ulm university universität ulm

Statistical Computing 2011

Abstracts der 43. Arbeitstagung

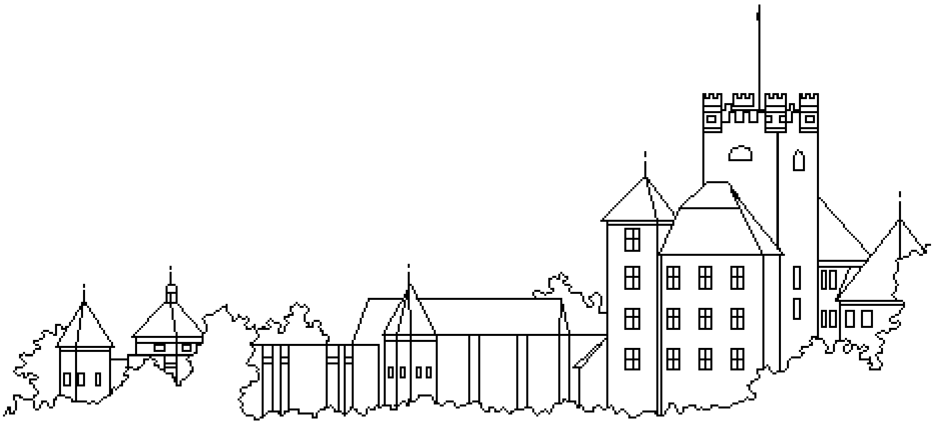
HA Kestler, H Binder, M Schmid
F Leisch, JM Kraus (eds)

Ulmer Informatik-Berichte

Nr. 2011-04
Mai 2011

International Graduate School
in Molecular Medicine Ulm

Statistical Computing 2011



43. Arbeitstagung

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

22.05.-25.05.2011, Schloss Reisensburg (Günzburg)

Workshop Program

Sunday, May 22, 2011

18:15-20:00		**Dinner**
20:00-21:00		**Chair: H.A. Kestler (Ulm)**
20:00-21:00	Wolfgang Huber (Heidelberg)	Differential expression analysis for sequence count data

Monday, May 23, 2011

8:50		**Opening of the workshop: H.A. Kestler, H. Binder**
09:00-12:00		**Chair: A. Benner (Heidelberg)**
09:00-09:30	Stefanie Hieke (Freiburg)	Min P-test: a resampling based gene region-level testing procedure for genetic case-control studies implemented in R
09:30-10:00	Michel Lang (Dortmund)	Survival models with preclustered gene groups as covariates
10:00-10:30	Aslihan Gerhold-Ay (Mainz)	Evaluation and validation of gene expression signatures for prognostic use in negative breast cancer patients
10:30-11:00		**Coffee break**
11:00-11:30	Khalid Abnaof (Bonn)	Differential expression analysis and cluster method for time course Microarray data
11:30-12:00	Lea Vaas (Braunschweig)	Phenotype Microarray - Data organization and analysis of respiration curve
12:15-14:00		**Lunch**
14:00-18:00		**Chair: B. Lausen (Essex)**
14:00-15:00	Luc De Raedt (Leuven)	Analyzing Structured Data - Symbolic and Probabilistic Approaches
15:00-15:30	Alfred Ultsch (Marburg)	Association of complex human pain phenotypes with complex pain genotypes using a self-organizing maps approach
15:30-16:00	Miriam Schmidt (Ulm)	Spectral graph features for the classification of graphs and graph sequences
16:00-16:30		**Coffee break**
16:30-18:00	Benjamin Hofner, Andreas Mayr, Matthias Schmid (Erlangen), Nikolay Robinzonov (München)	**Tutorial I**: “mboost - Model based Boosting in R”
18:15-20:00		**Dinner**
20:00-21:00		**Tutorial II**: “mboost - Model based Boosting in R”

Tuesday, May 24, 2011

09:00-12:00		**Chair: U. Ligges (Dortmund)**
09:00-09:30	Alexander Melkozerov (Ulm)	Tuning distance measures in k-nearest neighbor classification via evolution strategies
09:30-10:00	Alexander Pilhöfer (Augsburg)	optile: Optimizing k-dimensional graphical classification analysis via category reordering
10:00-10:30	Sebastian Krey (Dortmund)	Order constrained clustering for music structure analysis
10:30-11:00		**Coffee break**
11:00-11:30	Martin Hopfensitz (Ulm)	Inference of Boolean networks by fuzzy sets
11:30-12:00	Markus Maucher (Ulm)	Inferring Boolean network structure via correlation
12:15-14:00		**Lunch**
14:00-18:00		**Chair: H. Binder (Freiburg)**
14:00-15:00	Berthold Lausen (Essex)	Meta-analysis methods for gene expression profiles
15:00-15:30	Klaus Jung (Göttingen)	Connecting miRNA and mRNA expression profiles for medical classification problems
15:30-16:00	Manuela Zucknick (Heidelberg)	Integration of copy number variation and gene expression data in Bayesian models for prediction
16:00-16:30		**Coffee break**
16:30-17:00	Ludwig Lausser (Ulm)	On the utility of partially labeled data for classification in high-dimensional settings
17:00-18:00		Working groups meeting on **Statistical Computing 2012** and other topics (all welcome)
18:15-20:00		**Dinner**

Wednesday, May 25, 2011

09:00-12:00		**Chair: M. Schmid (Erlangen)**
09:00-10:30	Christoph Müssel, Ludwig Lausser, Markus Maucher, Hans A. Kestler (Ulm) Sergej Potapov, Werner Adler (Erlangen) Berthold Lausen (Essex)	**Hands-on Software Demos:** I) Multi-objective parameter selection for classifiers II) The Daim package - Diagnostic accuracy of classification models
10:30-11:00		**Coffee break**
11:00-11:30	Christoph Bernau (München)	Correcting the optimally selected resampling-based error rate: A smooth analytical alternative to nested cross-validation
11:30-12:00	Julia Schiffner (Dortmund)	Bias-variance analysis of local classification methods
12:15-14:00		**Lunch**

Differential expression analysis for sequence count data

Wolfgang Huber

EMBL Heidelberg
wolfgang.huber@embl.de

High-throughput DNA sequencing is a powerful and versatile new technology for obtaining comprehensive and quantitative data about RNA expression (RNA-Seq), protein-DNA binding (ChIP-Seq), and genetic variations between individuals. It addresses essentially all of the use cases that microarrays were applied to in the past, but produces more detailed and more comprehensive results.

