



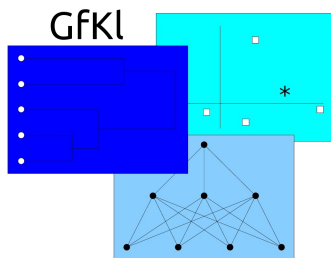
# Classification 2016 - Abstracts of the 5th German-Japanese Symposium

A Fürstberger, JM Kraus, HA Kestler (eds)

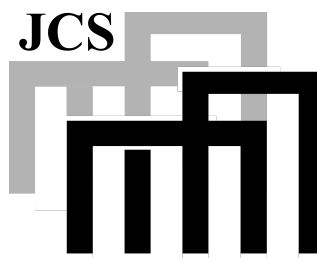
## Ulmer Informatik-Berichte

Nr. 2016-05  
September 2016

Supported by



German Classification Society



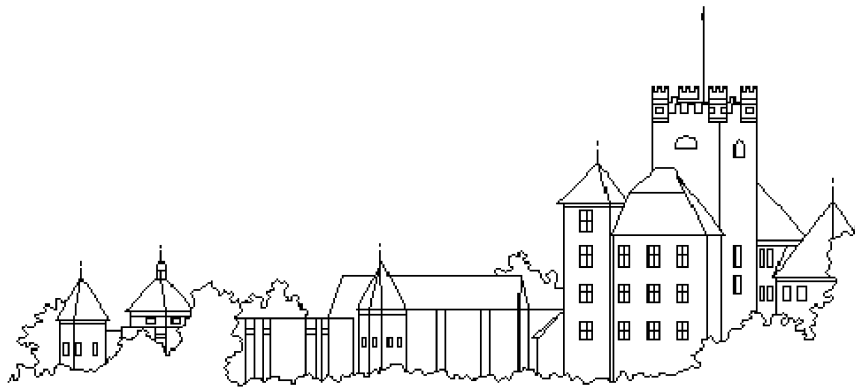
Japanese Classification Society



German Research Foundation



# 5th German-Japanese Symposium on Classification 2016



**Joint workshop of the**

**Japanese Classification Society (JCS)**

**and the**

**German Classification Society (GfKI)**

**11.09. - 13.09.2016, Schloss Reisenburg (Günzburg)**

# Workshop Program

## Sunday, September 11, 2016

17:00-17:50	Informal Meeting
18:00-21:00	Welcome Reception and Get Together

## Monday, September 12, 2016

08:50-09:00	Hans Kestler (Ulm)	Opening
<b>09:00-10:15</b>	<b>Chair: B. Lausen</b>	
09:00-09:25	Tadashi Imaizumi (Tokyo)	Model Based Clustering for a Sparse Similarity Matrix
09:25-09:50	Akinori Okada, Hiroyuki Tsurumi (Tokyo, Yokohama)	Estimating Market Share of New Brand by Analyzing Brand Switching Matrix Using Asymmetric Multidimensional Scaling
09:50-10:15	Ludwig Lausser (Ulm)	Exact relabeling tests for classification
<b>10:15-10:45</b>	<b>Break</b>	
<b>10:45-12:00</b>	<b>Chair: T. Imaizumi</b>	
10:45-11:10	Berthold Lausen (Essex)	Ensemble methods for clustering and classification
11:10-11:35	Sadaaki Miyamoto (Tsukuba)	Ward method applied to non-positive definite matrices with a model of unstrict users
11:35-12:00	Atsuhiko Nakayama (Tokyo)	Analysis of Trending Topics in Consumer Web Communication Data
<b>12:00-13:30</b>	<b>Lunch</b>	
<b>13:30-14:20</b>	<b>Chair: C. Weihs</b>	
13:30-13:55	Fumitake Sakaori (Tokyo)	Fully Bayesian Soft Impute for Matrix Completion
13:55-14:20	Johann Kraus (Ulm)	Tuning hierarchical clustering with domain knowledge
<b>14:20-14:50</b>	<b>Break</b>	
<b>14:50-15:40</b>	<b>Chair: A. Wilhelm</b>	
14:50-15:15	Claus Weihs (Dortmund)	Classification on Large-Scale Data: Systematic Testing in the many-features case
15:15-15:40	Yoshiro Yamamoto, Sanetoshi Yamada (Hiratsuka)	Visualization of cross tabulation using association plot
15:40-21:00	Social Program (Ulm)	Guided Tour and Dinner

