ulm university universität ulm

Statistical Computing 2012

Abstracts der 44. Arbeitstagung

HA Kestler, H Binder,

M Schmid, JM Kraus (eds)

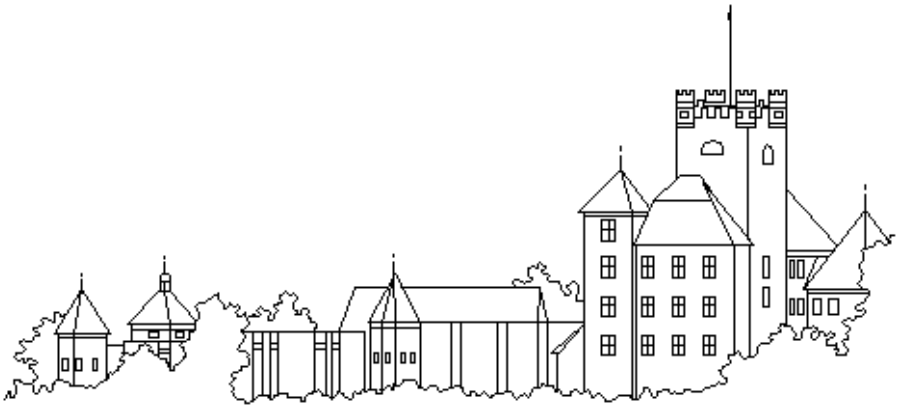
Ulmer Informatik-Berichte

Nr. 2012-02

June 2012

International Graduate School in Molecular Medicine Ulm

Statistical Computing 2012



44. Arbeitstagung

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

24.06.-27.06.2012, Schloss Reisensburg (Günzburg)

Workshop Program

Sunday, June 24, 2012

18:15-19:30		**Dinner**
19:30-20:30		**Chair: H.A. Kestler (Ulm)**
19:30-20:30	Korbinian Strimmer (Leipzig)	Statistical analysis of proteomics data

Monday, June 25, 2012

08:50-09:00		**Opening of the workshop: H.A. Kestler, H. Binder, M. Schmid**
09:00-12:00		**Chair: H. Binder (Freiburg)**
09:00-09:30	Andreas Leha (Göttingen)	Predicting ordinal therapy response with high-dimensional expression data
09:30-10:00	Andre Burkovski (Ulm)	Rank aggregation for candidate gene selection
10:00-10:30	Stefanie Hieke (Freiburg)	Integrated analysis of genome wide data sets related to several molecular measurements in risk prediction models
10:30-11:00		**Coffee break**
11:00-11:30	Alfred Ultsch (Marburg)	Functional abstraction for large gene ontology taxonomies
11:30-12:00	Melanie Grieb (Ulm)	Enrichment analysis based on a minimal attribute core
12:15-14:00		**Lunch**
14:00-18:00		**Chair: G. Füllen (Rostock)**
14:00-15:00	Rainer Spang (Regensburg)	Modelling cell intervention data
15:00-15:30	Florian Schmid (Ulm)	A transductive set covering machine
15:30-16:00	Daniela Herold (Regensburg)	Tumor progression modelling without likelihoods
16:00-16:30		**Coffee break**
16:30-18:00	Bernd Bischl, Michel Lang (Dortmund)	**Tutorial I**: "BatchJobs and BatchExperiments: Abstraction mechanisms for using R in batch environments"
18:15-19:30		**Dinner**
19:30-20:30		**Tutorial II**: "BatchJobs and BatchExperiments"

Tuesday, June 26, 2012

09:00-12:00		**Chair: R. Spang (Regensburg)**
09:00-09:30	Werner Adler (Erlangen)	Increasing the diversity of trees in classifier-ensembles to improve prediction accuracy
09:30-10:00	Murat Sariyar (Mainz)	The benefit of combining techniques for identifying molecular signatures
10:00-10:30	Isabell Hoffmann (Mainz)	Boosting molecular main effects signatures by random forests
10:30-11:00		**Coffee break**
11:00-11:30	Andreas Mayr (Erlangen)	Boosting generalized additive models for location, scale, and shape
11:30-12:00	Julia Schiffner (Dortmund)	Analysis of local classification methods
12:15-14:00		**Lunch**
14:00-18:00		**Chair: G. Sawitzki (Heidelberg)**
14:00-15:00	Georg Füllen (Rostock)	All kinds of support tools for the differential analysis of high-throughput data
15:00-15:30	Thomas Schnattinger (Ulm)	Structural RNA alignment by multi-objective optimization
15:30-16:00	Mohammed Sadeh (Regensburg)	Not confoundable partial network reconstruction with unknown-unknown players
16:00-16:30		**Coffee break**
16:30-18:00		Working groups meeting on **Statistical Computing 2013** and other topics (all welcome)
18:15-19:30		**Dinner**

Wednesday, June 27, 2012

09:30-12:00		**Chair: M. Schmid (Erlangen)**
09:00-09:30	Sebastian Krey (Dortmund)	Visualization and Clustering of Electricity Networks
09:30-10:00	Günther Sawitzki (Heidelberg)	The excess mass approach and diagnostics for modality
10:00-10:30	Matthias Schmid (Erlangen)	A comparison of cumulative/dynamic estimators to evaluate the prediction accuracy of survival models
10:30-11:00		**Coffee break**
11:00-11:30	Gunnar Völkel (Ulm)	Group-based ant colony optimization
11:30-12:00	Leo Geppert (Dortmund)	Efficient Bayesian analysis on the method of “merge and reduce”
12:15-14:00		**Lunch**

Statistical analysis of proteomics data

Korbinian Strimmer
joint work with Sebastian Gibb

Protemic mass spectrometry profiling is increasingly becoming an important tool in clinical diagnostics, for example to identify biomarkers for cancer. Similarly as with other high-throughput technologies, sophisticated statistical algorithms are essential in the analysis of spectrometry data.

In my talk I will discuss the statistical and algorithmic challenges involved in the analysis of clinical mass spectrometry data. I will discuss all steps required in a complete analysis from importing of raw data, preprocessing (e.g. baseline removal), peak detection, non-linear peak alignment to calibration of mass spectra and feature selection and classification.

In order to facilitate automated analyzes we have implemented this pipeline in the open source R package MALDIquant [1] which is freely available from CRAN [2].

References

1. http://strimmerlab.org/software/maldiquant/
2. http://cran.r-project.org/web/packages/MALDIquant/

Institut für medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig

Strimmer@uni-leipzig.de

Predicting ordinal therapy response with high-dimensional expression data

Andreas Leha, Klaus Jung, and Tim Beißbarth

Molecular diagnosis or prediction of clinical treatment outcome based on high-throughput genomics data is a modern application of machine learning techniques for clinical problems. In practice, clinical parameters, such as patient health status or toxic reaction, are often measured on an ordinal scale, as these parameters are frequently staged or graded (e.g. good, fair, poor). The most common approach to treat the prediction of ordinal end-points as a multi-class classification problem disregards the ordering information contained in the response. This may result in a loss of prediction accuracy. Classical approaches to model ordinal response directly, including for instance the cumulative logit model or the continuation ratio model, are typically not applicable to high-dimensional data where the number of samples is much smaller than the number of features. Although there have been some extensions of existing methods for response prediction tailored towards ordinal response and high-dimensional data (Ananth and Kleinbaum, 1997; Hechenbichler and Schliep, 2006; Chu and Keerthi, 2007; Archer and Mas, 2009), the choice of methodology is still limited and the field is still in focus of research (Archer and Williams, 2012). We compare several available and novel approaches for ordinal classification on both, real world data as well as simulated data. Our novel approaches include the forest of small trees (fost) inspired by Guo, Sarkar, and Peddada (2010) and the hierarchical towing (hi2) extending Frank and Hall (2001) that both combine the power of well-understood binary classification with ordinal response prediction and are targeted at high-dimensional data directly. Our findings suggest, that the classification performance of an algorithm is dominated by its ability to deal with the high-dimensionality of the data. Although the comparative evaluation do not show a clear winner, taking the ordinality of the response into account can improve the classification accuracy.