One of the basic statistical tasks is inference (testing, regression) on discrete count values (e.g., representing the number of times a certain type of mRNA was sampled by the sequencing machine). Challenges are posed by a large dynamic range, heteroskedasticity and small numbers of replicates. Hence, model-based approaches are needed to achieve statistical power.

I will present an error model that uses the negative binomial distribution, with variance and mean linked by local regression, to model the null distribution of the count data. The method controls type-I error and provides good detection power. I will also discuss how to use the GLM framework to detect alternative transcript isoform usage. A free open-source R software package, DESeq, is available from the Bioconductor project.

* joint work with Simon Anders

Min P test: a resampling based gene region-level testing procedure for genetic case-control studies implemented in R

Stefanie Hieke, Harald Binder, Alexandra Nieters
and Martin Schumacher

Institute of Medical Biometry and Medical Informatics,
Center of Chronic Immunodeficieny,
University Medical Center Freiburg,
Freiburg Center for Data Analysis and Modeling,
University Freiburg
hieke@imbi.uni-freiburg.de

Introduction Current technologies generate a huge number of single nucleotide polymorphism (SNP) genotype measurements in case-control studies. The resulting multiple testing problem can be ameliorated by considering candidate gene regions. The minPtest R package provides the first widely accessible implementation of a gene region-level summary for each candidate gene using the min P test.

Method The gene region-level summary, as the min P test, assesses the statistical significance of the smallest p-trend within each gene region and, therefore, considers a reduced number of tests. The min P test is a permutation-based method that can be based on several univariate tests per SNP. In permutation resampling, the observed variable (case/control status) is randomly re-assigned without replacement to "pseudo case/control status". A test statistic is then recomputed using the pseudo data and compared to the marginal test statistic in the original data set. This procedure is repeated B times. The inference is based on the permutation distribution of the minimum of the ordered p-values from the marginal test of each SNP. The gene region-level summary is mostly compatible with univariate statistical tests per SNP conducted separately over multiple loci.

Results Combining the p-values from tests in a permutations-based approach prevents an increase of the false-positive rates, as correlations of SNPs are automatically taken into account. We developed an R package that brings together three different kinds of tests that are scattered over several R packages, and automatically selects the most appropriate one for the design at hand. The implementation in the minPtest package integrates two different parallel computing packages, thus optimally leveraging available resources for speedy results. The package comprises a function to simulate SNP data with known structure, allowing the user to explore different scenarios and settings.

Conclusion The minPtest package provides a useful and feasible implementation of a gene region-level summary, using the min P test, controlling the false-positive rate and having higher power. In addition minPtest provides acceleration by parallel computing.

References

Chen,B.E. et al. (2006). Resampling-based multiple hypothesis testing procedures for genetic case-control association studies. Genetic Epidemiology, 30, 495-507.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing. ISBN 3-900051-07-0. url = http://www.R-project.org.

Westfall,P.H. et al.(2002). Multiple tests for genetic effects in association studies. Methods Mol Biol, 184, 143-168.

Westfall,P.H. and Young,S.S. (1993). Resampling-Based Multiple Testing: Exam- ple and Methods for p-Value Adjustment. Wiley, New York.

Survival models with preclustered gene groups as covariates

K. Kammers, M. Lang and J. Rahnenführer

Departments of Statistics,
TU Dortmund University
lang@statistik.tu-dortmund.de

An important application of high dimensional gene expression measurements is the prediction of survival times and the interpretation of the variables in the resulting regression models. When the response variables are censored survival times, an appropriate hazard framework is required. The largest problem in this context is the typically large number of genes compared to the number of observations (individuals). We thus apply feature selection procedures to construct predictive models for future patients. This approach aims at identifying models with high prediction accuracy and at the same time low model complexity. However, interpretability of the resulting models is still limited due to little knowledge on many of the remaining selected genes. In order to improve the interpretability of the estimated models, we summarize genes as gene groups defined by the hierarchically structured Gene Ontology (GO) and include these gene groups as covariates in the hazard regression models. Though the expression profiles present in GO groups are often heterogeneous, leading to several different expression profiles within one group. Preclustering genes within GO groups according to the correlation of their gene expression measurements leads to homogeneous subclasses. This allows the aggregation of each subclass to single covariates with predictive importance as well as, as a result of GO annotations, additional interpret- ability. Besides the genomic data, we include clinical information to reveal the real benefit of the preclustered genomic models. To evaluate the prediction performance of the models, we examine both Brier scores and p-values derived from the prognostic index in a nested cross-validation setup. Survival models with preclustered gene groups as covariates have similar prediction accuracy to models built only with single genes. Using only gene groups as covariates can lead to decreased prediction accuracy since many genes are not yet annotated to any corresponding function. However, integrating the preclustering information improves the interpretability of the models while prediction performance remains stable.

Evaluation and validation of gene expression signatures for prognostic use in node negative breast cancer patients

Aslihan Gerhold-Ay, Anja Victor and Marcus Schmidt

Institute of Medical Biostatistics, Epidemiology and Informatics,
University Medical Center of the Johannes Gutenberg University Mainz,
Merck KGaA, Darmstadt,
Department of Obstetrics and Gynaecology,
University Medical Center of the Johannes Gutenberg University Mainz
aslihan.gerhold-ay@unimedizin-mainz.de

Introduction The most widely used treatment guidelines for breast cancer are based on classical risk factors like the St. Gallen classification. The guidelines recommend adjuvant systemic therapy for almost all breast cancer patients because this therapy has greatly improved survival in early breast cancer. However, adjuvant therapy also has a lot of negative effects with respect to quality of life. For this reason there is a need to specify an individual risk profile for each patient to avoid over- as well as under treatment. To get useful risk profiles different predictors based on patients gene expression have been developed for breast cancer (1; 2; 3; 4; 5). Furthermore, two gene expression predictors are currently tested in prospective clinical trial (6; 7). The aim of our project is the evaluation and validation of these well known signatures on a cohort of Mainz.

Methods The cohort of Mainz consist of 199 node-negative breast cancer patients treated between 1989 and 1998 at the Department of Obstetrics and Gynaecology, Medical Center of the Johannes Gutenberg University Mainz. All patients were treated with surgery and did not receive any systemic therapy. Data that have been collected are classical risk factors and in addition the gene expression data from the Affmetrix chip HG-U133A (8). To analyse the effect of a signature on survival we apply log-rank test and uni- and multivariate Cox-regression. ROC curves with distant metastasis within 5 years as the defined endpoint were used to describe the quality of the signatures classification into low- and high risk group. Cluster analyses were performed to identify the intrinsic subtypes of breast cancer. Currently simulations are initiated to analyse the stability of the intrinsic subtype signature (3; 4; 5), which is based on previously reported molecular subtypes of breast cancer. Furthermore approaches were identified to develop a new tumor grade signature based on gene expression data. About half of the breast cancers are assigned histological grade 1 or 3. The other berast tumors are classified as histological grade 2, which is not informative for clinical decision mak- ing because of the intermediate risk of recurrence. To increase the prognostic value of tumor grade 2 new methods are necessary to classify them to tumor grade 1 or tumor grade 3.

