order to generate meta-examples for training, it is necessary to estimate the performance of the algorithms on the datasets. This typically requires running all candidate algorithms on all datasets, which is computationally very expensive. One approach to address this problem is the use of an active learning approach to meta-learning, termed active meta-learning. In this paper we investigate the combined use of an active meta-learning approach based on an uncertainty score and datasetoids. Based on our results, we conclude that the accuracy of our method is very good results with as little as 10% to 20% of the meta-examples labeled.

Learning Scheme for Complex Neural Networks Using Simultaneous Perturbation

Yutaka Maeda, Takahiro Yamada, Seiji Miyoshi and Hiroomi Hikawa

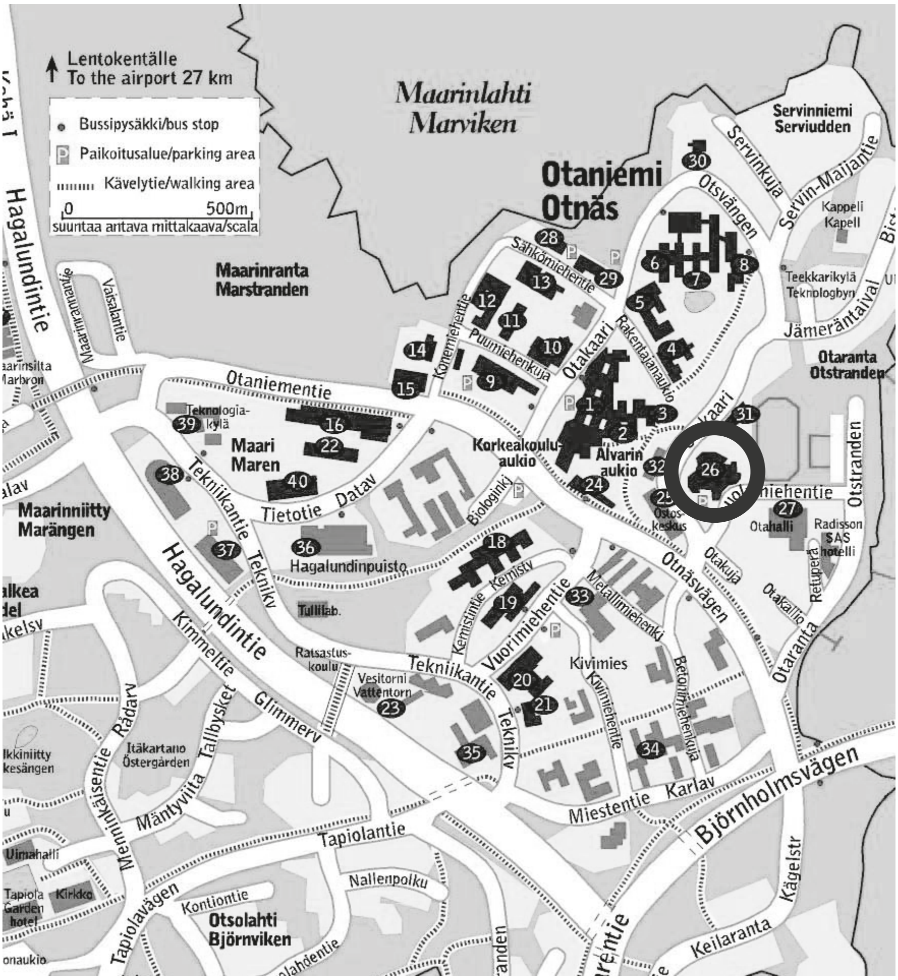
Usually, the back-propagation learning rule is widely used for complex-valued neural networks as well. On the other hand, in this paper, learning rule for complex-valued neural networks using the simultaneous perturbation optimization method is proposed. Comparison between the back-propagation method and the proposed simultaneous perturbation learning rule is made for some test problems. Simplicity of the proposed method results in faster learning speed.