## Tuesday, September 13, 2016

<b>09:00-10:15</b>		<b>Chair: J. Kraus</b>
09:00-09:25	Adalbert Wilhelm (Bremen)	The imbalanced class problem revisited
09:25-09:50	Masahiro Mizuta (Sapporo)	Analysis of environmental data with SDA and FDA
09:50-10:15	Andreas Geyer-Schulz (Karlsruhe)	Do Industrial End-Consumer Product-Configurators Use Rational Pricing? A Case Study
<b>10:15-10:45</b>		<b>Break</b>
<b>10:45-13:55</b>		<b>Chair: A. Okada</b>
10:45-11:10	Hans-Hermann Bock (Aachen)	Interaction-based co-clustering - but which interactions?
11:10-11:35	Kei Kurakawa, Yasumasa Baba (Tokyo)	Tensor based relational learning for the author name disambiguation
11:35-12:00	Jun Tsuchida (Kyoto)	Majorization algorithm for dominance point model
<b>12:00-13:30</b>		<b>Lunch</b>
13:30-13:55	Hans-Joachim Mucha (Berlin)	Big Data Clustering: Is Subsampling Better than Fast Pre-clustering?
13:55-14:20	Hans Kestler (Ulm)	Closing Remarks
15:40		Departure



## Synopsis

This volume comprises abstracts of talks presented at the 5th German-Japanese Symposium on Classification hosted by the Institute of Medical Systems Biology, Ulm University and the German Classification Society (GfKI) at the Reisenburg Castle in Günzburg 2016. President Prof. Dr. B. Lausen (University of Essex) and 1st Vice President Prof. Dr. H. A. Kestler (Ulm University) of the GfKI as well as President Tadashi Imaizumi (Tama University) and other members of the Japanese Classification Society (JCS) attended the symposium. The conference was supported by the GfKI as well as the German Research Foundation (DFG).

The scientific program of the symposium includes presentations covering a broad range of topics. A special emphasis is laid on research and development of tools, techniques, and strategies that address challenges of data science using computational, mathematical, statistical and data analytical methods for classification.

With half of the participants coming from outside of Germany, the symposium corroborates the spirit of international networks and cooperations.

Selected papers will be published in a special issue of the "Archives of Data Science".

### Past Symposiums:

- Tokyo 2005
- Berlin 2006
- Karlsruhe 2010
- Kyoto 2012

September 2016

Hans Kestler  
Director, Institute of Medical Systems Biology  
Ulm University





# Contents

Model Based Clustering for a Sparse Similarity Matrix . . . . .	1
Estimating Market Share of New Brand by Analyzing Brand Switching Matrix Using Asymmetric Multidimensional Scaling . . . . .	2
Exact relabeling tests for classification . . . . .	3
Ensemble methods for clustering and classification . . . . .	4
Ward method applied to non-positive definite matrices with a model of unstrict users . . . . .	5
Analysis of Trending Topics in Consumer Web Communication Data . . . . .	6
Fully Bayesian Soft Impute for Matrix Completion . . . . .	7
Tuning hierarchical clustering with domain knowledge . . . . .	8
Classification on Large-Scale Data: Systematic Testing in the many-features case	9
Visualization of cross tabulation using association plot . . . . .	10
The imbalanced class problem revisited . . . . .	11
Analysis of environmental data with SDA and FDA . . . . .	12
Do Industrial End-Consumer Product-Configurators Use Rational Pricing? A Case Study. . . . .	13
Interaction-based co-clustering - but which interactions? . . . . .	14
Tensor based relational learning for the author name disambiguation . . . . .	15
Majorization algorithm for dominance point model . . . . .	16
Big Data Clustering: Is Subsampling Better than Fast Pre-clustering? . . . . .	17
List of technical reports published by the University of Ulm . . . . .	19