Results The Mainz cohort is similar to the described populations used for the gene signature development with respect to classical risk factors. Not all of the published prognostic values of the gene signatures could be validated on the cohort of Mainz.

Discussion Gene signatures can provide a powerful tool for identification of patients with high risk of recurrence. Many potential sources of bias (dye bias, sampling bias, time lag bias and publication bias (9; 10)) can make the transmission of the methods into practice difficult. Based on our results we recommend prospective studies to test the validity of the signatures.

References

[1] Y Wang, J G M Klijn, Y Zhang, A M Sieuwerts, M P Look, F Yang, D Talantov, M Timmermans, M E Meijer-van Gelder, J Yu, T Jatkoe, E M J J Berns, D Atkins, J A Foekens. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005; 365:671–79.

[2] C Sotiriou, P Wirapati, S Loi, A Harris, S Fox, J Smeds, H Nordgren, P Farmer, V Praz, B Haibe- Kains, C Desmedt, D Larsimont, F Cardoso, H Peterse, D Nuyten, M Buyse, M J Van de Vijver, J Bergh, M Piccart, M Delorenzi. Gene-expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade To Improve Prognosis. J Natl Cancer Inst. 2006; 98:262-72.

[3] C M Perou, T Srlie, M B Eisen, M va de Rijn, S Jeffrey, C A Rees, J R Pollack, D T Ross, H Johnsen, L A Akslen, O Fluge, A Pergamenschikov, C Williams, S X Zhu, P E Lnning, A L Brresen-Dale, P O Brown,D Botstein. Molecular portraits of human breast tumors Nature 2000; 406(6797):747-52.

[4] Z Hu, C Fan, D S Oh, J S Marron, X He, B F Qaqish, C Livasy, L A Carey, E Reynolds, L Dressler, A Nobel, J Parker, M G Ewend, L R Sawyer, J Wu, Y Liu, R N, M Tretiakova, A Ruiz Orrico, D Dreher, J P Palazzo,L Perreard, E Nelson, M Mone, H Hansen, M Mullins, J F Quackenbush, M J Ellis, O I Olopade, P S Bernard, C M Perou. The molecular Portraits of Breast Tumors Are Conserved Across Microarray Platforms.

BMC Genomics. 2006; 7:96.

[5] M Smid, Y Wang, Y Zhang, A M Sieuwerts, J Yu, J G M Klijn, J A Foekens, J W M Martens. Subtypes of Breast Cancer Show Preferential Site of relapse. Cancer Res. 2008; 1;68(9):3108-14.

[6] S Paik, S Shak, G Tang, C Kim, J Baker, M Cronin, F L Baehner, M G Walker, D Watson, T Park, W Hiller, E R Fisher, D Wickerham, J Bryant, N Wolmark. A multigeneassay to predict recurrence of tamoxifen -treated, node-nagtive breast cancer N Engl J Med 2004, 351(27):2817- 2826.

[7] MJ van de Vijver,YD He, LJ vant Veer,H Dai, AAM Hart, DW Voskuil, GJ Schreiber,JL Peterse, C Roberts, M J Marton, M Parrish, D Atsma, A Witteveen, A Glas, L Delahaye, T van der Velde, H Bartelink, S Rodenhuis, E T Rutgers, S H Friend, R Bernards. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002, 347(25):1999-2009.

[8] M Schmidt, D Bhm, C von Trne, E Steiner, A Puhl, H Pilch, H Lehr, J G Hengstler, H Klbl, M Gehrmann. The Humoral Immune System Has a Key Prognostic Impact in Node-Negative Breast. Cancer. Cancer Research 2008; 68:5405–5413.

[9] K K Dobbin, E S Kawasaki, D W Petersen, R M Simon. Characterizing dye bias in microarray expreiments. Bioinformatics 2005; 15;21(10):2430-7.

[10] J P Ioannidis, E E Ntzani, T A Trikalinos, D G Contopoulos-Ioannidis. Replication validity of genetic association studies. Nat Genet. 2001; 29(3):306-9.

Differential Expression Analysis and Cluster Method for Time Course Microarray Data

Khalid A. Abnaof and Holger Fröhlich

Bonn-Aachen International Center for Information Technology (B-IT),
Bonn University
Institute of Molecular Biotechnology
RWTH Aachen University
abnaof@bit.uni-bonn.de

Understanding the mechanism by which transcription factors dynamically regulate genes and other transcription factors (TF) in multicellular organisms is a very important and interesting task in the research activities of molecular biology. However this task is not easy to tackle, as the dynamic process underlying this regulatory system is very complex. Here, we were particularly interested in TF-target gene networks (transcriptional programs) in multipotent progenitor (MPP) and common dendritic progenitor cells (CDP) in mice, which are dependent on TGF-β stimulation.

The data used in this study consisted of time series microarray data at six time points. We applied a Bayesian approach to determine differentially expressed time series between stimulated and unstimulated cells and between cell types (1). A logistic regression model was used to perform an analysis of differential expression on pathway level (2). Afterwards time courses for each cell type were grouped into clusters utilizing an EM based approach, that describes mean curves within a cluster via smoothing spline models (3). This approach, unlike conventional clustering methods, considers the dynamics of expressions changes. Further analysis of enriched transcription factor binding sites (TFBS) within clusters allowed for partial reconstruction of gene regulatory modules. Hypotheses on the dependencies between these gene regulatory modules may be derived in the future via Dynamic Bayesian Networks (DBNs). A meta-analysis of TFBS enriched in MPPs, but not in CDPs points towards gene regulatory mechanisms, that drive stem cell development in dendritic progenitor cells.

References

Martin J. Aryee, José A. Gutiérrez-Pabello, Igor Kramnik, Tapabrata Maiti and John Quackenbush (2009): An improved empirical bayes approach to estimating differential gene expression in microarray time-course data. BMC Bioinformatics, 9(1).

Montaner D, Dopazo J, 2010 Multidimensional Gene Set Analysis of Genomic Data. PLoS ONE 5(4): e10348. doi:10.1371/journal.pone.0010348.