Department of Medical Statistics, University Medical Center Göttingen

Andreas.Leha@med.uni-goettingen.de

References

1. Ananth, C V and D G Kleinbaum (1997). Regression models for ordinal responses: a review of methods and applications. In: International Journal of Epidemiology 26.6, pp. 1323–1333.
2. Archer, K J and V R Mas (Dec. 2009). Ordinal response prediction using bootstrap aggregation, with application to a high-throughput methylation data set. In: Stat Med 28.29, pp. 3597–610.
3. Archer, K.J. and A.A.A. Williams (2012). L 1 penalized continuation ratio models for ordinal response prediction using high-dimensional datasets. In: Statistics in Medicine.
4. Chu, Wei and S. Sathiya Keerthi (2007). Support Vector Ordinal Regression. In: Neural Computation 19.3, pp. 792–815.
5. Frank, Eibe and Mark Hall (2001). A simple approach to ordinal classification. In: Proc 12th Europ Conf on Machine Learning. Springer, pp. 145–156.
6. Guo, W, S K Sarkar, and S D Peddada (2010). Controlling false discoveries in multidimensional directional decisions, with applications to gene expression data on ordered categories. In: Biometrics 66.2, pp. 485–92.
7. Hechenbichler, Klaus and Klaus Schliep (2006). Weighted k-nearest-neighbor techniques and ordinal classification. In: Discussion Paper 399, SFB 386.

Rank aggregation for candidate gene selection

Andre Burkovski, Ludwig Lausser, and Hans A. Kestler

Biomolecular processes are highly affected by the RNA concentrations (gene expression level) of the involved cells. Expression profiles are screened for differentially expressed genes in order to explain changes in such systems. The top scored genes for these profile are considered for further analysis. This ranking process may be influenced by variation in measurement or other sources of noise leading to selection of genes which may not be related to the biological process.

Our approach is the analysis of relative ranks within gene expression profiles which may provide different results and interpretation of gene lists. Since the measurements of a single profile are taken under more or less the same conditions, they are more likely to be comparable to each other, than the measurements of different profiles. The ranking over a single profile is therefore more reliable than the ranking of an aggregated value. Relationships that are stable over all samples can be extracted by aggregating the individual rankings into a consensus ranking. To reveal changes in gene activity we compare intra- and inter-class aggregated rankings. Set operations, like intersection, are applied to the aggregates to create a specific features sets, which may provide candidate genes for further analysis or investigation.

We will compare statistical and positional rank aggregation approaches and their application to artificial as well as real world data. The resulting consensus rankings may be used to identify specifically expressed genes and differences between groups.

RG Bioinformatics and Systems Biology, University of Ulm, Germany

Hans.Kestler@uni-ulm.de

Integrated analysis of genome wide data sets related to several molecular measurements in risk prediction models

Stefanie Hieke[1,2], Thomas Hielscher[3], Richard F. Schlenk[4], Martin Schumacher[1], Axel Benner[3], Lars Bullinger[4], and Harald Binder[1,5]

In recent years, high-throughput microarray technology has become available which allows to measure various molecular features in parallel. Integration of such multiple genome wide data sets in risk prediction models with regard to clinical endpoints could potentially help to improve therapy management for future patients. We systematically investigate statistical strategies to connect several molecular sources with partial overlap in the biological samples. A promising strategy is to take biological hierarchies into account using several sources in different roles. We adapt an approach which considers first one molecular source and keeps the information from this source fixed in the model when incorporating the second source. We illustrate this strategy in an application to survival data from acute myeloid leukemia patients. Specifically, we consider microarray-based gene expression profiling and single nucleotide polymorphism microarrays with relatively small overlap in the biological samples. While each of the molecular sources could be considered as a first or second source statistically, we will highlight how a specific combination corresponds to relevant biological questions. Particularly, certain molecular entities are seen to only emerge in clinical risk prediction signatures after taking the other molecular levels explicitly into account. These results indicate how in general statistical procedures can be adapted for connecting different molecular sources in a way that is related to the underlying biology, resulting in a potentially improved basis for individual therapy management.

[1]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg
[2]Freiburg Center for Data Analysis and Modeling, University Freiburg
[3]Division of Biostatistics, German Cancer Research Center, Heidelberg
[4]Department of Internal Medicine III, University Hospital of Ulm
[5]Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center Johannes Gutenberg University Mainz

Hieke@imbi.uni-freiburg.de

Functional abstraction for large Gene Ontology taxonomies

Joern Loetsch und Alfred Ultsch

The comparison of problem specific lists of gene product functions with the set of all genes of an organism using Over Representation Analysis (ORA) results often in large taxonomies consisting of several hundreds of Gene Ontology (GO) terms. To maximize the information about different mechanisms in which particular genes are involved, an abstraction of these huge taxonomies is required. This can be obtained by using the Shannon Information on the size of the gene set associated with each GO-term and combining it with the certainty that a GO-term describes the biological process involved. This combination provides a measure of the remarkableness of a GO-term. The distribution of remarkableness can be used to identify GO terms that are both highly remarkable and as specific as possible. The terms identified by this procedure describe disjunct functional areas in which the genes under consideration are involved. The approach is tested on gene lists for human pain and micro-RNAs.

Datenbionik, Philipps-Universität Marburg

Ultsch@informatik.uni-marburg.de

Enrichment analysis based on a minimal attribute core

Melanie B. Grieb[1], Johann M. Kraus[1], Karl Lenhard Rudolph[2], and Hans A. Kestler[1]

Merging information about cell function with information about the corresponding pathway or disease is one of the major challenges in the life sciences. Functional information is often represented by experimental data, e.g. differentially expressed genes. Enrichment analysis is a common approach to evaluate if experimental data is connected to an attribute set of interest. The topic of our research is the development of a new enrichment analysis method that uses the information of databases connecting phenotypes and attributes. The method is based on a minimal set of attributes which cover all phenotypes in the database and therefore is a prototype for the attributes associated with all phenotypes in the database. The construction algorithm uses a set P of phenotypes and a set A of attributes and creates the minimal set A_{min} for A as the minimal number of attributes $A_{min} \in A$ that covers all phenotypes in P. The construction of A_{min} works on a mapping matrix of P and is based on the greedy algorithm for the set covering problem. The method was applied to databases of cancer types associated with chromosomal locations of mutations where a minimal set of chromosomal locations covering all cancer types was constructed. The minimal set was validated by an enrichment analysis using tumor suppressor genes as a positive control and inflammation genes as a negative control. The enrichment analysis results in statistically significant values for the tumor suppressor genes and to nonsignificant values for the inflammation genes.

In an application of this minimal set we investigate a possible connection of transient telomere dysfunction with Hepatocellular Carcinoma (HCC) in a cross-species analysis of human and mural genes. The minimal set is used to find the mural genes specific to human Hepatocellular Carcinoma (HCC+).

