Statistical Computing 2013

Abstracts der 45. Arbeitstagung

HA Kestler, M Schmid, F Schmid

M Maucher, JM Kraus (eds)

Ulmer Informatik-Berichte

International Graduate School
in Molecular Medicine Ulm

SyStaR

Statistical Computing 2013

45. Arbeitstagung

der Arbeitsgruppen Statistical Computing (GMDS/IBS-DR),
Klassifikation und Datenanalyse in den Biowissenschaften (GfKl).

23.06.-25.06.2013, Schloss Reisensburg (Günzburg)

Workshop Program

Sunday, June 23, 2013

18:15-19:30		**Dinner**
19:30-20:30		**Chair: H.A. Kestler (Ulm)**
19:30-20:30	Uwe Ligges (Dortmund)	R-3.0.x and beyond

Monday, June 24, 2013

08:50-09:00		**Opening of the workshop: H.A. Kestler, M. Schmid**
09:00-12:00		**Chair: U. Ligges (Dortmund)**
09:00-09:30	Günther Sawitzki (Heidelberg)	The In and Out of R: Byte Code, Profiling and Optimization
09:30-10:00	Helena Kotthaus (Dortmund)	Runtime and memory consumption analyses for machine learning R programs
10:00-10:30	Florian Schmid (Ulm)	gsaTools - A toolbox for gene set enrichment analysis in R

10:30-11:00		**Coffee break**

11:00-11:30	Axel Fürstberger (Ulm)	Extended pairwise local alignment of wild card DNA/RNA sequences using dynamic programming
11:30-12:00	Sebastian Behrens (Ulm)	Using VennMaster to evaluate and analyse shRNA data

12:15-14:00		**Lunch**

14:00-16:00		**Chair: M. Maucher (Ulm)**
14:00-15:00	Tim Beißbarth (Göttingen)	Methods for the integration of biological network knowledge into classification algorithms
15:00-15:30	André Burkovski (Ulm)	Aggregating diverse high-throughput data for the identification of common differences between young and old
15:30-16:00	Alfred Ultsch (Marburg)	Functional genomic and transcriptomic analysis of the human olfactory bulb

16:00-17:00		**Poster Session**

17:00-18:00		Working groups meeting on **Statistical Computing 2014** and other topics (all welcome)

18:15-19:30		**Dinner**

Tuesday, June 25, 2013

09:00-12:00		**Chair: M. Schmid (Erlangen)**
09:00-09:30	Anna Telaar (Dortmund)	A pragmatic procedure for stepwise feature selection
09:30-10:00	Benjamin Hofner (Erlangen)	Controlling false discoveries in high dimensional situations: Boosting with stability selection
10:00-10:30	Gunnar Völkel (Ulm)	Information Theoretic Measures for Strategy Evaluation in Ant Colony Optimization

10:30-11:00		**Coffee break**

11:00-11:30	Johann Kraus (Ulm)	Subscan - a cluster algorithm for identifying statistically dense subspaces with application to biomedical data
11:30-12:00	Sebastian Krey (Dortmund)	Power network clustering in modern protection systems

12:15-14:00		**Lunch**

14:00-17:30		**Chair: A. Groß (Ulm)**
14:00-14:30	Vito Baccelliere (Ulm)	Reconstructing gene regulatory networks: deducing the coefficients of stochastic differential equations
14:30-15:00	Markus Maucher (Ulm)	A critical noise level for learning Boolean functions
15:00-15:30	Andreas Mayr (Erlangen)	Boosting sonographic birth weight prediction
15:30-16:00	Michel Lang (Dortmund)	Automatic model selection and configuration for high dimensional survival analysis

16:00-16:30		**Coffee break**

16:30-17:00	Ludwig Lausser (Ulm)	Exhaustive biomarker selection for small and medium sized datasets
17:00-17:30	Werner Adler (Erlangen)	Diversity based ensemble pruning for higher interpretability

18:15-19:30		**Dinner**

R-3.0.x and beyond

Uwe Ligges

Fakultät Statistik, TU Dortmund

Ligges@statistik.tu-dortmund.de

The In and Out of R: Byte Code, Profiling and Optimization

Günther Sawitzki

Runtime and Memory Consumption Analyses for Machine Learning R Programs

Helena Kotthaus, Michel Lang, Jörg Rahnenführer and Peter Marwedel

R is a multi-paradigm language with a dynamic type system, different object systems and functional charac- teristics. These characteristics support the development of statistical algorithms at a high level of abstraction. Although R is commonly used in the statistics domain, a big disadvantage are its performance problems when handling computation-intensive algorithms. Especially in the domain of machine learning, for ex- ample when analyzing high-dimensional genomic data, the execution of R programs is often unacceptably slow. Morandat et al. [2] analyzed R programs from different fields of statistics and were able to show major performance issues. Our goal is to overcome these issues, particularly focusing on machine learning pro- grams [1]. As a first step towards this goal, we used the traceR tool [2] to analyze the bottlenecks arising in this domain. Here, we measured the runtime and overall memory consumption on a well-defined set of classical machine learning applications and gained detailed insights into the performance issues of these programs. Bottlenecks we identified inculde: boxing of scalar values, vector allocation and copy, environment and promise creation and in particular vector subsetting. As in R every value is a vector, scalar values are boxed into single-element vectors, which results in an excessively high number of vector allocations. Variable and function symbols are stored in environments representing the interpreter state. Environments are allocated on the heap of the interpreter, which also applies to variables local to the current environment. Function calls are lazily evaluated. For each function call a promise, representing a function closure with all parameters, is allocated on the heap. Those characteristics cause high runtime and memory management overhead. In this talk we present the results of our runtime and memory consumption analyses and outline approaches to overcome the identified bottlenecks.

TU Dortmund

{Helena.Kotthaus, Michel.Lang, Joerg.Rahnenfuehrer, Peter.Marwedel}@tu-dortmund.de

References

1. Kotthaus, H., S. Plazar, and P. Marwedel (2012). A JVM-based Compiler Strategy for the R Language. Research Poster at The 8th International R User Conference.