Ping Ma, Cristian I. Castillo-davis , Wenxuan Zhong and Jun S. Liu (2006): A data-driven clustering method for time course gene expression data. Nucleic Acids Research, 34(4):1261-1269.

Phenotype Microarray – Data organisation and analysis of respiration curves

Lea A. I. Vaas, Johannes Sikorski and Markus Göker

DSMZ German Collection of Microorganisms and Cell Cultures GmbH,
Braunschweig
Lea.Vaas@dsmz.de

Recently, the set of techniques generating so called -omics data was augmented by yet another one, Phenotype Microarrays (PM). In contrast to the existing major technologies, i.e. DNA Microarrays, 2D-Proteomic and chromatographic applications, PM monitors cell respiration over time. Through a redox reaction that alters the colour of a tetrazolium dye in the presence of respiration, kinetic response curves are generated. This provides a high throughput means to characterize microbial metabolism. Yet, the system consists of about 2000 assays for monitoring the cells respiration in the presence of macro- and micronutrients or their reactions to osmotic stress factors, ion or pH effects. The application of a number of chemicals, such as antibiotics, antimetabolites, membrane-active agents, respiratory inhibitors and toxic metals to investigate the cells sensitivity is also possible. Beside the application in identification and drug screening scenarios, where mainly presence-absence calls from each assay are of interest, many research projects emphasize the interest in more sophisticated comparisons of phenotypes of different strains, isolates, mutants, etc. The in-depth evaluation of redox kinetics should gain knowledge about the metabolic differences and provide indications on which genetic features of the investigated organisms these differences are based. The main steps in those analyses would be (1) data organisation and graphical presentation of the curves, (2) application of methods for the reliable estimation of the growth parameters of each curve, and (3) extraction and comparison of such growth curve characteristics. Based on a summary of the available statistical tools covering large parts of the demanded analyses, an application strategy of these ready-to-use software tools for the data analysis of the PM data in R will be proposed. In addition to the statistical challenges this new type of high dimensional data brings along, this talk will give an outlook on features to be provided for a more convenient data analysis pipeline comprehending the organisation of meta-data, concepts for handling the raw data, data analysis and presentation of the results.

Analysing structured data – symbolic and probabilistic appraoches.

Luc De Raedt

Katholieke Universiteit Leuven,
luc.deraedt@cs.kuleuven.be

Structured data in the form of graphs, networks or relational databases are omnipresent across numerous application-areas such as biology, chemistry, the internet, social or bibliographic networks, robotics, vision, etc. The machine learning and data mining literature has devoted a lot of attention to coping with such data giving rise to a class of techniques that is known under the name of graph and network mining or relational learning. In this talk, an introduction will be given to this class of techniques, which often extend or upgrade more traditional techniques for dealing with "flat" data (that is, data in feature vector format) towards graph-based and relational data. This talk will introduce and motivate the problems and techniques of relational and graph-based learning and look into both symbolic and probabilistic methods. Symbolic methods attempt to identify patterns in the form of subgraphs in graph data (e.g., in molecular datasets) while probabilistic methods extent graphical models (like Bayesian and Markov networks) for dealing with relational data. The approaches will be illustrated and motivated by several real-life examples.

Association of complex human pain phenotypes with complex pain genotypes using a self-organizing maps approach

Jörn Lötsch and Alfred Ultsch

pharmazentrum frankfurt/ZAFES, Institute of Clinical Pharmacology,
Johann Wolfgang Goethe University Hospital,
Data Bionics Research Group,
University of Marburg
ultsch@mathematik.uni-marburg.de

BACKGROUND Pain is a complex trait. While clinical pain syndromes can already be diagnosed by a set of neurological parameters, the complexity of experimental pain is only incompletely accounted for, which often impedes associations of pain data with clinical or genetic parameters.
METHODS Pain phenotype markers (n = 8) and genotype markers (n = 30) were available from previous assessments in 125 healthy volunteers. A U-Matrix on an Emergent Self organizing map (ESOM) was used for visualization of the distance structures in the data. Subsequently, the prediction of the clusters by the genetic markers was assessed using a classification and regression tree (CART) approach.
RESULTS On the U-Matrix of the pain phenotypes, eight clusters were identified. This clustering showed advantages over a Ward clustering on the same data. Rules could be derived to describe the cluster contents that corresponded to three basic types of pain thresholds: low, mean and high sensitivity. In the mean and low sensitivity (stoical) phenotypes, subgroups could be identified. A cluster consisting of persons with high overall pain threshold but selectively low resistance to heat, the predictive accuracy of the classifiers was 84.56%. Among the genetic variants that were used for the CART decision in that cluster were polymorphisms in TRPV1, a gene coding for a heat sensor.
CONCLUSIONS ESOM-based clustering of pain data provides biologically meaningful results and satisfying the complexity of pain. The thus obtained clusters seem to facilitate the otherwise only insufficiently successful genotype phenotype association in common pain.

Spectral graph features for the classification of graphs and graph sequences

Miriam Schmidt, Günther Palm and Friedhelm Schwenker

Institute of Neural Information Processing,
University of Ulm
{miriam.k.schmidt,friedhelm.schwenker}@uni-ulm.de

Spectral graph theory is an important branch in the area of graph classification. Matrices associated with graphs such as the adjacency matrix or the Laplacian matrix contain essential information about graph connectivity (Cvetković 1998). In this study the power of the principal eigenvalues of adjacency matrices for classification tasks is investigated. In order to illustrate the proposed method, a toy problem to classify 2D-objects has been defined. The goal was to discriminate between two classes: circles and squares. An object is represented by a set of 2D points, describing the object's outer shape. These points are considered as the graph's nodes, and the graph's adjacency matrix is defined through the pairwise Euclidean distances of the points. From this adjacency matrix principal eigenvalues are computed as the object's characteristic features. This method is then evaluated on a problem of optical character recognition. For this, the capital letters data set from Bern repository of graph data sets (see Riesen et al 2008) has been selected. Additionally, this method has been applied to the problem of human activity recognition based on sequences of camera images. In this task, hidden Markov models are modeling the sequential structure of the data. In the first step locations of the person's body parts (hand, head, etc.) and objects (table, cup, etc.), relevant for the human activity, have to be estimated in each camera image. Subsequently distances between all pairs of detected objects and body parts are computed and the eigenvalues of this Euclidean distance matrix are calculated. These eigenvalues serve as inputs to Gaussian mixture models estimating the emission probabilities of the hidden Markov models.