[1]RG Bioinformatics and Systems Biology, University of Ulm
[2]Institute of Molecular Medicine and Max-Planck-Research Department on Stem Cell Aging, University of Ulm

Hans.Kestler@uni-ulm.de

The mural genes expressed in tumors of mice with transient telomere dysfunction (TTD+) and the mural genes expressed in tumors of wild-type mice were then intersected with HCC+. The intersection of HCC+ and TTD+ was more than 8 times larger than the intersection of HCC+ and TTD-, which is consistent with mouse experiments, where more and larger HCC tumors were found in TTD+ mice than in TTD- mice.

References

1. M. Ackermann and K. Strimmer. A general modular framework for gene set enrichment analysis. *BMC Bioinformatics*, 10:47, 2009.
2. M. Baudis and M.L. Cleary. Progenetix.net: an online repository for molecular cytogenetic aberration data. *Bioinformatics*, 17(12):1228–1229, 2001.
3. Y. Begus-Nahrmann, D. Hartmann, J.M. Kraus, P. Eshraghi, A. Scheffold, M. Grieb, V. Rasche, P. Schirmacher, H.W. Lee, H.A. Kestler, A. Lechel, and K.L. Rudolph. Transient telomere dysfunction induces chromosomal instability and promotes carcinogenesis. *The Journal of Clinical Investigation*, 122(6):2283–2288, 2012.
4. F. Mitelman, B. Johansson, F. Mertens, et al. Mitelman database of chromosome aberrations and gene fusions in cancer. *Available at http://cgap.nci.nih.gov/Chromosomes/Mitelman*, 2010.
5. A. Subramanian, P. Tamayo, V.K. Mootha, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–15550, 2005.

Modeling cell intervention data

Rainer Spang

Functional genomics has a long tradition of inferring the inner working of a cell through analysis of its response to various perturbations. Observing cellular features after knocking out or silencing a gene reveals which genes are essential for an organism or for a particular pathway. A key obstacle to inferring genetic networks from perturbation screens is that phenotypic profiles generally offer only indirect information on how genes interact.

I will briefly summarize the biological concept of signalling pathways and their crosstalk. The rest of my lecture will be on statistical models of pathway structure. Special attention will be given to strategies for controlling network complexity and the influence of hidden confounders. I will demonstrate the power of our methods in the context of modelling disrupted Wnt signalling in colorectal cancers.

Institute for Functional Genomics, University of Regensburg

Rainer.Spang@klinik.uni-regensburg.de

A transductive set covering machine

Florian Schmid, Ludwig Lausser, and Hans A. Kestler

Classification of tissue samples on the basis of gene expression profiles is an essential task in molecular medicine. The corresponding decision boundary of such a classification can not only be used for diagnostic purposes. It also allows insights into the set of involved features. It should therefore be based on a relative small set of genes.

A classifier which is based on such a small set of predictive genes is the Set Covering Machine (SCM) with data dependent rays [1]. This Set Covering Machine combines a minimal set of single threshold classifiers in order to minimize the empirical risk under the constrained of covering all positive samples while minimizing the number of errors on the negative ones. This combined ensemble of base classifiers offers a decision rule of very low dimensionality.

The SCM is trained according to the inductive learning scheme. Its training is solely based on samples already classified by a specialist. In scenarios with high labeling costs, as the one described above, the amount of labeled samples is very low. As a consequence, inductive classifiers tend to overfit in this setting. The trained models often do not generalize. An alternative in this setting is the transductive learning scheme. In these scheme also the information given by unlabeled samples can be included in the training process. We developed a transductive version of the SCM with data dependent rays. The new classifier is tested on artificial data and real microarray datasets with different ratios of labeled and unlabeled samples.

References

1. Kestler, HA, Lausser, L, Lindner, W, Palm, G: On the fusion of threshold classifiers for categorization and dimensionality reduction. Comp Stat 26(2), 321–340 (2011)

Research Group Bioinformatics and Systems Biology, University of Ulm
Hans.Kestler@uni-ulm.de

Tumor progression modelling without likelihoods

Daniela Herold, Giusi Moffa, and Rainer Spang

The emergence and progression of cancer is driven by the accumulation of multiple mutations in the genome. The successive occurrence of mutations follows rules that can be modeled statistically. In the last decades, enormous progress has been made in modelling tumor progression. A recently developed method called Conjunctive Bayesian Networks (CBN) [1, 2] models tumor progression as a directed acyclic graph (DAG). Thus cancer progression allows for both branching and convergence of progression pathways.

Tree or network topologies are estimated by maximizing the likelihood of networks for a given dataset of observed mutations. This is embedded in a simulated annealing framework, where in each iteration the likelihoods of two competing models need to be compared. With complex models, likelihood calculations can be time consuming. In the CBN example, it involves a nested EM-algorithm. Together with the simulated annealing loop this computational burden confines the applicability of CBN to only small progression models.

We discuss an alternative approach that keeps the simulated annealing framework but does not calculate likelihoods explicitly. In a nutshell: When having to decide which of two competing models fit the data better, we simulate multiple sets of artificial data from both candidates, train a support vector machine (SVM) to distinguish them and finally apply this SVM to the real data.

We compare the performances of our approach and the CBN methodology [3] with respect to network reconstruction in a simulation study. Positive predictive values of network reconstruction are comparable to those obtained by likelihood calculations. More importantly, our approach is much faster and allows for the estimation of much larger models.

Institute of Functional Genomics, Computational Diagnostics Group, University of Regensburg

{Daniela.Herold,Rainer.Spang}@klinik.uni-regensburg.de

References

1. Beerenwinkel, N., Eriksson, N., and Sturmfels, B. (2006). Evolution on distributive lattices. Journal of theoretical biology, 242(2), 409–420.
2. Beerenwinkel, N., Eriksson, N., and Sturmfels, B. (2007). Conjunctive bayesian networks. Bernoulli, 13(4), 893–909.
3. Gerstung, M., Baudis, M., Moch, H., and Beerenwinkel, N. (2009). Quantifying cancer progression with conjunctive Bayesian networks. Bioinformatics, 25(21), 2809.

BatchJobs and BatchExperiments: Abstraction mechanisms for using R in batch environments

Michel Lang, Bernd Bischl, Olaf Mersmann,
Jörg Rahnenführer, and Claus Weihs

Empirical analysis of statistical algorithms often demands time-consuming experiments which are best per- formed on high performance computing clusters. While distributed computing environments provide immense computational power, they are not easy to use for the non-expert. We present two R packages [1, 2] which greatly simplify working in batch computing environments. The package BatchJobs implements the basic objects and procedures to control a batch cluster from within R. It is structured around cluster versions of the well-known higher order functions Map/Reduce/Filter from functional programming. The second package, BatchExperiments, is tailored for the still very general scenario of analyzing arbitrary algorithms on problem instances. It extends BatchJobs by letting the user define an array of jobs of the kind apply algorithm A on problem instance P and store results R. It is possible to associate statistical designs with parameters of algorithms and problems and therefore systematically study their influence on the algorithms performance. An important feature is that the state of computation is persistently available in a database. The user can query the status of jobs, submit subsets of the experiments to the batch system and in a later stage add or remove experiments. As SQLite is used as a backend, all data required to reproduce the experimental study, including internally handled seeds, resides in a single directory. The abstraction mechanism used to separate source code from cluster specific code makes the study portable to other batch systems. Both packages are written in such a way that they can be used in principle on any batch system. We currently support standalone multicore machines, SSH clusters, Torque/PBS, LSF and Sun Grid Engine.

Fakultät für Statistik, TU Dortmund

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

References

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

Increasing the diversity of trees in classifier-ensembles to improve prediction accuracy

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

Classification ensembles tend to show higher prediction accuracy, when the base classifiers constituting the ensemble show less correlation to each others, i.e. when the ensembles are more diverse (Kuncheva & Whitaker, 2003). This lead to the development of several strategies to increase the diversity of base classifiers, e.g. bagging (Breiman, 1996), or random forests (Breiman, 2001). For the special case of a repeated measurements setup, Adler et al. (2011) examined modified bootstrap strategies for bagged classification trees and random forests. These strategies lead to an improved performance, which was assumed to be due to an increased diversity among the single trees.