2. Morandat, F., B. Hill, L. Osvald, and J. Vitek (2012). Evaluating the Design of the R Language. In Proceedings of the 26th European Conference on Object-Oriented Programming.

gsaTools - A toolbox for gene set analysis in R

Florian Schmid, Johann M. Kraus and Hans A. Kestler

The measurements gained from microarray or deep sequencing experiments are often analyzed for gene set enrichment. There is a wide range of methods available for performing these analyses. The aim is to check the measured data for enrichment or overrepresentation of gene sets connected to any kind of process in a cell. The R-package `gsaTools` embraces these various methods in one toolbox and allows their application by using a generic pipeline which is based on the modular framework given by Ackermann and co-workers [5].

In contrast to other currently available approaches (i.e. [4,2]) this modular framework provides the user the possibility to integrate his own methods into the generic pipeline of the `gsaTools` package. An advantage over available web interfaces (i.e. [3]) is the complete implementation of the package in R which allows a fully automated analysis of data sets. Parts of the package apply computer intensive tests on the given data. To ensure an efficient analysis the package provides the possibility of analyzing a list of gene sets in parallel.

Gene sets which are already known and published are available in databases like the MSigDB [1]. They can be downloaded and used for analysis. If there is no gene set available for a certain pathway or process it can be of interest to create a new gene set. The information sources, such as hints from the literature, which are used to create this new gene set are highly uncertain. To handle this a robustness analysis, which is also a feature of the `gsaTools` package, can be applied to the gene sets. A resampling experiment is performed to quantify the included uncertainty.

Research Group Bioinformatics and Systems Biology, Ulm University, 89069 Ulm, Germany

{Florian-1.Schmid, Johann.Kraus, Hans.Kestler}@uni-ulm.de

References

1. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P.: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sci- ences of the United States of America 102(43), 15,54515,550 (2005)

2. Zeeberg, B., Feng, W., Wang, G., Wang, M., Fojo, A., Sunshine, M., Narasimhan, S., Kane, D., Reinhold, W., Lababidi, S., Bussey, K., Riss, J., Barrett, J., Weinstein, J.: Gominer: a resource for biological interpretation of genomic and proteomic data. Genome Biology 4(4), R28 (2003)

3. Zhang, B., Kirov, S., Snoddy, J.: Webgestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Research 33(suppl 2), W741W748 (2005)

4. Väremo, L., Nielsen, J., Nookaew, I.: Enriching the gene set analysis of genome- wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic Acids Research 41(8), 43784391 (2013)

5. Ackermann, M., Strimmer, K.: A general modular framework for gene set enrich- ment analysis. BMC Bioinformatics 10(1), 47 (2009)

Extended pairwise local alignment of wild card DNA/RNA-sequences with wild-type protein-sequences using dynamic programming

Axel Fürstberger and Hans A. Kestler

Sequence alignment and mutation analysis are essential tasks in modern molecular medicine. Besides finding homologies to identify a common ancestor or reconstruct evolutionary changes, genotypic resistance testing is a another important area of sequence alignment analysis.

Supporting the complete IUPAC nomenclature leads to a new step within the alignment-algorithm, which performs a best case or worst case wild card analysis to calculate the optimal local alignment, depending on the intended use. Supporting the different forms of nucleotide- and amino acid-mutations as well as common and individual scoring matrices, with the genotypic resistance testing the extended pairwise local alignment tool SWAT also has an additional focus on the mutated positions of alignments.

We present an algorithm for the extended pairwise sequence alignments, which covers the problem of input-data-wild cards, offers a high flexible set of parameters and display a detailed alignment output and a compact representation of the mutated positions of the alignment.

Research Group Bioinformatics and Systems Biology, Ulm University, 89069 Ulm, Germany

{Axel.Fuerstberger, Hans.Kestler}@uni-ulm.de

Using VennMaster to evaluate and analyse shRNA data

Sebastian Behrens and Hans A. Kestler

VennMaster is a tool to visualise set data in area proportional Venn diagrams. Since exact solutions of this problem typically do not exist for more then three sets, VennMaster uses an error function and a heuristic particle swarm optimisation algorithm to approximate a solution. We here report the extension of VennMaster to serve as a tool for statistical analysis of set data derived from next generation sequencing experiments. New features include generic criterion adjustment to select for elements included in the graph, based on sequencing counts and an analysis of the significance of the corresponding overlaps. Given a selection criterion for elements, a resampling test is used to assess the significance of each overlap between sets. Overlap significance is then visualised within the graph via colour coding.

A data set of shRNA knockdown experiments on human cancer cell lines was analysed to identify genes playing a key role in proliferation of these cells. VennMaster was used to visualise gene set overlap for different minimal sequencing counts between technical and biological replicates as well as between different cell lines.

While originally intended for, and improved for shRNA library and sequencing data, these new VennMaster features can also be of great use in any other scenario, where overlap between multiple sets need to be visualised and statistically verified.

Additionally a parallelized implementation of the optimisation algorithm was realised, greatly speeding up the diagram generation on machines with multiple cpu-cores making VennMaster suitable for the analysis of growing amounts of next generation sequencing data.

RG Bioinformatics and Systems Biology, University of Ulm

{Sebastian.Behrens, Hans.Kestler}@uni-ulm.de

References

1. Kestler, HA et al. "VennMaster: area-proportional Euler diagrams for functional GO analysis of microarrays." BMC bioinformatics 9.1 (2008): 67.

2. Kestler, HA et al. "Generalized Venn diagrams: a new method of visualizing complex genetic set relations." Bioinformatics 21.8 (2005): 1592-1595.

3. Metzker, ML "Sequencing technologiesthe next generation." Nature Reviews Genetics 11.1 (2009): 31-46.

4. Poli, R, Kennedy J, and Blackwell T. "Particle swarm optimization." Swarm intelligence 1.1 (2007): 33-57.

Methods for the integration of biological network knowledge into classification algorithms

Tim Beißbarth

————————————————

Department of Medical Statistics, University Medical Center Göttingen

Tim.Beissbarth@med.uni-goettingen.de

Aggregating diverse high-throughput data for the identification of common differences between young and old