Practical Information

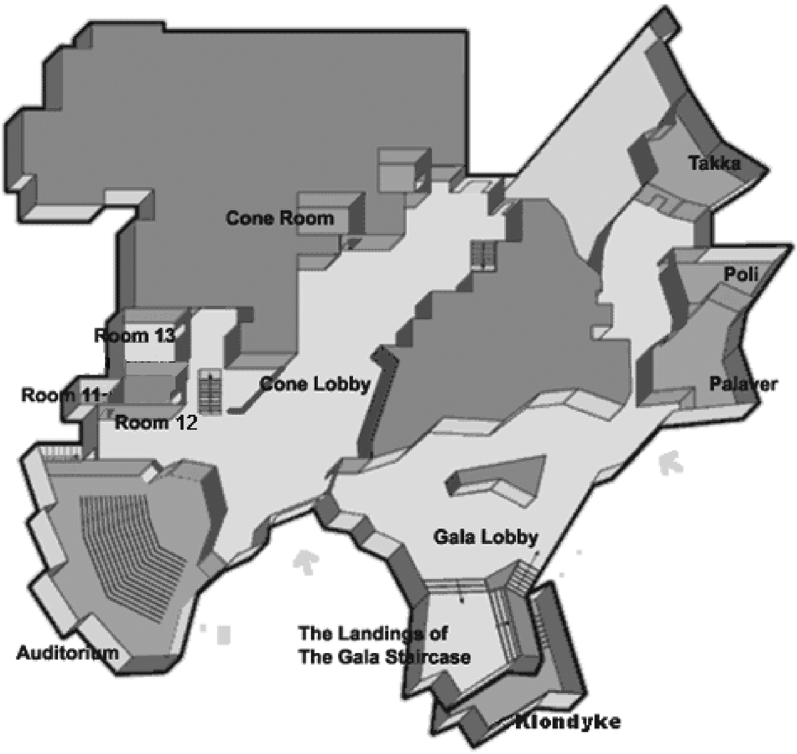
Otaniemi Campus

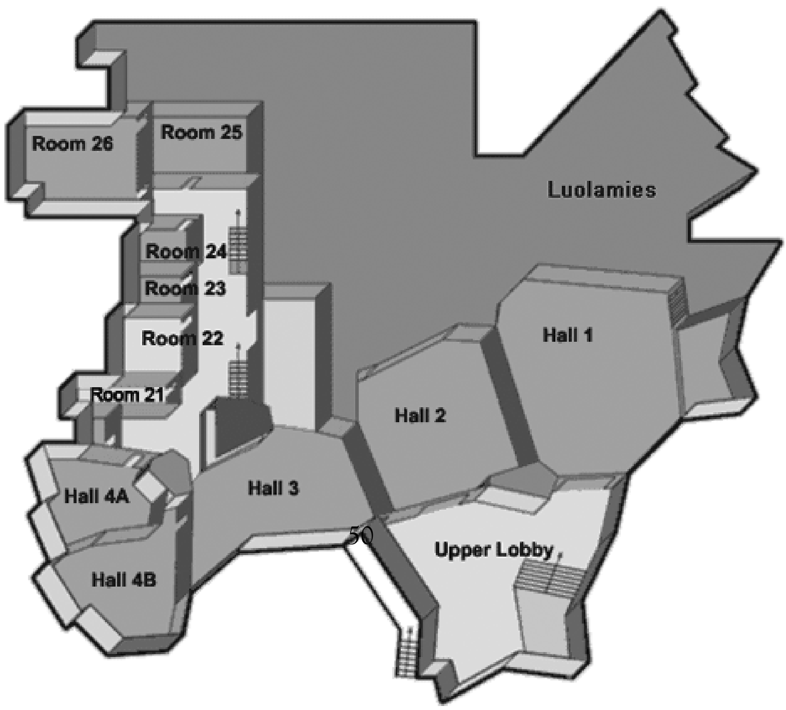
The venue is the Dipoli Congress Center (number 26, circled on the map below) located on the campus of Aalto University (former Helsinki University of Technology), in Espoo (8km West from the city centre of Helsinki).

Dipoli Congress Center,
Otakaari 24, 02150 Espoo, Finland



Dipoli Congress Center Floor Plans





Bus Information

Buses 102 and 103 shuttle from the early morning till midnight between Helsinki bus station (Kamppi) and Aalto University Campus. They depart from piers 41 and 42, respectively, in Kamppi bus station building, which is co-located with Kamppi metro station. The trip takes 13-20 minutes, 3-5 minutes more if the bus goes via Lauttasaari (marked by T on the bus). One-way ticket: 4 Euros.