In this talk we will verify this assumption based on a number of diversity measures, examined e.g. by Kuncheva & Whitaker (2003) or Tang et al. (2006). In the next step, appropriate measures are used to apply a weighting scheme to the single trees of a classifier ensemble based on their diversity to each others. The effect on the ensemble accuracy is tested by a simulation study and clinical data sets.

References

1. Adler, W., Potapov, S., and Lausen, B. (2011): Classification of repeated measurements data using tree-based ensemble methods. *Computational Statistics*, 26(2), 355–369.
2. Breiman, L. (1996): Bagging Predictors. *Machine Learning*, 24(2), 123–140.
3. Breiman, L. (2001): Random Forests. *Machine Learning*, 45, 5–32.
4. Kuncheva, L.I., Whitaker, C.J. (2003): Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning*, 51, 181–207.

[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

The benefit of combining techniques for identifying molecular signatures

Murat Sariyar, Isabell Hoffmann, and Harald Binder

Different model families can exploit different kind of information in the molecular data. This can be ascribed to unequal decision boundaries. Combinations of several methods might therefore increase the prediction performance and the generalizability of the results. In practice, such combinations are already used in an exploratory way. However, systematic assessments of possible benefits are still lacking. We provide a preliminary framework to determine if a combination of specific methods for molecular signature development is promising or not. The presentation of the framework is case-based: we investigate whether random survival forests and a Cox model with regularized estimation are beneficial for each other in a high-dimensional time-to-event setting. Following diagnosis tools are used in a subsampling context: variable selection frequencies, prediction performance measured by the Brier score, model rebuilding based on variable selections of the other method, and model-based simulation of new data. Rebuilding models based on information of the other model leads to relevant insights. For variable selection by regularized estimation our results show a stable shift towards different, potentially interesting genes in an application to microarray data for a time-to-event endpoint. For prediction performance, no such benefit is diagnosed. We furthermore indicate how more knowledge regarding the combination of techniques can be gained by theoretical investigations of the underlying decision boundaries.

Institute for Medical Biostatistics, Epidemiology and Informatics, University Medical Center, Johannes Gutenberg-University Mainz

Murat.Sariyar@unimedizin-mainz.de

Boosting molecular main effects signatures by random forests

Isabell Hoffmann, Murat Sariyar, and Harald Binder

When linking high-dimensional molecular measurements to some clinical endpoint, sparse regression techniques are destined to provide a short molecular signature of main effects. While interactions are highly likely to be present in molecular applications, there is no standard method for determining interactions that should be added to a list of main effects. We specifically consider component-wise boosting as a sparseness inducing main effects regression method for the Cox proportional hazards model in a time-to-event settings. It combines variable selection with model estimation. For obtaining interaction information, random survival forests are used. Using resampling techniques on data from patients with diffuse large B-cell lymphoma, we investigate whether it is possible to extract interactions from the forests for injecting them into the Cox model. The search for interactions is based on variable importance measures, their empirical distribution and resulting inclusion frequencies. After determining potential interaction terms, we consider several ways of incorporating them into the Cox model together with the main effects. These are evaluated according to prediction performance. Our results show that there is no straightforward way of improving performance by adding interactions. One reason could be that interactions are not relevant on our data. A more serious reason might be that adding relevant and manageable interactions terms to main effects signatures through random forests is generally difficult.

Institute for Medical Biostatistics, Epidemiology and Informatics, University Medical Center, Johannes Gutenberg-University Mainz

Murat.Sariyar@unimedizin-mainz.de

Boosting generalized additive models for location, scale and shape

Andreas Mayr[1], Nora Fenske[2], Benjamin Hofner[1],
Thomas Kneib[3], and Matthias Schmid[1,2]

Generalized additive models for location, scale and shape (GAMLSS) are a popular semi-parametric modelling approach that, in contrast to conventional GAMs, regress not only the expected mean but every distribution parameter (e.g. location, scale and shape) to a set of covariates. Current fitting procedures for GAMLSS are infeasible for high-dimensional settings and require variable selection based on (potentially problematic) information criteria. This presentation describes a boosting algorithm for high-dimensional GAMLSS that was developed to overcome these limitations. Specifically, the new algorithm was designed to allow the simultaneous estimation of predictor effects and variable selection. The proposed algorithm was applied to data of the Munich Rental Guide, which is used by landlords and tenants as a reference for the average rent of a flat depending on its characteristics and spatial features. The net-rent predictions that resulted from the high-dimensional GAMLSS were found to be highly competitive while covariate-specific prediction intervals showed a major improvement over classical GAMs. The algorithm is available on CRAN with the R add-on package gamboostLSS.

[1]Institut für Medizininformatik, Biometrie und Epidemiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2] Institut für Statistik, Ludwig-Maximilians-Universität München
[3] Wirtschaftswissenschaftliche Fakultät, Georg-August-Universität Göttingen

Andreas.Mayr@imbe.med.uni-erlangen.de

References

1. Mayr, A., N. Fenske, B. Hofner, T. Kneib and M. Schmid (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.
2. Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). Applied Statistics 54(3), 507–554.

Analysis of local classification methods

Julia Schiffner and Claus Weihs

There are various ways to structure the vast amount of classification methods that are nowadays available. For example parametric and nonparametric approaches or predictive and generative methods are usually distinguished. Another distinction can be made between local and global classification methods. Recently there is an increasing interest in local methods in the data mining and machine learning community. Localized versions of nearly all standard classification techniques like linear discriminant analysis [1], logistic regression [2], support vector machines [4], or boosting [5] are available.

The term 'local' is not clearly defined and rather used in an intuitive way by most authors. The probably best-known example of a local classification approach is the k nearest neighbors method where 'local' refers to the environment of the point to be predicted. But there are also many other types of local methods in the literature. We will give an overview of existing approaches. Especially, we will identify several properties that can be used for grouping the various local approaches and may affect their performance.

Our central question of interest is: How do local methods perform – in general as well as in comparison with global counterparts? This surely depends on a number of different factors, on the one hand the properties of the employed methods, on the other hand the properties of the classification problem. Hence related questions are: How important is which type of local method is used? How important is the base classification method itself? What are the effects of certain properties of the data? For which types of classification problems may local methods be beneficial?

In this talk we present our efforts to investigate these issues and partly answer the questions mentioned above.

Department of Statistics, TU Dortmund

{Schiffner,Weihs}@statistik.tu-dortmund.de

As performance measure we consider the misclassification rate since it is usually of interest. Moreover, we use the bias-variance decomposition of the error rate in order to gain deeper insight into the behavior of local learning algorithms. The bias-variance decomposition of prediction error was originally introduced for quadratic loss functions, but generalizations to other loss functions, especially zero-one loss, have been developed in the last 15 years, e.g. [3].

In order to systematically address the questions regarding the influence of the locality type and the base method on the performance we consider 6 global classification methods of different complexity (a simple classifier that always predicts the most frequent class, linear and quadratic discriminant analysis, logistic regression, support vector machines and neural networks) and investigate 4 different local versions of each method.

For illustration and in order to check our intuitions regarding the performance of local methods we first present some results obtained on 2-dimensional artificial classification problems where we think or where it is claimed that local methods may be beneficial.