# Model Based Clustering for a Sparse Similarity Matrix

Tadashi Imaizumi<sup>1</sup>

When a given (dis)similarity matrix  $\mathbf{S}$  of  $n \times n$ , a hierarchical clustering methods is useful for us to understand the relationship of objects in data. However, it is difficult to choose the "proper" clustering results, without prior information of the number of clusters and distance metric between clusters etc. Bock(1996) discussed on probabilistic clustering and Fraley and Raftery(2003) discussed on the model-based cluster analysis. These probabilistic model-based clustering approach will be more appropriate to find the hidden clusters in data matrix of objects  $\times$  attributes.

It is also applicable for a similarity matrix after obtaining a dimensional representation of objects(configuration). However this approach is not appropriate one as it consists of two different phases. And how to derive this configuration becomes the key point. We propose a probabilistic model-based clustering approach for a similarity matrix. The normal distribution is assumed to derive a configuration for a fixed number of clusters  $G$ . An object configuration will be obtained by a method of multi-dimensional scaling with a distributional assumption,

$$X_i \sim N(\mu_g, \Sigma_g), i = 1, 2, \dots, n. \quad (1)$$

The covariance matrix of each cluster  $\Sigma_g$  is assumed to be decomposed as

$$\Sigma_g = \lambda_g \mathbf{D}_k \mathbf{A}_k \mathbf{D}_k^t, g = 1, 2, \dots, G \quad (2)$$

where  $\mathbf{D}_k$  is the orthogonal matrix of eigenvectors. And the similarity  $s_{ij}$  between two objects  $i$  and  $j$  of same cluster are related as the distance between two corresponding points within cluster. And those between two objects of different clusters are related as the distance between two cluster means of each cluster. The configuration is updated with using the estimated covariance matrices. When a given sparse similarity matrix is one of the block-diagonal matrix, then the above procedure works well as the covariance matrix of each cluster will be identified, but, we need to pay attention when similarity matrix is more complex one. In that case, the structure of the covariance matrix will be restricted.

## References

- 1 Bock, H.-H. Probabilistic models in cluster analysis, Computational Statistics & Data Analysis, 23,5-28, 1996.
- 2 Fraley, C. and Raftery, A. E. Enhanced software for model-based clustering, density estimation, and discriminant analysis: MCLUST. Journal of Classification, 20, 263– 286, 2003.

---

<sup>1</sup> Tama University 4-1-1 Hijirigaoka, Tama, Tokyo, JAPAN, 206-0022

# Estimating Market Share of New Brand by Analyzing Brand Switching Matrix Using Asymmetric Multidimensional Scaling

Akinori Okada<sup>1</sup> and Hiroyuki Tsurumi<sup>2</sup>

When a new brand is introduced into market, the new brand may dominate existing brands and takes a lot of customers away from the existing brands or the new brand may not dominate the other brands and takes only small number of the customers away from the existing brands. A brand switching matrix among existing brands is analyzed by asymmetric multidimensional scaling of Okada & Tsurumi (2014) to derive configuration of existing brands. The analysis represents one configuration along each dimension, where asymmetric relationships among existing brands are represented by the outward tendency and the inward tendency. The outward tendency of a brand represents the tendency of switching from the brand to the other brands, and the inward tendency of a brand represent the tendency of switching from the other brands to the brand. The characteristics of a new brand are compared with those of the existing brands along dimensions, and the location of the new brand in the configuration is determined. The location of the new brand is used to estimate the frequency of brand switching to/from th new brand with existing brands and the market share of the new brand as well. The procedure is applied to frequent shoppers program data analyzed in Okada & Tsurumi (2014).