Andre Burkovski, Johann M. Kraus and Hans A. Kestler

High-throughput studies provide rankings of genes related to a study of interest, like gene activity difference in aging. It is often the case that the rankings of genes vary in each study because of differences in the experimental setup. Also use of different high-throughput technologies hinders a direct comparison of different rankings. Rank aggregation is used to infer a ranking that shows which genes are considered high ranked across all studies. In this regard, rank aggregation combines diverse information to build up a consensus ranking of differentially expressed genes. In this study we integrate several aging related studies for the identification of common differences between young and old phenotype: (i) mouse comparison be- tween old (wildtype liver, 4 samples) and young phenotype (p53ko liver, 4 samples) (Katz et al.), (ii) mouse experiment old (TTD- liver, 6 samples) and young phenotype (TTD+ liver, 6 samples) experiments (Begus-Nahrman et al.), (iii) mouse experiment old (CMP cell type, 22 months, 8 samples) and young phenotype (CMP cell type, 3 months, 8 samples), (iv) mouse experiment from old (Crypts cell type, irradiated, 8 samples) and young phenotype (non irradiated, 8 samples), and (v) Drosophila melanogaster experiment old (gut, 3 samples) and young phenotype (gut, 3 samples). Applying rank aggregation and further analysing the consensus ranking via enrichment analysis reveals several enriched pathways common to all studies that are not reported when analysing each study individually. These pathways can be interpreted as the common difference in aging related experiments. We conclude, that aggregating results from different studies related to the research question can lead to a higher confidence in the gene signature and the discovery of new information in enrichment analysis.

Research Group Bioinformatics and Systems Biology, Institute of Neural Information Processing, Ulm University, 89069 Ulm, Germany

{Andre.Burkovski, Johann.Kraus, Hans.Kestler}@uni-ulm.de

References

1. Katz, S. et. al.: Disruption of Trp53 in Livers of Mice Induces Formation of Carcinomas With Bilineal Differentiation. In Gastroenterology, 2012, 142, 1229-1239.e3.

2. Begus-Nahrmann, Y. et al. Disruption of Trp53 in Livers of Mice Induces Formation of Carcinomas With Bilineal Differentiation. In The Journal of Clinical Investigation, The American Society for Clinical Investigation, 2012, 122, 2283??2288.

Functional genomic and transcriptomic analysis of the human olfactory bulb

Alfred Ultsch[1], Jörn Lötsch[2] et.al.

This reports the analysis of the transcriptome of the human olfactory bulb via RNA quantification intersected with the set of expressed transcriptomic genes using independently available proteomic expression data.

To obtain a functional genomic perspective, this intersection was analyzed for higher-level organization of gene products into biological pathways established in the gene ontology database. We report that a fifth of genes expressed in adult human olfactory bulbs serve functions of nervous system or neuron development, half of them functionally converging to axonogenesis but no other non-neurogenetic biological processes. Other genes were expectedly involved in signal transmission and response to chemical stimuli. This provides a novel, functional genomics perspective supporting the existence of neurogenesis in the adult human olfactory bulb.

[1]Datenbionik, Philipps-Universität Marburg
[2]Klinische Pharmakologie, Universitätsklinikum Frankfurt

Ultsch@ulweb.de

A generic method to find disease-associated genes in databases

Melanie B. Grieb, Johann M. Kraus, Sebastian Behrens, Karl L. Rudolph and Hans A. Kestler

Background

The improvement and cheaper availability of microarray and gene sequencing technologies has led to an increasing number of studies about the association of gene mutations with diseases. Multiple databases collect different types of results of these studies. We developed a generic method that extracts a minimal set of common disease genes from any database containing mutations in genes and samples to a disease. The resulting minimal set can then be applied in enrichment analysis to predict if an arbitrary gene set is associated with the disease.

The method is based on finding a minimal set of genes that covers all samples in the database. Mathematically, the solution to this question is a set covering problem. As it is computationally infeasible to compute the exact solution, we use the greedy heuristic. It starts with the set covering the most elements and adds in each step the set covering the most still uncovered elements, breaking ties by chosing the set covering the most total elements. The algorithm traverses all remaining ties (same number of uncovered elements and same number of total covered elements), until 99% of all elements are covered.

Institute of Neural Information Processing, Institute of Molecular Medicine, University of Ulm, Fritz-Lipmann-Institute, Jena, Germany

Hans.Kestler@uni-ulm.de

Results

To evaluate the validity of the approach on a biological dataset, the algorithm was applied to the Catalogue of Somatic Mutations in Cancer (COSMIC) [1]. The COSMIC has its origin in the census of human cancer genes [1]. It is a database of somatic mutations in cancer based on publications. The current COSMIC version (64) contains more than 800000 entries of mutations in more than 20000 genes for more than 180000 samples.

The algorithm found one solutions containing 108 genes, resulting in one minimal set. Vogelstein et al [2] used a cancer-specific approach to identify Tumor suppressor genes and oncogenes from the COSMIC database. These Tumor suppressor genes and oncogenes were used as a positive control for our analysis. The minimal set found with our generic approach has a true positive rate of 88%, covering 72% of all tumor suppressor genes and 81% of all oncogenes from Vogelstein.

References

1. S. Forbes, N. Bindal, S. Bamford, C. Cole, C. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton, and F. P. A., "Cosmic: mining complete cancer genomes in the catalogue of somatic mutations in cancer," *Nucleic acids research*, vol. 39, no. suppl 1, pp. D945–D950, 2011.
2. B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler, "Cancer genome landscapes," *Science*, vol. 339, pp. 1546–1558, Mar. 2013.

A model of the crosstalk between the Wnt and the IGF signaling pathways

Shuang Wang[1], Michael Kühl[2] and Hans A. Kestler[1]

The Wnt signaling pathway regulates various aspects of cellular processes, such as cell growth and differentiation. The IGF signaling pathway has been shown to tightly influence aging and longevity. Previous research suggests that several molecules function as regulators or downstream effectors which present in both signaling pathways. Signaling pathways are supposed to interact with each other, forming a more complex network. Thus, revealing the crosstalk between the Wnt and the IGF-1 signaling pathways can be helpful to understand aging and lifespan determination. The aim of this study is to establish mathematical multiscale models to analyze the mutual regulation of the Wnt signaling pathway and the IGF signaling pathways. Based on published results, we constructed an overview network of these two signaling pathways which showed that Yap, GSK-3b and Akt are potential nexuses for coupling the Wnt and the IGF signaling pathways. FoxO1 is a critical downstream effector of both pathways in the control of cell aging. The initial overview model can be extended by including stochastic effects, kinetics and time delays in the future.