Moreover, we present an extensive simulation study. Here we attempt to artificially generate data in order to cover a wide range of possible classification problems. The obtained misclassification rates as well as bias and variance are related to properties of the data sets as well as the classification methods.

References

1. I. Czogiel, K. Luebke, M. Zentgraf, and C. Weihs. Localized linear discriminant analysis. In R. Decker and H.-J. Lenz, editors, *Advances in Data Analysis*, volume 33 of *Studies in Classification, Data Analysis, and Knowledge Organization*, pages 133–140, Berlin Heidelberg, 2007. Springer.
2. D. J. Hand and V. Vinciotti. Local versus global models for classification problems: Fitting models where it matters. *The American Statistician*, 57(2):124–131, 2003.
3. G. M. James. Variance and bias for general loss functions. *Machine Learning*, 51(2): 115–135, 2003.
4. N. Segata and E. Blanzieri Fast and scalable local kernel machines. *Journal of Machine Learning Research*, 11:1883–1926, June 2010.
5. C.-X. Zhang and J.-S. Zhang. A local boosting algorithm for solving classification problems. *Computational Statistics & Data Analysis*, 52:1928–1941, 2008.

All kinds of support tools for the differential analysis of high-throughput data

Georg Füllen

We would like to demonstrate how similarity methods can be applied to "differential data", that is, to the results of differential analyses of high-throughput data, with or without replicates.

More specifically, clustering, similarity searches and machine learning can all be applied to "differential data". We also demonstrate how we can profit from a formal ontological framework for "differential data".

Application examples are taken from the domains of stem cells, cancer, ageing and autoimmunity diseases. For example, knocking down the Nanog gene in embryonic stem cells triggers specific regulatory changes affecting the Esrrb gene. As another example, the differential analysis of human and mouse embryonic and epiblast stem cells implicates histone acetylation as a relevant factor.

Universitätsmedizin Rostock

Fuellen@uni-rostock.de

Structural RNA alignment by multi-objective optimization

Thomas Schnattinger[1], Uwe Schöning[1], and Hans A. Kestler[2]

Functional RNA molecules are often more conserved in their secondary structure than in their primary sequence. To examine functional similarity, sequence and structure must both be taken into account. Sequence alignment tools fail when the sequence similarity is too low. Therefore, dynamic programming algorithms have been developed to integrate RNA structure information in the aligning process by constructing a consensus secondary structure along with the sequence alignment. It is not clear, however, how sequence and structure information should be weighted in the objective function, different kinds of RNA may even require different parameters.

To overcome this fixed weighting in these algorithms, which is their major drawback, we split the objective function for structural RNA alignment into two parts. The first objective measures the quality of sequence alignment by penalizing gaps and mismatches and rewarding sequence matches. The second objective is to maximize the stability of the constructed consensus secondary structure.

Having to satisfy more than one objective function, the concept of optimality has to be generalized. We say a vector a dominates b if all entries a_i are greater than or equal to b_i and if there is at least one entry a_j which is strictly larger than b_j. We are interested in all solutions S to our problem with are Pareto-optimal, in other words there is no other solution which dominates S. This task is accomplished by multi-objective dynamic programming.

We present an algorithm for the multi-objective RNA sequence-structure alignment problem. It allows the user to generate and choose from a set of Pareto-optimal solutions, which do not depend on previously set parameters for the weighting between primary sequence and secondary structure.

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

{Thomas.Schnattinger,Uwe.Schoening,Hans.Kestler}@uni-ulm.de

References

1. I. L. Hofacker, S. H. F. Bernhart, and P. F. Stadler. Alignment of RNA base pairing probability matrices. Bioinformatics, 20(14):2222–2227, 2004.
2. H. Laux. Entscheidungstheorie. Springer-Verlag, Berlin, Germany, 6th edition, 2005.
3. T. Schnattinger, U. Schöning, and H. Kestler. Pareto-optimal RNA Sequence-Structure alignments. In 9th International Workshop on Computational Systems Biology 2012 (WCSB 2012), Ulm, Germany, June 2012.
4. S. Will, K. Reiche, I. L. Hofacker, P. F. Stadler, and R. Backofen. Inferring noncoding RNA families and classes by means of genome-scale structure- based clustering. PLoS Comput Biol, 3(4):e65, 2007.

Not confoundable partial network reconstruction with unknown-unknown players

Mohammad J. Sadeh, Giusi Moffa, and Rainer Spang

Our current understanding of virtually all cellular signaling pathways is almost certainly incomplete. We miss important but sofar unknown players in the pathways. Moreover, we only have a partial account of the molecular interactions and modifications of the known players. When analyzing the cell, we look through narrow windows leaving potentially important events in blind spots. Much network reconstruction methods are based on investigating unknown relations of known players assuming there are not any unknown players. This might severely bias both the computational and manual reconstruction of underlying biological networks. Here we ask the question, which features of a network can be confounded by incomplete observations and which cannot. In the context of nested effect model based network reconstruction, we show that in the presence of missing observations or hidden factors with their unknown effects (unknown-unknown players), a reliable reconstruction of the full network is not feasible. Nevertheless, we can show that certain characteristics of signaling networks like the existence of cross talk between certain branches of the network can be inferred in a notconfoundable way. We derive a simple polynomial test for inferring such not-confoundable characteristics of signaling networks. We also define a set of edges to partially reconstruct the signaling networks when the unknown players exist and evaluate its performance both on simulated data and in the context of a recent study on Wnt-signaling in colorectal cancer cells.

Institute of Functional Genomics, Computational Diagnostics Group, University of Regensburg

{Mohammed.Sadeh,Rainer.Spang}@klinik.uni-regensburg.de

References

1. Markowetz, F and Bloch, J and Spang, R (2005) Non-transcriptional pathway features reconstructed from secondary effects of RNA interference Bioinformatics 21:4026-32.
2. Markowetz, F and Kostka, D and Troyanskaya, O G and Spang, R (2007) Nested effects models for high- dimensional phenotyping screens Bioinformatics 13:i305-12.
3. Tresch A, Markowetz F (2008) Structure learning in nested effects models. Stat Appl Genet Mol Biol 7.
4. Fröhlich H, Beissbarth T, Tresch A, Kostka D, Jacob J, et al. (2008) Analyzing gene perturbation screens with nested effects models in R and Bioconductor. Bioinformatics 24: 2549–2550.
5. Anchang B, Sadeh M, Jacob J, Tresch A, Vlad M, et al. (2009) Modeling the temporal interplay of molecular signaling and gene expression by using dynamic nested effects models. Proceedings of the National Academy of Sciences 106: 6447.

Visualization and Clustering of Electricity Networks

Sebastian Krey, Dirk Surmann, Uwe Ligges, and Claus Weihs

In modern interconnected powersystems (often comprising whole continents) the control of the power flows to provide an uninterrupted supply with energy is a very complex task. The ongoing replacement of classical power plants (nuclear and coal) with renewable energy provided by highly distributed and relatively small solar, wind and hydrodynamic power stations introduces a whole collection of challenges. Beside the prediction of the generated power the distribution of the renewable energy to the industrial centers is one of the major challenges.

On heavily loaded power supply systems the occurence of low frequency oscillations, introduced by dynamic load changes, can have severe impacts. Therefore it is necessary to understand the behaviour of the network to predict the occurence of such oscillations and detect possible overloads.

As our power systems cover large areas and are heavily interconnected, we cannot use measurements from every possible metering point for the state estimation of the network (computational- and communication time constraints). Hence a selection of important stations for the state of the whole network and local modelling is necessary.

In this work we want to present graphical methods to unterstand the oscillation behaviour of the network, clustering of the network graph to find areas for the fitting of local methods and first ideas for the selection of important metering points.