## References

Okada A. and Tsurumi, H. (2014): Evaluating the Effect of new Brand by Asymmetric Multidimensional Scaling. In: D. Vicari, A. Okada, G. Ragozini and C. Weihs (Eds.): *Analysis and Modeling of Complex Data in Behavioral and Social Sciences*. Springer, Heidelberg, 147–156.

<sup>1</sup> Tama University Research Institute, 4-1-1 Hijirigaoka Tama-shi Tokyo Japan 206-0022

<sup>2</sup> College of Business Administration, Yokohama National University, 79-4 Tokiwadai, Hodogaya-ku, Yokohama-shi Japan 240-8501

okada@rikkyo.ac.jp, tsurumi@ynu.ac.jp

# Exact relabeling tests for classification

*Ludwig Lausser<sup>1</sup>, Alexander Groß<sup>1</sup>, Hans A. Kestler<sup>1</sup>*

An object can typically be assigned to many different concepts. It is therefore not a sample for one particular class or category. This eclectic character of an object can be a hindrance for the process of supervised learning. Being not aware of the characteristic patterns of a categorization, a classifier might be diverted by the patterns of a different one.

For a chosen classification task, the suitability of a set of samples is therefore not only characterized by the classification accuracy achieved by the corresponding classifier. It is also characterized by the classification accuracy achieved for any other (possibly unknown) classification of the dataset. A relative measure of these two quality scores is needed.

In this work, we propose a relabeling strategy that allows the exhaustive evaluation of all binary labelings in the leave one out cross-validation experiments. The strategy utilizes the characteristics of  $k$ -nearest neighbor classifiers and the corresponding graph structures for fast and efficient calculation. It can be applied for the construction of an exact permutation test which relates the classification performance achieved for a chosen categorization to any other categorization. We utilize this strategy in artificial scenarios and the domain of functional genomics.

---

<sup>1</sup> Institute of Medical Systems Biology, Ulm University, 89069 Ulm, Germany

`ludwig.lausser@uni-ulm.de, alexander.gross@uni-ulm.de, hans.kestler@uni-ulm.de`

# Ensemble methods for clustering and classification

Berthold Lausen<sup>1</sup>, Kaloyan Stoyanov<sup>2</sup>, Luca Citi<sup>3</sup>, Rolando Medellin<sup>1,2</sup>,  
Henrik Nordmark<sup>2</sup>, Aris Perperoglou<sup>1</sup>

We review methods to use ensembles of selected classifiers to achieve classification rules with increased accuracy (Gul et al. 2016, Khan et al. 2016). Feature selection methods are often used as preprocessing method. For example after preprocessing microarray data with 500 000 probes and 22125 features (probesets) which represent genes, we use a proposal to improve feature selection of microarray data based on a proportional overlapping score (Mahmoud et al. 2014).

We investigate ensemble methods for cluster analysis. Using ensemble concepts Stoyanov (2015) developed an R package to use hierarchical clustering as preprocessing for k-means clustering. In addition we discuss proposals to use nonparametric and parametric bootstrap resampling and distance based variance estimation (Felsenstein 1985; Lausen & Degens, 1987; Degens, Lausen & Vach 1990) to derive a statistical evaluation of clusters.