[1]Research Group Bioinformatics and Systems Biology,

[2]Institute for Biochemistry and Molecular Biology, Ulm University, 89069 Ulm, Germany

Hans.Kestler@uni-ulm.de

Investigation of Fuzzy Support Vector Machines

Markus Frey, Martin Schels, Michael Glodek, Sascha Meudt, Markus Kächele and Friedhelm Schwenker

In many pattern recognition applications, such as medical diagnosis, stress or affect recognition, the ground truth might be hidden. Labeling data sets in such scenarios is difficult, time consuming and error-prone, and therefore experts may disagree on class label of the given input data. The results are typically soft or fuzzy labels. In this study we investigate methods for handling this type of labeled data set. There are several approaches to model uncertainty in classification. Some classification methods - typically regression-based approaches - are suited to handle fuzzy or soft labeled data directly, for instance radial basis function neural networks or multi layer perceptrons. For other classifiers the standard training algorithm must be enhanced to obtain a fuzzy-input fuzzy-output behavior, for example the fuzzy Learning-Vector-Quantization classifier introduced in [1]. Here our aim is to study algorithms that are able to handle fuzzy labels in the training phase, and to compute a fuzzy label in the classification phase.

The fuzzy-input fuzzy-output support vector machine F^2SVM introduced in [2] can be applied in such pattern recognition tasks, for example in [3] it is used for classification of voice qualities in spoken language. The quality of the voice refers to the timbre or coloring of a speaker's voice, this is a typical pattern recognition task where fuzzy labels appear. As shown, in this case F^2SVM can outperform the standard SVM classifiers. In [4] the F^2SVM is investigated on different artificial datasets and the performance of the classifier is compared between fuzzy and hard input data. Furthermore, optimization in l_1 and l_2 sense is investigated to train parameters of a logistic transfer function in order to fuzzify the SVM's output.

Institute of Neural Information Processing, Ulm University, 89069 Ulm, Germany

Friedhelm.Schwenker@uni-ulm.de

References

1. Christian Thiel, Britta Sonntag, and Friedhelm Schwenker. Experiments with supervised fuzzy lvq. *Artificial Neural Networks in Pattern Recognition*, pages 125–132, 2008.
2. Christian Thiel, Stefan Scherer, and Friedhelm Schwenker. Fuzzy-input fuzzy-output one-against-all support vector machines. *Knowledge-based intelligent information and Engineering Systems*, pages 156–165, 2007.
3. Stefan Scherer, John Kane, Christer Gobl, and Friedhelm Schwenker. Investigating fuzzy-input fuzzy-output support vector machines for robust voice quality classification. *Computer Speech & Language*, 27(1):263–287, January 2013.
4. Markus Frey. Optimization of Membership functions in Fuzzy Support Vector Machines. *Bachelor Thesis, Ulm University*, 2013.

A pragmatic procedure for stepwise feature selection

Anna Telaar and Carmen Theek

For classification studies like for example in biomarker research often high dimensional data sets are provided. The challenge is to find a small set of features out of this high dimensional setting which allows the classification of unknown samples into pre-defined groups with the greatest possible accuracy.

Many feature selection approaches exist facing this problem with different focus, but often these are very complex multivariate methods.

We propose a simple algorithm using a stepwise forward selection approach: As classification characteristics can be determined for all single features the AUC (area under curve) calculated within a ROC (receiver operating characteristic) analysis is chosen as the primary criterion for feature selection. Starting with the feature with the highest AUC, features are added successively which improve the classification within a logistic regression model. Selection is done with respect to the improvement in the AUC, and at the same time the new feature has to show a low correlation to the features selected previously.

This pragmatic procedure determines a feature set in a transparent way and does not require a complex theoretical statistical background or time consuming computational resources.

We apply our proposed approach on median fluorescence intensity data provided by the Luminex(R) xMAP technology for individual serum samples to identify biomarker candidates for stratification of diseases. This approach is compared to other feature selection techniques like L_1-penalized logistic regression and Random Forest. The sets of features selected and the accuracy of classification results are considered.

Protagen AG, Otto-Hahn-Strasse 15, 44227 Dortmund, Germany

Anna.Telaar@protagen.com

Controlling false discoveries in high dimensional situations: Boosting with stability selection

Benjamin Hofner[1] and Michael Drey[2]

Modern biotechnologies often result in high-dimensional data sets with much more variables than observations ($n >> p$). These data sets pose new challenges to statistical analysis: Variable selection becomes one of the most important tasks in this setting. Recently, Meinshausen and Bühlmann (2010) proposed a flexible framework for variable selection called stability selection. By the use of resampling procedures, stability selection adds a finite sample error control to high dimensional variable selection procedures such as Lasso or boosting. We consider the combination of boosting and stability selection and present results from a detailed simulation study that presents insights on the usefulness of this combination. Limitations will be discussed and guidance on the specification and tuning of stability selection will be given. The results will then be used for the analysis of an high dimensional data set. All methods are implemented in the R package mboost (Hofner et al., 2012; Hothorn et al., 2010, 2013).

[1]Institut für Medizininformatik, Biometrie und Epidemiologie; Friedrich-Alexander-Universität Erlangen-Nürnberg; Germany
[2]Klinikum Fürth, Fachabteilung Geriatrie; Germany

Benjamin.Hofner@fau.de

References

1. Hofner, B., A. Mayr, N. Robinzonov, and M. Schmid (2012). Model-based boosting in R A hands-on tutorial using the R package mboost. Computational Statistics. Online First.

2. Hothorn, T., P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner (2010). Model-based boosting 2.0. Journal of Machine Learning Research 11, 21092113.

3. Hothorn, T., P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner (2013). mboost: Model-Based Boosting. R package version 2.2-1.

4. Meinshausen, N. and P. Bühlmann (2010). Stability selection (with discussion). Journal of the Royal Statistical Society. Series B 72, 417473.