Fakultät Statistik, TU Dortmund

krey@statistik.tu-dortmund.de

The excess mass approach and diagnostics for modality

Günther Sawitzki

The basic idea of the excess mass approach is to measure the amount of probability mass not fitting a given statistical model. It came up first in the context of testing for a treatment effect, was later applied to inference about the modality of a distribution. The framework has been extended to other problems, such as density estimation or regression.

For testing, given a hypothesis about the modality of a distribution, the excess mass test and related methods give a satisfactory solution.

For exploratory data analysis, with no or only tentative hypotheses, more flexibilityis needed. The shorth plot is a tool in the spirit of the excess mass approach. For the one dimensional situation, the shorth plot is a graphical representation of the length of the shorth, the shortest interval covering a certain fraction of the distribution. Localising the shorth, i.e. requiring it to contain specific data points, makes it usable for diagnostics.

References

1. D.W. Müller, G. Sawitzki: Excess Mass Estimates and Tests for Multimodality. Journal of the American Statistical Association 86 (Sep. 1991), 738–746
2. J. Einmahl, M. Gantner, G. Sawitzki: The Shorth Plot. Journal of Computational and Graphical Statistics March 2010, Vol. 19, No. 1: 62–73.
3. J. Einmahl, M. Gantner, G. Sawitzki: Asymptotics of the Shorth Plot. Journal of Statistical Planning and Inference November 2010, Vol. 140, Issue 11, Pages 3003–3012.

See: http://lshorth.r-forge.r-project.org/

StatLab Heidelberg

gs@statlab.uni-heidelberg.de

A comparison of cumulative/dynamic estimators to evaluate the prediction accuracy of survival models

Matthias Schmid and Sergej Potapov

Time-dependent AUC curves have become a popular tool to evaluate the predictive performance of survival models with right-censored time-to-event outcome. A frequently used approach is to consider cumulative/dynamic AUC curves, which can be computed for each time point t under consideration and which can be used to discriminate between observations having an event at or before t and observations having an event after t.

In the literature, several approaches to estimate cumulative/dynamic AUC curves have been proposed. These approaches rely on a variety of assumptions (such as the proportional hazards assumption in Cox regression) that are needed to guarantee consistency and asymptotic normality of the estimators. In this talk we analyze four popular estimators of cumulative/dynamic AUC curves. Based on the results of a simulation study and the analysis of a set of gene expression data, we investigate the sensitivity of the four estimators with regard to violations of underlying regularity assumptions.

Institut für Medizininformatik, Biometrie und Epidemiologie, Universität Erlangen-Nürnberg

Matthias.Schmid@imbe.med.uni-erlangen.de

Group-based ant colony optimization

Gunnar Völkel[1], Uwe Schöning[1], 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. In general a solution consists of a sequence of solution components. For problems like the Traveling Salesman Problem (TSP) the linear solution en- coding of ACO as a sequence of components works well since a sequence of customers is a natural representation of the visiting order of those customers. The solutions of the Capacitated Vehicle Routing Problem (CVRP), a descendant of the TSP, usually consist of more than one route. A linear solution encoding for the CVRP has to consist of the component sequences of the individual routes interleaved with some end of route component. This is no natural encoding because it favors one route over another whereas the problem does not state such a preference. Generally, this applies to problems with a solution that is sub-structured into independent groups of components. We propose Group-Based Ant Colony Optimization (GBACO) whose solution encoding is sub-structured into groups each consisting of a sequence of components. The modified construction procedure selects one pair of group and component probabilistically and adds the selected component to the selected group. First experiments comparing ACO and GBACO on the commonly used Solomon benchmark instances (VRP with Time Windows) are presented.

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

{Gunnar.Voelkel,Uwe.Schoening,Hans.Kestler}@uni-ulm.de

References

1. Dorigo, M. and Stützle, T. (2009): Ant Colony Optimization: Overview and Recent Advances. Techreport, IRIDIA, Universite Libre de Bruxelles.
2. Gambardella, L. M. and Taillard, E. and Agazzi G. (1999): MACS-VRPTW - A Multiple Colony System For Vehicle Routing Problems With Time Windows. In: Corne, D. and Dorigo, M. et al. (Eds.): New Ideas in Optimization. McGraw-Hill, Maidenhead, England, 63–76.

Efficient Bayesian analysis based on the method "merge and reduce"

Leo Geppert and Katja Ickstadt

In our setting, we want to analyse a massive data set via Bayesian regression. The standard method for Bayesian regression is using Markov Chain Monte Carlo (MCMC) methods, which are demanding in terms of computing time and memory requirements. When dealing with data sets with a massive number of observations ($n >> p$), using standard MCMC-methods quickly becomes unfeasible. To avoid this problem we propose applying the technique of merge and reduce. Merge and reduce has been used in classification and clustering problems for large data sets as well as in streaming settings. In the setting of Bayesian regression, the basic task is to merge two (or more) MCMC samples without increasing the complexity. We present different ideas on how to achieve this. One approach is to consider summary values only. This fairly intuitive idea is taken from meta-analysis, but it might be too simple, depending on the aim of the analysis. As an alternative, we characterise the MCMC samples using core sets. The calculation of the core sets can be implemented in an efficient and scalable way, which makes them especially suitable for massive data sets.

Fakultät für Statistik, TU Dortmund

Geppert@statistik.uni-dortmund.de

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

91-01 *Ker-I Ko, P. Orponen, U. Schöning, O. Watanabe*
Instance Complexity

*91-02** *K. Gladitz, H. Fassbender, H. Vogler*
Compiler-Based Implementation of Syntax-Directed Functional Programming

*91-03** *Alfons Geser*
Relative Termination

*91-04** *J. Köbler, U. Schöning, J. Toran*
Graph Isomorphism is low for PP

91-05 *Johannes Köbler, Thomas Thierauf*
Complexity Restricted Advice Functions

*91-06** *Uwe Schöning*
Recent Highlights in Structural Complexity Theory

*91-07** *F. Green, J. Köbler, J. Toran*
The Power of Middle Bit

*91-08** *V.Arvind, Y. Han, L. Hamachandra, J. Köbler, A. Lozano, M. Mundhenk, A. Ogiwara, U. Schöning, R. Silvestri, T. Thierauf*
Reductions for Sets of Low Information Content

*92-01** *Vikraman Arvind, Johannes Köbler, Martin Mundhenk*
On Bounded Truth-Table and Conjunctive Reductions to Sparse and Tally Sets

*92-02** *Thomas Noll, Heiko Vogler*
Top-down Parsing with Simulataneous Evaluation of Noncircular Attribute Grammars

92-03 *Fakultät für Informatik*
17. Workshop über Komplexitätstheorie, effiziente Algorithmen und Datenstrukturen

*92-04** *V. Arvind, J. Köbler, M. Mundhenk*
Lowness and the Complexity of Sparse and Tally Descriptions

*92-05** *Johannes Köbler*
Locating P/poly Optimally in the Extended Low Hierarchy

*92-06** *Armin Kühnemann, Heiko Vogler*
Synthesized and inherited functions -a new computational model for syntax-directed semantics

*92-07** *Heinz Fassbender, Heiko Vogler*
A Universal Unification Algorithm Based on Unification-Driven Leftmost Outermost Narrowing

*92-08** *Uwe Schöning*
On Random Reductions from Sparse Sets to Tally Sets

*92-09** *Hermann von Hasseln, Laura Martignon*
Consistency in Stochastic Network

92-10 *Michael Schmitt*
A Slightly Improved Upper Bound on the Size of Weights Sufficient to Represent Any Linearly Separable Boolean Function