References

Cvetković, D.M., Doob, M., Horst, S.: Spectra of Graphs. Theory and Applications. Vch Verlagsgesellschaft Mbh (1998)

Riesen, K., Bunke, H.: IAM Graph Database Repository for Graph Based Pattern Recognition and Machine Learning. Structural, Syntactic, and Statistical Pattern Recognition (LNCS 5342), 287–297 (2008)

Tuning distance measures in k-nearest neighbor classification via evolution strategies

Alexander Melkozerov, Ludwig Lausser and Hans A. Kestler

Institute of Neural Information Processing,
University of Ulm
{alexander.melkozerov,ludwig.lausser,hans.kestler}@uni-ulm.de

Molecular high-throughput technologies usually generate data with a high dimensionality and a low cardinality. In the context of classification this poses a serious problem as many classical (model based) methods turn out to be too complex for this task. This stimulated the development of new classifiers that are of a lower complexity thus attaining higher generalization performance by additional regularization terms and more rigid model assumptions.

Some classification methods completely omit the usage of model assumptions. For example the transductive k nearest neighbor ($k-NN$) classifiers directly predict the class label of a datapoint x according to the k training examples closest to x. The performance of this technique is always coupled to the chosen distance measure, which is often, due to the lack of a better choice, the Euclidian one. Other, non-standard distance measures may be more suitable for this task. Here, we investigate the usability of optimized distance measures of type

$$d_{\vec{w}}(\vec{x},\vec{y}) = \sqrt{\sum_{i=1}^{n} w_i(x_i - y_i)^2}$$

for $k-NN$ classification. The weights $\vec{w}$ are optimized to minimize the empirical risk of the current classifier. Two types of evolutionary strategies (ES) were utilized: the standard self-adaptive ES with intermediate recombination (the so-called $(\mu/\mu_I, \lambda)$-σSA-ES) and the covariance matrix adaptation ES (the (μ_W, λ)-CMA-ES). While the former is a simple ES which provides baseline performance and serves as a reference in our comparison, the (μ_W, λ)-CMA-ES is the state-of-the-art algorithm for continuos optimization showing very good performance in recent experimental benchmarks. Observing that the fitness function under consideration is multi-modal, the following restart strategy was used for the $(\mu/\mu_I, \lambda)$-σSA-ES: if no improvement of the best fitness function value occurs for 300 generation or the mutation strength gets too small, the ES starts the search again from a random point.

The advanced (μ_W, λ)-CMA-ES uses the following restart criteria in addition to the check of the best fitness function value improvement:

- the standard deviation $\sigma^{(g)}$ of the normal distribution used to sample new points or evolution path is smaller than given value;
- numerical precision problems: the mean $\langle \mathbf{y} \rangle^{(g)}$ of newly sampled points does not change when adding to $\langle \mathbf{y} \rangle^{(g)}$ a $0.1\sigma^{(g)}$-vector in a principal axis direction of the covariance matrix or a $0.2\sigma^{(g)}$ to each coordinate of $\langle \mathbf{y} \rangle^{(g)}$;
- the condition number of the covariance matrix is too large.

After each restart, the (μ_W, λ)-CMA-ES runs from a random point with the population size increased by factor of 2.

The performance of these modified $k - NN$ techniques is investigated within a comparative study on different microarray datasets. The new classifiers are compared to other well known classification techniques on their generalization ability, robustness and sparsity.

optile: Optimizing k-dimensional graphical classification analysis via category reordering

Alexander Pilhöfer and Alexander Gribov

Department of Computer Oriented Statistics and Data Analysis,
Institute of Mathematics,
University of Augsburg
alexander.pilhoefer@math.uni-augsburg.de

In cluster analysis it is good practice to regard several different clustering methods instead of focusing on only one type. The interest lies in the agreements as well as the differences between the clusterings. A good way to interpret the results is to use visualizations such as fluctuation diagrams or categorical parallel coordinate plots Pilhoefer and Unwin (2011). Clustering classifications usually are of a nominal categorical structure and therefore the variable orders can be changed to improve the displays and make their interpretation easier. Different seriation methods (see Chen et al., 2008) have been proposed to improve the displays for 2-dimensional problems, mostly using one-mode optimizations like the Anti-Robinson-Criterion in distance matrices which do not directly account for the associations between the variables. The talk will present a family of criteria and related optimization algorithms which can be used to choose the category orders for 2- and k-dimensional categorical classification data with respect to their multidimensional associations using the concept of agreement in a pseudo-diagonal form. An effective optimization algorithm will be presented and the applicability to both table-like plots such as fluctuation diagrams and line-based plots such as parallel coordinates plots will be discussed. A special modification of the algorithm which takes account of hierarchical classification structures will be presented using implementations in the software package Seurat Gribov (2010). The talk will use real data clustering results from the US Current Population Survey (CPS) for illustration.

References

Chen, C.-h., W. Haerdle, A. Unwin, H.-M. Wu, S. Tzeng, and C.-h. Chen (2008). Matrix visualization. In Handbook of Data Visualization, Springer Handbooks Comp.Statistics, pp. 681708. Springer Berlin Heidelberg.

Gribov, A. (2010, June). Seurat - visual analytics for the integrated analysis of microarray data. http: //seurat.r-forge.r-project.org/. Hofmann, H. (2000). Exploring categorical data: Interactive mosaic plots. Metrika 51(1), 1126.

Pilhoefer, A. and A. Unwin (2011). Multiple barcharts for relative frequencies and parallel coordinates plots for categorical data - package extracat. Journal of Statistical Software. submitted.

Order constrained clustering for music structure analysis

Sebastian Krey, Uwe Ligges and Friedrich Leisch

Fakultät Statistik,
Technische Universität Dortmund,
Universität für Bodenkultur Wien
krey@statistik.tu-dortmund.de

In music structure analysis, unsupervised machine learning methods are desirable to get a first overivew and to segment into parts that may be relevant for further analyses.

Traditional unconstrained clustering methods may yield unstable and uninterpretable results, particularly when used on sound features derived of recordings of real music. Therefore, it is helpful to constrain the possible solutions in a way that frequently alternating cluster assignments are suppressed.

One intuitive constraint is the temporal order of the recording which implies that a sensible cluster consists of connected time segments. Steinley and Hubert [1] describe a method to introduce an order constraint in clustering. Using an R implementation [2,3] of this idea, we get promising results. Due to the exponential runtime in the number of clusters, we propose a recursive tree-based approach of order constrained clustering with small cluster numbers (e.g. 2).