Information Theoretic Measures for Strategy Evaluation in Ant Colony Optimization

Gunnar Völkel[1], Markus Maucher[2] and Hans A. Kestler[2]

Ant Colony Optimization (ACO) is a meta-heuristic for combinatorial optimization problems. The main idea of ACO is that in each iteration a fixed number of solutions is constructed probabilistically based on a pheromone matrix which evolves between the iterations. When implementing an ACO algorithm for an optimization problem, sev- eral strategic implementation choices are left to the researcher. Those strate- gic choices include the decision on a pheromone update rule and the selection of parameters, e.g. the evaporation rate, the number of solutions constructed per iteration and the number of iterations. The evaluation of those choices often consists of average fitness comparisons of black box runs of the ACO algorithm on selected problem instances. We use information theoretic measures applied to the pheromone matrices of ACO algorithm runs on problem instances to evaluate strategic choices. The internal state of the ACO variants we consider consists of the pheromone matrix and the best-so-far solution. During one iteration of an ACO algorithm all probabilities for the construction choices are derived from the current pheromone matrix. Hence, the pheromone matrix corresponds to a set of random variables. We use the entropy of those random variables during the iterations as first measure of the internal state of the ACO algorithm. As a second measure we propose m-path entropy, i.e. the entropy of constructing a sequence of m solution components probabilistically. Those entropy measures indicate how much exploration is still possible in the corresponding iteration of the algorithm. This can be used to reason about the strategic choices. We compare ClassicACO and GB-ACO, two ACO variants introduced in [2], based on information theoretic measures. Furthermore, we compare the different choices for empty group punishment in `GB-ACO`.

---

[1]Institute of Theoretical Computer Science, University of Ulm
[2]RG Bioinformatics and Systems Biology, University of Ulm

`{Gunnar.Voelkel, Markus.Maucher, Hans.Kestler}@uni-ulm.de`

References

1. Marco Dorigo and Thomas Stützle (2009): Ant Colony Optimization: Overview and Recent Advances. Techreport, IRIDIA, Universite Libre de Bruxelles.

2. Gunnar Völkel, Markus Maucher, Hans A. Kestler: Group-Based Ant Colony Optimization, Proceedings of the fifteenth international conference on Genetic and evolutionary computation conference, ACM, 2013. accepted

Subscan - a cluster algorithm for identifying statistically dense subspaces with application to biomedical data

Johann M. Kraus and Hans A. Kestler

Cluster analysis is an important technique of initial explorative data mining. Recent approaches in clustering aim at detecting groups of data points that exist in arbitrary, possibly overlapping subspaces. Generally, subspace clusters are neither exclusive nor exhaustive, i.e. subspace clusters can overlap as well as data points are not forced to participate in clusters. In this context subspace clustering supports the search for meaningful clusters by including dimensionality reduction in the clustering process. Subspace clustering can overcome drawbacks from searching groups in high-dimensional data sets, as often observed in clustering biological or medical data. In the context of microarray or next-generation sequencing data this refers to the hypothesis that only a small number of genes is responsible for different tumor subgroups. We generalize the notion of scan statistics to multi-dimensional space and introduce a new formulation of subspace clusters as aggregated structures from dense intervals reported by single axis scans. We present a bottom-up algorithm to grow high-dimensional subspace clusters from one-dimensional statistically dense seed regions. Our approach objectifies the search for subspace clusters as the reported clusters are of statistical relevance and are not artifacts observed by chance. Our experiments demonstrate the applicability of the approach to both low-dimensional as well as high-dimensional data.

Research Group Bioinformatics and Systems Biology, Ulm University, 89069 Ulm, Germany

{Johann.Kraus, Hans.Kestler}@uni-ulm.de

Power network clustering in modern protection systems

Sebastian Krey, Sebastian Brato, Uwe Ligges, Claus Weihs and Jürgen Götze

In modern interconnected powersystems (often comprising whole continents) the protection against blackouts is a very complex task. The ongoing replacement of classical power plants with renewable energy provided by highly distributed and relatively small power stations introduces a lot of challenges to maintain the currently in Europe very high quality of the supply with electricity.

In case of an emergency the automatic protection systems in electricity networks divide the system into regions to limit the impact of a component failure to specific region. For a successful reaction it is necessary that the created regions have a balanced energy production and consumption. In times of renewable energies with a large production variation, the classical static defense plans can not always provide an acceptable solution (for example during the large European blackout in November 2006).

In this work we present methods to cluster the network graph of the electricity network into regions based on the current network topology and static information about the electrical characteristics of the network components. We will also present ideas to incorporate the results of a stability assessment of the network using dynamic measurements of voltage and frequency as well as the current power flow (energy production and consumption) into the clustering process. We will compare the results of these different methods on different test systems with each other and a manual clustering based on the expertise of an electrical engineer.

For the selection of a specific clustering result the compliance with regulatory conditions on quality and redundancy of the network is an important factor. We present the most important aspects of this question and how the presented methods perform in this regard.

Fakultät Statistik, TU Dortmund

{Krey, Ligges, Weihs}@statistik.tu-dortmund.de
{Sebastian.Brato, Juergen.Goetze}@tu-dortmund.de

Reconstructing gene regulatory networks: deducing the coefficients of stochastic differential equations

Vito Baccelliere and Ulrich Stadtmüller

The reconstruction of the dependence structure in gene regulatory networks is one of the most challenging tasks in system biology. As shown by Gillespie (2007) and El Samad et al. (2005) randomness is a basic characteristic, when observing a small volume of reactants in a chemical reaction system, which can be described by SDEs (Stochastic Differential Equations). The modeling with SDEs captures the volatility of the system in a flexible way. But for a small sample size, which is the case for experimental datasets, the classical estimation techniques as e.g. Maximum-Likelihood are not feasible for a multi-dimensional system. In my talk I will present some stochastic modeling approaches of chemical reaction systems, motivating the use of SDEs for gene network analysis. And I will discuss some estimation methods for multivariate parametric SDEs.

Institut für Zahlentheorie und Wahrscheinlichkeitstheorie, University of Ulm

Pasquale.Baccelliere@uni-ulm.de

References

1. Gillespie D. T. (2007). Stochastic Simulation of Chemical Kinetics. Annual Review of Physical Chemistry, 58:35-55.

2. El Samad H., Khammash M., Petzold L. and Gillespie D. (2005). Stochastic modelling of gene regulatory networks, 15:691-711.

A critical noise level for learning Boolean functions

Markus Maucher and Hans A. Kestler

The inference of gene regulatory systems from time series measurements is a challenging task to reveal the global functionality of a cell. Among several reconstruction methods Boolean networks have been successfully applied to such data. As time-resolved gene expression measurements at different stages of a cell are difficult and expensive, all reconstruction methods are faced with a relative small number of time points compared to the number of genes. In addition to this dimension problem, biological systems as well as measurement techniques are subject to noise.