92-11 *Johannes Köbler, Seinosuke Toda*
On the Power of Generalized MOD-Classes

92-12 *V. Arvind, J. Köbler, M. Mundhenk*
Reliable Reductions, High Sets and Low Sets

92-13 *Alfons Geser*
On a monotonic semantic path ordering

*92-14** *Joost Engelfriet, Heiko Vogler*
The Translation Power of Top-Down Tree-To-Graph Transducers

93-01 *Alfred Lupper, Konrad Froitzheim*
AppleTalk Link Access Protocol basierend auf dem Abstract Personal Communications Manager

93-02 *M.H. Scholl, C. Laasch, C. Rich, H.-J. Schek, M. Tresch*
The COCOON Object Model

93-03 *Thomas Thierauf, Seinosuke Toda, Osamu Watanabe*
On Sets Bounded Truth-Table Reducible to P-selective Sets

93-04 *Jin-Yi Cai, Frederic Green, Thomas Thierauf*
On the Correlation of Symmetric Functions

93-05 *K.Kuhn, M.Reichert, M. Nathe, T. Beuter, C. Heinlein, P. Dadam*
A Conceptual Approach to an Open Hospital Information System

93-06 *Klaus Gaßner*
Rechnerunterstützung für die konzeptuelle Modellierung

93-07 *Ullrich Keßler, Peter Dadam*
Towards Customizable, Flexible Storage Structures for Complex Objects

94-01 *Michael Schmitt*
On the Complexity of Consistency Problems for Neurons with Binary Weights

94-02 *Armin Kühnemann, Heiko Vogler*
A Pumping Lemma for Output Languages of Attributed Tree Transducers

94-03 *Harry Buhrman, Jim Kadin, Thomas Thierauf*
On Functions Computable with Nonadaptive Queries to NP

94-04 *Heinz Faßbender, Heiko Vogler, Andrea Wedel*
Implementation of a Deterministic Partial E-Unification Algorithm for Macro Tree Transducers

94-05 *V. Arvind, J. Köbler, R. Schuler*
On Helping and Interactive Proof Systems

94-06 *Christian Kalus, Peter Dadam*
Incorporating record subtyping into a relational data model

94-07 *Markus Tresch, Marc H. Scholl*
A Classification of Multi-Database Languages

94-08 *Friedrich von Henke, Harald Rueß*
Arbeitstreffen Typtheorie: Zusammenfassung der Beiträge

94-09 *F.W. von Henke, A. Dold, H. Rueß, D. Schwier, M. Strecker*
Construction and Deduction Methods for the Formal Development of Software

94-10 *Axel Dold*
Formalisierung schematischer Algorithmen

94-11 *Johannes Köbler, Osamu Watanabe*
New Collapse Consequences of NP Having Small Circuits

94-12 *Rainer Schuler*
On Average Polynomial Time

94-13 *Rainer Schuler, Osamu Watanabe*
Towards Average-Case Complexity Analysis of NP Optimization Problems

94-14 *Wolfram Schulte, Ton Vullinghs*
Linking Reactive Software to the X-Window System

94-15 *Alfred Lupper*
Namensverwaltung und Adressierung in Distributed Shared Memory-Systemen

94-16 *Robert Regn*
Verteilte Unix-Betriebssysteme

94-17 *Helmuth Partsch*
Again on Recognition and Parsing of Context-Free Grammars:
Two Exercises in Transformational Programming

94-18 *Helmuth Partsch*
Transformational Development of Data-Parallel Algorithms: an Example

95-01 *Oleg Verbitsky*
On the Largest Common Subgraph Problem

95-02 *Uwe Schöning*
Complexity of Presburger Arithmetic with Fixed Quantifier Dimension

95-03 *Harry Buhrman,Thomas Thierauf*
The Complexity of Generating and Checking Proofs of Membership

95-04 *Rainer Schuler, Tomoyuki Yamakami*
Structural Average Case Complexity

95-05 *Klaus Achatz, Wolfram Schulte*
Architecture Indepentent Massive Parallelization of Divide-And-Conquer Algorithms

95-06 *Christoph Karg, Rainer Schuler*
Structure in Average Case Complexity

95-07 *P. Dadam, K. Kuhn, M. Reichert, T. Beuter, M. Nathe*
ADEPT: Ein integrierender Ansatz zur Entwicklung flexibler, zuverlässiger kooperierender Assistenzsysteme in klinischen Anwendungsumgebungen

95-08 *Jürgen Kehrer, Peter Schulthess*
Aufbereitung von gescannten Röntgenbildern zur filmlosen Diagnostik

95-09 *Hans-Jörg Burtschick, Wolfgang Lindner*
On Sets Turing Reducible to P-Selective Sets

95-10 *Boris Hartmann*
Berücksichtigung lokaler Randbedingung bei globaler Zieloptimierung mit neuronalen Netzen am Beispiel Truck Backer-Upper

95-11 *Thomas Beuter, Peter Dadam*:
Prinzipien der Replikationskontrolle in verteilten Systemen

95-12 *Klaus Achatz, Wolfram Schulte*
Massive Parallelization of Divide-and-Conquer Algorithms over Powerlists

95-13 *Andrea Mößle, Heiko Vogler*
Efficient Call-by-value Evaluation Strategy of Primitive Recursive Program Schemes

95-14 *Axel Dold, Friedrich W. von Henke, Holger Pfeifer, Harald Rueß*
A Generic Specification for Verifying Peephole Optimizations

96-01 *Ercüment Canver, Jan-Tecker Gayen, Adam Moik*
Formale Entwicklung der Steuerungssoftware für eine elektrisch ortsbediente Weiche mit VSE

96-02 *Bernhard Nebel*
Solving Hard Qualitative Temporal Reasoning Problems: Evaluating the Efficiency of Using the ORD-Horn Class

96-03 *Ton Vullinghs, Wolfram Schulte, Thilo Schwinn*
An Introduction to TkGofer

96-04 *Thomas Beuter, Peter Dadam*
Anwendungsspezifische Anforderungen an Workflow-Mangement-Systeme am Beispiel der Domäne Concurrent-Engineering

96-05 *Gerhard Schellhorn, Wolfgang Ahrendt*
Verification of a Prolog Compiler - First Steps with KIV

96-06 *Manindra Agrawal, Thomas Thierauf*
Satisfiability Problems

96-07 *Vikraman Arvind, Jacobo Torán*
A nonadaptive NC Checker for Permutation Group Intersection

96-08 *David Cyrluk, Oliver Möller, Harald Rueß*
An Efficient Decision Procedure for a Theory of Fix-Sized Bitvectors with Composition and Extraction

96-09 *Bernd Biechele, Dietmar Ernst, Frank Houdek, Joachim Schmid, Wolfram Schulte*
Erfahrungen bei der Modellierung eingebetteter Systeme mit verschiedenen SA/RT–Ansätzen

96-10 *Falk Bartels, Axel Dold, Friedrich W. von Henke, Holger Pfeifer, Harald Rueß*
Formalizing Fixed-Point Theory in PVS

96-11 *Axel Dold, Friedrich W. von Henke, Holger Pfeifer, Harald Rueß*
Mechanized Semantics of Simple Imperative Programming Constructs

96-12 *Axel Dold, Friedrich W. von Henke, Holger Pfeifer, Harald Rueß*
Generic Compilation Schemes for Simple Programming Constructs

96-13 *Klaus Achatz, Helmuth Partsch*
From Descriptive Specifications to Operational ones: A Powerful Transformation Rule, its Applications and Variants