This way, on a piece of popular music, it is possible to get clusters which represent musical parts like intro, verse, refrain, bridge. This is promising in the sense that it is typically more benefical to train a musical recognition system with a characteristic part of a song rather than with artifical chosen segements.

In addition, it is possible to segment separate tones into attack, sustain, decay and silence - without splitting some of these tone phases into several clusters.

References

[1] Douglas Steinley, Lawrence Hubert (2008). "Order-constrained solutions in k-means clustering: Even better than being globally optimal", Psychometrika, Vol. 73, No. 5, pp. 647-664.

[2] Sebastian Hoffmeister (2009). ”Partitionierende Clusterverfahren unter Ordnungs-Nebenbedingungen”, Diplomarbeit, Institut für Statistik, Ludwig-Maximilians-Universität München.

[3] Friedrich Leisch (2006). ”A Toolbox for K-Centroids Cluster Analysis”, Computational Statistics and Data Analysis, Vol. 51, No. 2, pp. 526-544.

Inference of Boolean networks by fuzzy sets

Martin Hopfensitz, Markus Maucher and Hans A. Kestler

Internal Medicine I,
University Hospital Ulm,
Institute of Neural Information Processing,
Ulm University
{martin.hopfensitz, markus.maucher, hans.kestler}@uni-ulm.de

Molecular systems biology usually refers to integrated experimental and computational approaches for studying biomolecular networks, such as signal transduction, gene regulation or metabolic systems. At the core of systems biology research lies the identification of gene-regulatory networks from experimental data via reverse-engineering methods. Network inference algorithms can assist life scientists in unraveling gene-regulatory systems on a molecular level. In this context, Boolean networks (Kaufmann, 1969) provide a well founded framework for reverse-engineering and analysis of gene-regulatory networks (Hickman et al., 2009). In a Boolean network, a gene is modeled as a Boolean variable that can attain two alternative levels: expressed (1) or not expressed (0). In spite of this restriction, the behaviour of real genetic networks can be described well by this "coarse-grained" model (Bornholdt, 2005). To infer a Boolean network solely from quantitative time series data, the continuous data have to be binarized. But the binarization is often unreliable, since noise on gene expression data and the low number of temporal measurement points frequently lead to an uncertain binarization of values. We developed a novel reverse-engineering method based on Boolean networks that incorporates this uncertainty in the binarized data for the inference process.

First, we binarize the data with the fuzzy-2-means algorithm in order to obtain a binarization and a membership coefficient p_{ij} for each Boolean value indicating the reliability of the binarization. Based on the fuzzy model, multiple binarized time series are sampled via randomized rounding, using the coefficients as probabilities of membership. For each of these binarizations and each of the genes in the network, we infer possible dependencies by scoring all combinations of input genes. The scoring of an input gene combination is based on the error of the best Boolean function and on the number of involved genes. An accumulated score for each input gene combination over all binarizations is calculated, and the dependencies are modeled by the best-ranked combinations.

By incorporating uncertainty into the reverse-engineering process, we improve the accuracy in terms of state transitions and network wiring. For validation, our new approach was applied on artificial data and yeast expression time series data.

References

KAUFFMAN, S. A. (1969): Metabolic Stability and Epigensis in Randomly Constructed Genetic Nets. *Journal of Theoretical Biology, 22(3):437–467.*

HICKMAN, G. J. and HODGMAN T.C. (2009). Inference of gene regulatory networks using Boolean-network inference methods. *Journal of Bioinformatics and Computational Biology, 7(6):1013–29.*

BORNHOLDT, S. (2005): Systems Biology. Less is More in Modeling Large Genetic Networks. *Science, 310(5747): 449–451.*

Inferring Boolean network structure via correlation

M. Maucher, B. Kracher, M. Kühl, H.A. Kestler

Institute of Neural Information Processing,
Institute for Biochemistry and Molecular Biology,
University of Ulm
{markus.maucher,hans.kestler}@uni-ulm.de

The dynamic behavior of genetic regulatory networks can be described and analyzed using Boolean network models. The reconstruction of such a Boolean network from time series data requires the identification of dependencies within the network. To facilitate the dependency structure of such a network, one can take advantage of the fact that in a gene regulatory network a specific transcription factor often will consistently either activate or inhibit a specific target gene. In this case, the observed regulatory behavior can be modeled by the use of monotone functions.

We show that Pearson correlation can identify the dependencies in a Boolean network from time series data if that network consists of monotone Boolean functions. This approach enables fast inference of Boolean networks based on an intuitive correlation measure. In experiments, we could reconstruct large fractions of both a published E. coli transcriptional regulatory and metabolic network from simulated data and a yeast cell cycle network from microarray data.

Meta-analysis Methods for Gene Expression Profiles

Berthold Lausen

Department of Mathematical Sciences
University of Essex
blausen@essex.ac.uk

A fast increasing amount of public available gene expression data sets allows the use of meta analysis techniques to validate and to identify molecular signatures. I review several recent approaches. An important condition for preprocessing methods is that the data analysis work flow should not be influenced by properties of other data sets included in the meta analysis. For example a preprocessing method of one Affymetrix cel file should be invariant under different sets of cel files included in the meta analysis. I illustrate the talk with gene expression data sets of colorectal cancer.

References

BUFFA, F.M., HARRIS, A.L., WEST, C.M., MILLER, C.J. (2010): Large Meta- analysis of Multiple Cancers Reveals a Common, Compact and Highly Prognostic Hypoxia Metagene. British Journal of Cancer, 102, 428–35.

CRONER, R., FÖRTSCH, T., BRÜCKL, W., RÖDEL, F., et al. (2008): Molecular Signature for Lymphatic Metastasis in Colorectal Carcinomas. Annals of Surgery 247, 803–810.

GORLOV, I.P., SIRCAR, K., ZHAO, H., et al. (2010): Prioritizing genes associated with prostate cancer development. BMC Cancer 10:599.

MCCALL, M.N., BOLSTAD, B.M., IRIZARRY, R.A. (2010): Frozen robust multi-array analysis (fRMA). Biostatistics 11, 2, 242–253.

MPINDI, J.P., SARA, H., HAAPA-PAANANEN, S. et al. (2011): GTI: A Novel Algorithm for Identifying Outlier Gene Expression Profiles from Integrated Microarray Datasets. PLOSone 6, 2, e17259.