In this work, we present an analysis of the reconstructability of Boolean networks in the case of noisy data. We introduce the notion of the critical noise level, a function characteristic which measures the complexity of the reconstruction of a function from noisy time series data. This measure constitutes a natural upper bound for the noise probability under which a function can still be reconstructed, but can also be incorporated into the reconstruction process to improve reconstruction results. We show how to efficiently compute the critical noise level of any given Boolean function and present experimental data that shows how it can be used to improve the best-fit extension algorithm for the reconstruction of a Boolean network from noisy time series data

---

RG Bioinformatics and Systems Biology, University of Ulm

{Markus.Maucher,Hans.Kestler}@uni-ulm.de

References

1. Kauffman, S.A. (1993). The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press.

2. Lähdesmäki, H., Shmulevich, I., Yli-Harja, O. (2003). On learning gene regulatory networks under the boolean network model. Machine Learning, 52(1-2):147–167.

Boosting sonographic birth weight prediction

Andreas Mayr[1], Florian Faschingbauer[2], Matthias Schmid[1]

Sonographic measurements of the fetus are often used to predict birth weight as it is the most important indicator for possible complications during delivery. Statistical challenges in the modelling of these kind of prediction models include multicollinearity and variable selection. The aim of our investigation is to analyze if modern variable selection and regularization tools like component-wise boosting algorithms (Bu?hlmann and Hothorn, 2007), which can cope with both of those issues, can improve existing prediction formulas. We therefore consider boosting generalized additive models (GAM) as well as the recently proposed, more flexible gamboostLSS algorithm (Mayr et al., 2012) for boosting generalized additive models for location, scale and shape (GAMLSS). As distribution-free competitor we applied additive quantile regression boosting (Fenske et al., 2011).

[1]Department of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2]Department of Obstetrics and Gynecology, Universitätsklinikum Erlangen

Andreas.Mayr@fau.de

References

1. Bühlmann, P. and Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. Statistical Science, 22, 477-522.

2. Mayr, A., Fenske, N., Hofner, B., Kneib, T. and Schmid, M. (2012). Generalized additive models for location, scale and shape for high-dimensional data - a flexible approach based on boosting. Applied Statistics 61(3): 403-427.

3. Fenske, N., Kneib, T. and Hothorn, T. (2011). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. Journal of the American Statistical Association, 106(494): 494-510.

Automatic model selection and configuration for high dimensional survival analysis

Michel Lang, Bernd Bischl, Claus Weihs and Jörg Rahnenführer

Many different models for the analysis of high dimensional survival data have been developed over the past years. While some of the models and implementations come with internal parameter tuning automatisms, others require the user to accurately adjust defaults which often feels like a guessing game. Exhaustively try- ing out all model and parameter combinations will quickly become tedious or infeasible in high dimensional and computational intensive settings, even if parallelization is employed. Therefore, we propose to use modern algorithm configuration approaches to efficiently move through the hypothesis space. Two well-known methods for this purpose (which to our knowledge have not been applied to survival analysis before) are iterated F-racing [10] and sequential model-based optimization [6, 9]. In our presentation we will compare the different pros and cons of these two approaches for our problem at hand. In our application we study four lung cancer microarray datasets. For these, we select and configure a predictor based on four survival models in combination with six feature selection filters. As filter methods we consider: two literature based gene lists, two simple univariate scoring filters and two multivariate filters based on mRMR [4, 11] and nearest shrunken centroids [12], respectively. On the model side, we choose the cox proportional hazards model [3] on clinical covariates for reference purposes and compare to random survival forests [7, 8], boosted cox regression [1] and elastic net cox regression [5]. We optimize important hyperparameters of the filters and models with respect to their predictive performance. We parallelize our approaches with the BatchJobs R package [2]. Currently, our optimization targets each data set separately, but we also plan to configure models for multiple data sets in order to obtain predictors that perform well across the whole domain of lung cancer.

Fakultät für Statistik, TU Dortmund

{Lang, Bischl, Weihs, Rahnenfuehrer}@statistik.tu-dortmund.de

References

1. H. Binder, A. Allignol, M. Schumacher, and J. Beyersmann. Boosting for high-dimensional time-to- event data with competing risks. Bioinformatics, 25(7):8906, 2009.

2. B. Bischl, M. Lang, O. Mersmann, J. Rahnenführer, and C. Weihs. Computing on high performance clusters with R: Packages BatchJobs and BatchExperiments. Technical Report 1/2012, TU Dortmund University, 2012. Available at: http://sfb876.tu-dortmund.de/PublicPublicationFiles/bischl_etal_2012a.pdf.

3. D.R. Cox. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological), 34(2):187220, 1972.

4. C. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. In J Bioinform Comput Biol, pages 523529, 2003.

5. J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent, 2009.

6. F. Hutter, H. H. Hoos, and K. Leyton-Brown. Sequential model-based optimization for general algo- rithm configuration. In Proc. of LION-5, page 507523, 2011.

7. H. Ishwaran and U.B. Kogalur. Random survival forests for r. R News, 7(2):2531, October 2007.

8. H. Ishwaran, U.B. Kogalur, E.H. Blackstone, and M.S. Lauer. Random survival forests. Ann. Appl. Statist., 2(3):841860, 2008.

9. P. Koch, B. Bischl, O. Flasch, T. Bartz-Beielstein, C. Weihs, and W. Konen. Tuning and evolution of support vector kernels. Evolutionary Intelligence, 5(3):153170, 2012.

10. M. López-Ibáñẽz, J. Dubois-Lacoste, T. Stützle, and M. Birattari. The irace package, iterated race for automatic algorithm configuration. Technical Report TR/IRIDIA/2011-004, IRIDIA, Université Libre de Bruxelles, Belgium, 2011. URL http://iridia.ulb.ac.be/IridiaTrSeries/IridiaTr2011-004.pdf.

11. H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: criteria of max- dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Ma- chine Intelligence, 27:12261238, 2005.

12. R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proceedings of the National Academy of Sciences of the United States of America, 99(10):656772, 2002.

Exhaustive biomarker selection for small and medium sized datasets

Ludwig Lausser and Hans A. Kestler

Feature selection algorithms are essential for increasing the interpretability of a classification model. By removing distracting measurements, they reduce the search space to a subset of potentially, discriminatory features. Designed for feature spaces of high dimensionality state-of-the-art selection algorithms are based on heuristics that only evaluate a small number of feature combinations. Usually, such heuristic algorithms cannot guarantee to find a globally optimal feature set.