97-01 *Jochen Messner*
Pattern Matching in Trace Monoids

97-02 *Wolfgang Lindner, Rainer Schuler*
A Small Span Theorem within P

97-03 *Thomas Bauer, Peter Dadam*
A Distributed Execution Environment for Large-Scale Workflow Management Systems with Subnets and Server Migration

97-04 *Christian Heinlein, Peter Dadam*
Interaction Expressions - A Powerful Formalism for Describing Inter-Workflow Dependencies

97-05 *Vikraman Arvind, Johannes Köbler*
On Pseudorandomness and Resource-Bounded Measure

97-06 *Gerhard Partsch*
Punkt-zu-Punkt- und Mehrpunkt-basierende LAN-Integrationsstrategien für den digitalen Mobilfunkstandard DECT

97-07 *Manfred Reichert, Peter Dadam*
$ADEPT_{flex}$ - Supporting Dynamic Changes of Workflows Without Loosing Control

97-08 *Hans Braxmeier, Dietmar Ernst, Andrea Mößle, Heiko Vogler*
The Project NoName - A functional programming language with its development environment

97-09 *Christian Heinlein*
Grundlagen von Interaktionsausdrücken

97-10 *Christian Heinlein*
Graphische Repräsentation von Interaktionsausdrücken

97-11 *Christian Heinlein*
Sprachtheoretische Semantik von Interaktionsausdrücken

97-12 *Gerhard Schellhorn, Wolfgang Reif*
Proving Properties of Finite Enumerations: A Problem Set for Automated Theorem Provers

97-13 *Dietmar Ernst, Frank Houdek, Wolfram Schulte, Thilo Schwinn*
Experimenteller Vergleich statischer und dynamischer Softwareprüfung für eingebettete Systeme

97-14 *Wolfgang Reif, Gerhard Schellhorn*
Theorem Proving in Large Theories

97-15 *Thomas Wennekers*
Asymptotik rekurrenter neuronaler Netze mit zufälligen Kopplungen

97-16 *Peter Dadam, Klaus Kuhn, Manfred Reichert*
Clinical Workflows - The Killer Application for Process-oriented Information Systems?

97-17 *Mohammad Ali Livani, Jörg Kaiser*
EDF Consensus on CAN Bus Access in Dynamic Real-Time Applications

97-18 *Johannes Köbler,Rainer Schuler*
Using Efficient Average-Case Algorithms to Collapse Worst-Case Complexity Classes

98-01 *Daniela Damm, Lutz Claes, Friedrich W. von Henke, Alexander Seitz, Adelinde Uhrmacher, Steffen Wolf*
Ein fallbasiertes System für die Interpretation von Literatur zur Knochenheilung

98-02 *Thomas Bauer, Peter Dadam*
Architekturen für skalierbare Workflow-Management-Systeme - Klassifikation und Analyse

98-03 *Marko Luther, Martin Strecker*
A guided tour through *Typelab*

98-04 *Heiko Neumann, Luiz Pessoa*
Visual Filling-in and Surface Property Reconstruction

98-05 *Ercüment Canver*
Formal Verification of a Coordinated Atomic Action Based Design

98-06 *Andreas Küchler*
On the Correspondence between Neural Folding Architectures and Tree Automata

98-07 *Heiko Neumann, Thorsten Hansen, Luiz Pessoa*
Interaction of ON and OFF Pathways for Visual Contrast Measurement

98-08 *Thomas Wennekers*
Synfire Graphs: From Spike Patterns to Automata of Spiking Neurons

98-09 *Thomas Bauer, Peter Dadam*
Variable Migration von Workflows in *ADEPT*

98-10 *Heiko Neumann, Wolfgang Sepp*
Recurrent V1 – V2 Interaction in Early Visual Boundary Processing

98-11 *Frank Houdek, Dietmar Ernst, Thilo Schwinn*
Prüfen von C–Code und Statmate/Matlab–Spezifikationen: Ein Experiment

98-12 *Gerhard Schellhorn*
Proving Properties of Directed Graphs: A Problem Set for Automated Theorem Provers

98-13 *Gerhard Schellhorn, Wolfgang Reif*
Theorems from Compiler Verification: A Problem Set for Automated Theorem Provers

98-14 *Mohammad Ali Livani*
SHARE: A Transparent Mechanism for Reliable Broadcast Delivery in CAN

98-15 *Mohammad Ali Livani, Jörg Kaiser*
Predictable Atomic Multicast in the Controller Area Network (CAN)

99-01 *Susanne Boll, Wolfgang Klas, Utz Westermann*
A Comparison of Multimedia Document Models Concerning Advanced Requirements

99-02 *Thomas Bauer, Peter Dadam*
Verteilungsmodelle für Workflow-Management-Systeme - Klassifikation und Simulation

99-03 *Uwe Schöning*
On the Complexity of Constraint Satisfaction

99-04 *Ercument Canver*
Model-Checking zur Analyse von Message Sequence Charts über Statecharts

99-05 *Johannes Köbler, Wolfgang Lindner, Rainer Schuler*
Derandomizing RP if Boolean Circuits are not Learnable

99-06 *Utz Westermann, Wolfgang Klas*
Architecture of a DataBlade Module for the Integrated Management of Multimedia Assets

99-07 *Peter Dadam, Manfred Reichert*
Enterprise-wide and Cross-enterprise Workflow Management: Concepts, Systems, Applications. Paderborn, Germany, October 6, 1999, GI–Workshop Proceedings, Informatik '99

99-08 *Vikraman Arvind, Johannes Köbler*
Graph Isomorphism is Low for ZPP^{NP} and other Lowness results

99-09 *Thomas Bauer, Peter Dadam*
Efficient Distributed Workflow Management Based on Variable Server Assignments

2000-02 *Thomas Bauer, Peter Dadam*
Variable Serverzuordnungen und komplexe Bearbeiterzuordnungen im Workflow-Management-System ADEPT

2000-03 *Gregory Baratoff, Christian Toepfer, Heiko Neumann*
Combined space-variant maps for optical flow based navigation

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

2011-01 *Stephan Buchwald, Thomas Bauer, Manfred Reichert*
Flexibilisierung Service-orientierter Architekturen

2011-02 *Johannes Hanika, Holger Dammertz, Hendrik Lensch*
Edge-Optimized À-Trous Wavelets for Local Contrast Enhancement with Robust Denoising

2011-03 *Stefanie Kaiser, Manfred Reichert*
Datenflussvarianten in Prozessmodellen: Szenarien, Herausforderungen, Ansätze

2011-04 *Hans A. Kestler, Harald Binder, Matthias Schmid, Friedrich Leisch, Johann M. Kraus (eds):*
Statistical Computing 2011 - Abstracts der 43. Arbeitstagung

2011-05 *Vera Künzle, Manfred Reichert*
PHILharmonicFlows: Research and Design Methodology

2011-06 *David Knuplesch, Manfred Reichert*
Ensuring Business Process Compliance Along the Process Life Cycle

2011-07 *Marcel Dausend*
Towards a UML Profile on Formal Semantics for Modeling Multimodal Interactive Systems

2011-08 *Dominik Gessenharter*
Model-Driven Software Development with ACTIVECHARTS - A Case Study

2012-01 *Andreas Steigmiller, Thorsten Liebig, Birte Glimm*
Extended Caching, Backjumping and Merging for Expressive Description Logics

2012-02 *Hans A. Kestler, Harald Binder, Matthias Schmid, Johann M. Kraus (eds):*
Statistical Computing 2012 - Abstracts der 44. Arbeitstagung

Ulmer Informatik-Berichte
ISSN 0939-5091

Herausgeber:
Universität Ulm
Fakultät für Ingenieurwissenschaften und Informatik
89069 Ulm