SHI, F., ABRAHAM, G., LECKIE, C., HAVIV, I., KOWALCZYK, A. (2011): Meta-analysis of gene expression microarrays with missing replicates. BMC Bioinformatics 12,84

Connecting miRNA and mRNA expression profiles for medical classification problems

Klaus Jung, Tim Beißbarth and Mathias Fuchs

Department of Medical Statistics,
University Medical Center Göttingen,
Department of Bioinformatics,
University Medical Center Göttingen
kjung1@uni-goettingen.de

In biomedical research, it is by now very common that different types of high-dimensional molecular data are studied in parallel. In the past, many studies concentrated for example only on gene expression, protein expression or genetic data, due to the high cost of each technique or because techniques were just established in many research groups. Meanwhile, techniques have become cheaper and more common, so that they can be applied in parallel to study the same biological sample, e.g. a tumour biopsy or a cell line. A typical question for studying molecular data is to find differences in the samples from different biological groups, e.g. individuals with different phenotypes or different response to a therapy. In particular, many studies aim at finding molecular signatures that can be used for diagnosis or prediction. In this context, we currently study methods for connecting the information from mRNA and miRNA expression data for classification problems in medicine. Our particular questions are as follows. Should we first merge mRNA and miRNA data and search then for a common signature? Or should we first search individual signatures and combine then the correspond-ing classification rules (Kittler et al., 1998)? Is a merged classifier better than an individual one? Is there a benefit of having both data sources available? We evaluate the different approaches within a simulation study and on several publicly avail-able data (e.g. Peng et al., 2009). More precisely, we compare the prediction accuracies obtained with each approach. In addition, we discuss several technical difficulties of each approach, for example a common normalization of mRNA and miRNA data.

References

Kittler, J., Hatef, M., Duin, R.P.W. and Matas, J. (1998) On combining classifiers. IEEE Transactions on pattern analysis and machine intelligence, 20, 226–239.

Peng, X., Li, Y.,Walters, K.A., Rosenzweig, E.R., Lederer, S.L., Aicher, L.D., Proll, S. and Katze, M.G. (2009) Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers. BMC Genomics, 10: 373.

Integration of copy number variation and gene expression data in Bayesian models for prediction

Manuela Zucknick, Stefan Pfister and Axel Benner

Division of Biostatistics (C060),
Division Molecular Genetics (B060),
German Cancer Research Center, Heidelberg,
Department of Pediatric Oncology, Hematology and Immunology,
University Hospital Heidelberg
m.zucknick@dkfz-heidelberg.de

Bayesian variable selection models are an alternative to well-known regularisation methods like lasso regression and boosting for prognostic modelling based on high-dimensional input spaces. A typical application is prediction of clinical endpoints such as therapy response using microarray gene expression data.

High-throughput microarray technologies are available for many other types of genomic data in addition to gene expression, and in recent years clinical researchers have begun to systematically collect genome-wide data from various sources on the DNA- and RNA-level as well as epigenetic data. If data from several sources are available for the same set of biological samples, the data can be analysed together in an integrative manner, with the aim of providing a more comprehensive picture of the disease biology as well as improving the performance of clinical prediction models.

For example, the integration of copy number variation data into classical gene expression based prognostic models promises to improve both prognostic value and interpretability of the model, because genomic deletions and amplifications are known to affect expression levels of genes located in the corresponding genomic regions. In fact, the deletion of chromosomal regions harbouring important tumour suppressor genes is a well-known cause of certain cancers.

In contrast to methods like lasso and boosting, Bayesian variable selection models are very flexible in their setup and are naturally well-suited to extensions allowing for the integration of additional data sources.

We will propose a hierarchical Bayesian variable selection model, which combines whole-genome information on copy number variation and gene expression in a manner that is intuitive from a biological point of view. The model setup will be demonstrated, as well as aspects of the MCMC sampling algorithm and posterior inference. The model will be further illustrated in an application to pediatric brain tumour data.

On the utility of partially labeled data for classification in high dimensional settings

Ludwig Lausser, Florian Schmid and Hans A. Kestler

Institute of Neural Information Processing,
University of Ulm,
Department of Internal Medicine I,
University Hospital Ulm
{ludwig.lausser,hans.kestler}@uni-ulm.de

Initial results gained by high throughput technologies such as microarrays or deep sequencing are common starting points for investigations within molecular medicine or biology. They allow the tracking of several thousand signals within a single experiment. The data produced by such technologies is of usually high dimensionality but also of a low cardinality. Many inferences regarding these data can be formulated as clustering or classification tasks. In a clustering scenario the task is to find groups in a sample of data points. It is an example for unsupervised learning. The sample does not contain explicit information on the involved classes or groups; especially it does not contain class labels. In a classification scenario the involved categories are known a priori. The task is to predict the correct category of a unseen data point. Classification is an example of a supervised learning task. Here the training set contains examples labeled according the these categories.

Supervised and unsupervised learning are widely used paradigms within the analysis of microarray data. Other methodologies that bridge these paradigms are often neglected. These methods are based on partially labeled datasets and incorporate information gained from labeled and unlabeled data. Examples for such concepts are transductive learning and semi-supervised learning. In this work we investigate the benefit of these learning schemes within the classification of microarray datasets.

We compare several supervised algorithms to their transductive (or semi-supervised) counterparts in real and artificial settings. Aim of the study is the investigation of the influence of the high dimensionality on the generalization ability and the robustness of the algorithms.

Multi-objective parameter selection for classifiers

C. Müssel, L. Lausser, M. Maucher and H. A. Kestler

Institute of Neural Information Processing,
University of Ulm
{christoph.muessel,ludwig.lausser,markus.maucher,hans.kestler}@uni-ulm.de

The choice of appropriate values for parameters is an essential step in classifier training and can have a major influence on the classification performance. Often, such parameters are set according to rules of thumb. Parameter tuning is an automated way of adapting parameters. Most frequently, parameters are tuned according to single criterion, such as the cross-validation error, which can be a good estimate of the generalization performance. However, it is sometimes desirable to obtain parameter values that optimize several concurrent criteria at the same time. For example, sensitivity and specificity are important – but usually conflicting – characteristics of a classifier. Dominance-based selection procedures allow for a simultaneous optimization of multiple objectives. They leave the ultimate decision on the desired trade-off of objectives to the human expert.

We devised the R package *TunePareto* for multi-objective selection of parameters for classifiers. The software chooses candidate parameter configurations according to sophisticated sampling strategies and search heuristics, such as quasi-random sequences and evolutionary algorithms. It then determines the optimal configurations using Pareto dominance. The package provides flexible interfaces for classifiers and objective functions. The decision making process is supported by various visualizations as well as the formal definition of desired and undesired objective values.

We present a tutorial on the functionality and usage of the TunePareto package.