We present a feature selection algorithm based on the cross-validation accuracy of a k-Nearest Neighbor classifier. By taking advantage of the structure of this classifier, we can evaluate the quality of different feature sets in a highly efficient way. This allows for an exhaustive evaluation of all feature subsets of datasets of small or medium dimensionality in a reasonable time span. The current implementation of the algorithm can evaluate approximately 1,100,000 feature sets per minute in 10×10 cross-validations on a dataset of 100 samples (CPU: 3.2 GHz).

The method can be seen as a prototype of a fast exhaustive feature selection algorithm suitable for a variety of distances. By indexing all subspaces in an appropriate order, the scheme can be parallelized easily

Research Group Bioinformatics and Systems Biology, Ulm University, 89069 Ulm, Germany

{Ludwig.Lausser, Hans.Kestler}@uni-ulm.de

Diversity Based Ensemble Pruning for Higher Interpretability.

Werner Adler[1], Zardad Khan[2], Sergej Potapov[1] and Berthold Lausen[2]

Bootstrap aggregated ensembles of classification trees often show improved classification performance compared to single trees (Breiman, 1996). This comes to the cost of less interpretability which is an important aspect e.g. in medical applications, where decisions regarding future treatment of patients have to be justified. Several methods exist to combine both, improved performance and larger interpretability. For example Node Harvest proposed by Meinshausen (2010) is characterized by its interpretability and competitive performance in various situations.

A high diversity between individual base classifiers is deemed to be important in the performance of an ensemble (Kuncheva & Whitaker, 2003). It is our aim to reduce the number of trees without reducing the diversity in the ensemble, result- ing in higher interpretability with comparable performance. To obtain this goal, we examine several strategies to build sparse ensembles based on several diversity measurements. We report and discuss the results obtained using simulated data as well as example data.

[1]Department of Biometry and Epidemiology, University of Erlangen-Nuremberg, Germany
[2]Department of Mathematical Sciences, University of Essex, England

Werner.Adler@imbe.med.uni-erlangen.de

References

1. Breiman, L. (1996): Bagging predictors. Machine Learning, 26:2, 123140.

2. Kuncheva, L.I., Whitaker, C.J. (2003): Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Machine Learning, 51, 181207.

3. Meinshausen, N. (2010): Node Harvest. The Annals of Applied Statistics, 4(4), 20492072.

Liste der bisher erschienenen Ulmer Informatik-Berichte

Einige davon sind per FTP von `ftp.informatik.uni-ulm.de` erhältlich
Die mit * markierten Berichte sind vergriffen

List of technical reports published by the University of Ulm

Some of them are available by FTP from `ftp.informatik.uni-ulm.de`
Reports marked with * are out of print

2000-04   *Wolfgang Gehring*
Ein Rahmenwerk zur Einführung von Leistungspunktsystemen

2000-05   *Susanne Boll, Christian Heinlein, Wolfgang Klas, Jochen Wandel*
Intelligent Prefetching and Buffering for Interactive Streaming of MPEG Videos

2000-06   *Wolfgang Reif, Gerhard Schellhorn, Andreas Thums*
Fehlersuche in Formalen Spezifikationen

2000-07   *Gerhard Schellhorn, Wolfgang Reif (eds.)*
FM-Tools 2000: The 4th Workshop on Tools for System Design and Verification

2000-08   *Thomas Bauer, Manfred Reichert, Peter Dadam*
Effiziente Durchführung von Prozessmigrationen in verteilten Workflow-Management-Systemen

2000-09   *Thomas Bauer, Peter Dadam*
Vermeidung von Überlastsituationen durch Replikation von Workflow-Servern in ADEPT

2000-10   *Thomas Bauer, Manfred Reichert, Peter Dadam*
Adaptives und verteiltes Workflow-Management

2000-11   *Christian Heinlein*
Workflow and Process Synchronization with Interaction Expressions and Graphs

2001-01   *Hubert Hug, Rainer Schuler*
DNA-based parallel computation of simple arithmetic

2001-02   *Friedhelm Schwenker, Hans A. Kestler, Günther Palm*
3-D Visual Object Classification with Hierarchical Radial Basis Function Networks

2001-03   *Hans A. Kestler, Friedhelm Schwenker, Günther Palm*
RBF network classification of ECGs as a potential marker for sudden cardiac death

2001-04   *Christian Dietrich, Friedhelm Schwenker, Klaus Riede, Günther Palm*
Classification of Bioacoustic Time Series Utilizing Pulse Detection, Time and Frequency Features and Data Fusion

2002-01   *Stefanie Rinderle, Manfred Reichert, Peter Dadam*
Effiziente Verträglichkeitsprüfung und automatische Migration von Workflow-Instanzen bei der Evolution von Workflow-Schemata

2002-02   *Walter Guttmann*
Deriving an Applicative Heapsort Algorithm

2002-03   *Axel Dold, Friedrich W. von Henke, Vincent Vialard, Wolfgang Goerigk*
A Mechanically Verified Compiling Specification for a Realistic Compiler

2003-01   *Manfred Reichert, Stefanie Rinderle, Peter Dadam*
A Formal Framework for Workflow Type and Instance Changes Under Correctness Checks

2003-02   *Stefanie Rinderle, Manfred Reichert, Peter Dadam*
Supporting Workflow Schema Evolution By Efficient Compliance Checks

2003-03   *Christian Heinlein*
Safely Extending Procedure Types to Allow Nested Procedures as Values

2003-04   *Stefanie Rinderle, Manfred Reichert, Peter Dadam*
On Dealing With Semantically Conflicting Business Process Changes.

2003-05   *Christian Heinlein*
Dynamic Class Methods in Java

2003-06   *Christian Heinlein*
Vertical, Horizontal, and Behavioural Extensibility of Software Systems

2003-07   *Christian Heinlein*
Safely Extending Procedure Types to Allow Nested Procedures as Values
(Corrected Version)

2003-08   *Changling Liu, Jörg Kaiser*
Survey of Mobile Ad Hoc Network Routing Protocols)

2004-01   *Thom Frühwirth, Marc Meister (eds.)*
First Workshop on Constraint Handling Rules

2004-02   *Christian Heinlein*
Concept and Implementation of C+++, an Extension of C++ to Support User-Defined
Operator Symbols and Control Structures

2004-03   *Susanne Biundo, Thom Frühwirth, Günther Palm(eds.)*
Poster Proceedings of the 27th Annual German Conference on Artificial Intelligence

2005-01   *Armin Wolf, Thom Frühwirth, Marc Meister (eds.)*
19th Workshop on (Constraint) Logic Programming

2005-02   *Wolfgang Lindner (Hg.), Universität Ulm , Christopher Wolf (Hg.) KU Leuven*
2. Krypto-Tag – Workshop über Kryptographie, Universität Ulm

2005-03   *Walter Guttmann, Markus Maucher*
Constrained Ordering

2006-01   *Stefan Sarstedt*
Model-Driven Development with ACTIVECHARTS, Tutorial

2006-02   *Alexander Raschke, Ramin Tavakoli Kolagari*
Ein experimenteller Vergleich zwischen einer plan-getriebenen und einer
leichtgewichtigen Entwicklungsmethode zur Spezifikation von eingebetteten
Systemen

2006-03   *Jens Kohlmeyer, Alexander Raschke, Ramin Tavakoli Kolagari*
Eine qualitative Untersuchung zur Produktlinien-Integration über
Organisationsgrenzen hinweg

2006-04   *Thorsten Liebig*
Reasoning with OWL - System Support and Insights –

2008-01   *H.A. Kestler, J. Messner, A. Müller, R. Schuler*
On the complexity of intersecting multiple circles for graphical display

2008-02   *Manfred Reichert, Peter Dadam, Martin Jurisch,l Ulrich Kreher, Kevin Göser,*
                    Markus Lauer
                    Architectural Design of Flexible Process Management Technology

2008-03   *Frank Raiser*
                    Semi-Automatic Generation of CHR Solvers from Global Constraint Automata

2008-04   *Ramin Tavakoli Kolagari, Alexander Raschke, Matthias Schneiderhan, Ian Alexander*
                    Entscheidungsdokumentation bei der Entwicklung innovativer Systeme für
                    produktlinien-basierte Entwicklungsprozesse

2008-05   *Markus Kalb, Claudia Dittrich, Peter Dadam*
                    Support of Relationships Among Moving Objects on Networks

2008-06   *Matthias Frank, Frank Kargl, Burkhard Stiller (Hg.)*
                    WMAN 2008 – KuVS Fachgespräch über Mobile Ad-hoc Netzwerke

2008-07   *M. Maucher, U. Schöning, H.A. Kestler*
                    An empirical assessment of local and population based search methods with different
                    degrees of pseudorandomness

2008-08   *Henning Wunderlich*
                    Covers have structure

2008-09   *Karl-Heinz Niggl, Henning Wunderlich*
                    Implicit characterization of FPTIME and NC revisited

2008-10   *Henning Wunderlich*
                    On span-P^{cc} and related classes in structural communication complexity

2008-11   *M. Maucher, U. Schöning, H.A. Kestler*
                    On the different notions of pseudorandomness

2008-12   *Henning Wunderlich*
                    On Toda's Theorem in structural communication complexity

2008-13   *Manfred Reichert, Peter Dadam*
                    Realizing Adaptive Process-aware Information Systems with ADEPT2

2009-01   *Peter Dadam, Manfred Reichert*
                    The ADEPT Project: A Decade of Research and Development for Robust and Fexible
                    Process Support
                    Challenges and Achievements

2009-02   *Peter Dadam, Manfred Reichert, Stefanie Rinderle-Ma, Kevin Göser, Ulrich Kreher,*
                    Martin Jurisch
                    Von ADEPT zur AristaFlow® BPM Suite – Eine Vision wird Realität "Correctness by
                    Construction" und flexible, robuste Ausführung von Unternehmensprozessen

2009-03   *Alena Hallerbach, Thomas Bauer, Manfred Reichert*
Correct Configuration of Process Variants in Provop

2009-04   *Martin Bader*
On Reversal and Transposition Medians

2009-05   *Barbara Weber, Andreas Lanz, Manfred Reichert*
Time Patterns for Process-aware Information Systems: A Pattern-based Analysis

2009-06   *Stefanie Rinderle-Ma, Manfred Reichert*
Adjustment Strategies for Non-Compliant Process Instances

2009-07   *H.A. Kestler, B. Lausen, H. Binder H.-P. Klenk. F. Leisch, M. Schmid*
Statistical Computing 2009 – Abstracts der 41. Arbeitstagung

2009-08   *Ulrich Kreher, Manfred Reichert, Stefanie Rinderle-Ma, Peter Dadam*
Effiziente Repräsentation von Vorlagen- und Instanzdaten in Prozess-Management-Systemen

2009-09   *Dammertz, Holger, Alexander Keller, Hendrik P.A. Lensch*
Progressive Point-Light-Based Global Illumination

2009-10   *Dao Zhou, Christoph Müssel, Ludwig Lausser, Martin Hopfensitz, Michael Kühl, Hans A. Kestler*
Boolean networks for modeling and analysis of gene regulation

2009-11   *J. Hanika, H.P.A. Lensch, A. Keller*
Two-Level Ray Tracing with Recordering for Highly Complex Scenes

2009-12   *Stephan Buchwald, Thomas Bauer, Manfred Reichert*
Durchgängige Modellierung von Geschäftsprozessen durch Einführung eines Abbildungsmodells: Ansätze, Konzepte, Notationen

2010-01   *Hariolf Betz, Frank Raiser, Thom Frühwirth*
A Complete and Terminating Execution Model for Constraint Handling Rules

2010-02   *Ulrich Kreher, Manfred Reichert*
Speichereffiziente Repräsentation instanzspezifischer Änderungen in Prozess-Management-Systemen

2010-03   *Patrick Frey*
Case Study: Engine Control Application

2010-04   *Matthias Lohrmann und Manfred Reichert*
Basic Considerations on Business Process Quality

2010-05   *HA Kestler, H Binder, B Lausen, H-P Klenk, M Schmid, F Leisch (eds):*
Statistical Computing 2010 - Abstracts der 42. Arbeitstagung

2010-06   *Vera Künzle, Barbara Weber, Manfred Reichert*
Object-aware Business Processes: Properties, Requirements, Existing Approaches