The Daim package – Diagnostic accuracy of classification models

Sergej Potapov, Berthold Lausen and Werner Adler

Department of Biometry and Epidemiology
University of Erlangen-Nuremberg
Department of Mathematical Sciences
University of Essex
{sergej.potapov, werner.adler}@imbe.med.uni-erlangen.de

The `Daim` package contains several functions for evaluating the accuracy of classification models by ROC analysis (Fawcett, 2006). It provides the following performance measures: "cv", "bcv", "0.632" and "0.632+" estimation of the misclassification rate, sensitivity, specificity and AUC (Efron & Tibshirani, 1997; Adler & Lausen, 2009). The package provides a flexible interface to classifier functions and facilitates intuitive evaluation of predictive models. If an application is computationally intensive, parallel execution can be used in a simple manner to reduce the time taken.

References

EFRON, B. and TIBSHIRANI, R. (1997): Improvements on Cross-Validation: The .632+ Bootstrap Method. *JASA, 92(438), 548–560.*

FAWCETT, T. (2006): An introduction to ROC analysis. *Pattern Recognition Letters, 27(8).*

ADLER, W. and LAUSEN, B. (2009): Bootstrap estimated true and false positive rates and ROC curve. *Comput. Stat. Data Anal., 53(3), 718–729.*

Correcting the optimally selected resampling-based error rate: A smooth analytical alternative to nested cross-validation

Christoph Bernau, Thomas Augustin and Anne-Laure Boulesteix

Department of Medical Informatics, Biometry and Epidemiology (IBE),
University of Munich,
Department of Statistics,
University of Munich
bernau@ibe.med.uni-muenchen.de

Many statistical problems in bioinformatics are high-dimensional binary classification tasks, e.g. the classification of microarray samples into normal and cancer tissues. In this context, statistical learning methods usually incorporate a tuning parameter adjusting their complexity to the specific examined data set. By simply reporting the performance of the best tuning parameter value, overly optimistic prediction errors have been published in the past (Varma and Simon, 2006). A straightforward approach to avoid this tuning bias is nested cross-validation (CV).

In this talk we are addressing two objectives. Firstly, we develop a new method correcting for this tuning bias by embedding the tuning problem into a decision theoretic framework. The method is based on the decomposition of the unconditional error rate involving the tuning procedure. Our corrected error estimator can be reformulated as a weighted mean of resampling errors obtained using the different tuning parameter values. In this sense, it can be interpreted as a smooth version of nested CV. The smooth weighting additionally guarantees intuitive bounds for the corrected error. Secondly, we suggest to also use bias correction methods to address the bias resulting from the optimal choice of the learning method. The latter bias is particularly relevant to prediction problems based on high-dimensional ”omic” data. In the absence of standards, it is indeed common practice to apply several methods successively. This can lead to an optimistic bias similar to the tuning bias if one reports the performance of the optimal method only.

We demonstrate the performance of our new method to address both types of bias based on four microarray cancer data sets and compare it to existing methods. Our main result is that our approach yields intuitively bounded estimates similar to nested CV and at a dramatically lower computational price.

References

S. Varma and R. Simon. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7:91.

Bias-Variance Analysis of Local Classification Methods

Julia Schiffner and Claus Weihs

Department of Statistics,
TU Dortmund
{schiffner, weihs}@statistik.tu-dortmund.de

Nowadays a plethora of classification methods is available and new ones or modifications of established methods are regularly published. They can be grouped using different properties as e.g. parametric or nonparametric, distance-based or not, predictive or generative etc. Another distinction can be made between global and local methods. In recent years the amount of publications on local classification methods is increasing. Localized versions of nearly all standard classification techniques like linear discriminant analysis [1] and Fisher discriminant analysis [6], logistic regression [3, 7], support vector machines [5] or boosting [8] are available.

The term local is only vaguely defined and used in a rather intuitive way by most authors, referring to the position in some space, to a part of a whole or to something that is not general or widespread. Often it relates to the the neighborhood of the point where a prediction is required, with the k nearest neighbors method [2] as probably best-known example. But also other concepts of locality can be found in the relevant literature. For example Hand and Vinciotti [3] use the term local to refer to points close to the decision boundary. Most localization techniques can be applied in a generic manner to many different classification methods which results in a rather broad field of methods. We will give an overview of existing approaches and their properties.

A question of interest is how localization affects the performance of classification methods. The bias-variance decomposition of prediction error is conducive to gaining deeper insight into the behavior of learning algorithms. It was originally introduced for quadratic loss functions, but since in classification the misclassification rate is usually of interest, generalizations to zero-one loss have been developed in the last 15 years, e.g. [4]. This was particularly motivated by research on multi-classifier systems where variance-reduction was found as one explanation for the often good performance of multi-classifier systems.

In order to gain deeper insight into how local methods work we analyze local classification methods in terms of bias and variance of the error rate Our intuition that is supported by our recent experiments clearly is that local methods in general reduce the bias in comparison with global counterparts. We will show some toy examples for illustration of the decomposition and present some results for selected classification methods and localization types on simulated and real-world data sets.

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] E. Fix and J. L. Hodges. Discriminatory analysis – nonparametric discrimination: Consistency properties. Report 4, U.S. Airforce School of Aviation Medicine, Randolph Field, Texas, 1951.

[3] 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, May 2003.

[4] G. M. James. Variance and bias for general loss functions. *Machine Learning*, 51(2): 115–135, May 2003.

[5] N. Segata and E. Blanzieri Fast and scalable local kernel machines. *Journal of Machine Learning Research*, 11:1883–1926, June 2010.

[6] M. Sugiyama. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *Journal of Machine Learning Research*, 8:1027–1061, May 2007.

[7] G. Tutz and H. Binder. Localized classification. *Statistics and Computing*, 15:155–166, 2005.

[8] C.-X. Zhang and J.-S. Zhang. A local boosting algorithm for solving classification problems. *Computational Statistics & Data Analysis*, 52:1928–1941, 2008.

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

95-12 *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 $\mathrm{ZPP}^{\mathrm{NP}}$ 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 *Hans A. Kestler, Berthold Lausen, Harald Binder, Hans-Peter Klenk, Friedrich Leisch, Matthias 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 Beth, 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 *Hans A. Kestler, Harald Binder, Berthold Lausen, Hans-Peter Klenk, Matthias Schmid, Friedrich 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

Ulmer Informatik-Berichte
ISSN 0939-5091

Herausgeber:
Universität Ulm
Fakultät für Ingenieurwissenschaften und Informatik
89069 Ulm