## References

- Degens, P.O., Lausen, B., Vach, W. (1990), Reconstruction of phylogenies by distance data: Mathematical framework and statistical analysis, Lecture notes in biomathematics 84, 9-42. Felsenstein, J. (1985), Confidence limits on phylogenies: an approach using the bootstrap, *Evolution*, 783-791
- Gul, A., Perperoglou, A., Khan, Z., Mahmoud, O., Miftahuddin, M., Adler, W., Lausen, B. (2016), Ensemble of a subset of kNN classifiers, *Advances in Data Analysis and Classification*, online first DOI:10.1007/s11634-015-0227-5
- Khan, Z., Gul, A., Mahmoud, O., Miftahuddin, M., Perperoglou, A., Adler, W., Lausen, B. (2016), An ensemble of optimal trees for class membership probability estimation. In: Wilhelm, A., Kestler, H. A. (eds.), *Analysis of Large and Complex Data, European Conference on Data Analysis, Bremen, July, 2014, Series: Studies in Classification, Data Analysis, and Knowledge Organization*, Springer-Verlag Berlin.
- Lausen, B., Degens, P.O. (1988), Evaluation of the reconstruction of phylogenies with DNA-DNA hybridization data, in: Bock, H.H. (ed.), *Classification and related methods of data analysis, Proceedings First conference of the international federation of classification societies (IFCS), North Holland, Amsterdam, 367-374.*
- Mahmoud, O., Harrison, A.P., Perperoglou, A., Gul, A., Khan, Z., Metodiev, M., Lausen, B. (2014), A feature selection method for classification within functional genomics experiments based on the proportional overlapping score, *BMC Bioinformatics* 15 (1).
- Stoyanov, K. (2015), *Hierarchical k-means clustering and its application in customer segmentation.* Master dissertation, Department of Mathematical Sciences, University of Essex, UK.

<sup>1</sup> Department Mathematical Sciences, University of Essex, UK

<sup>2</sup> Profusion Ltd., London, UK

<sup>3</sup> School of Computer Science and Electronic Engineering, University of Essex, UK

blausen@essex.ac.uk

# Ward method applied to non-positive definite matrices with a model of unstrict users

Sadaaki Miyamoto<sup>1</sup>

It is well-known that the Ward method can only be applied to a set of objects as points in an Euclidean space and not to a general matrix of similarity/dissimilarity. More recently, it has also been proved that the Ward method can be applied to the class of semi-positive definite matrix of similarity by applying the theory of kernels. At the same time, the Ward method cannot be applied to the case of a non-positive definite matrix of similarity which has a negative eigenvalue. In spite of this fact, the aim of this study is to show the possibility to apply the Ward method even to the case of non-positive definite matrices of similarity  $S$  with a mild condition of  $S(i, i) = 1$  for  $i = 1, \dots, n$  and  $S(i, j) \leq S(i, i)$  for all  $j$ .

It is of course impossible that the Ward method is rigorously justified in the latter case. Instead, we introduce a model of *unstrict* users. Probably most researchers will agree that users of agglomerative hierarchical clustering in applications are rather rough or unstrict: they not care much about details of used similarities or very details of the output of dendrograms. Hence we assume a model of an unstrict user as follows: he does not care about the difference of two dissimilarities of  $D(i, j)$  and  $D'(i, j) = D(i, j) + \text{constant}$  when  $i$  and  $j$  are different. He does not care about the difference of two dendrograms when they produce the same sequence of the same clusters in the same order, but the merged levels of clusters may be different. We now have the following result:

Let  $S$  be a similarity matrix produced from  $D$  by a simple transformation of  $S = E - D$ , assuming  $0 \leq D(i, j) \leq 1$ .  $S' = S + aE$  will be positive-definite for sufficiently large constant  $a$ . Then the Ward method can be justified to  $S'$ . The Ward method applied to  $S$  cannot be justified in the rigorous way, but the algorithm of the Ward method to  $S$  still works and the dendrogram applied to  $S$  is 'isomorphic' to another dendrogram applied to  $S'$  in the sense that the two dendrograms will produce the same sequence of the same clusters in the same order, but the merged levels of clusters are different. It is also easy to see that  $D'(i, j) = D(i, j) + \text{constant}$ .

To summarize, the Ward method can be applied to a non positive-definite  $S$  under the assumption of the above model of unstrict users. This result make the applicability of the Ward method much broader, e.g., to the case of network clustering without the use of a particular type of kernels.

**Acknowledgment:** This study has partly been supported by the Grant-in-Aid for Scientific Research, JSPS, Japan, no.26330270.

---

<sup>1</sup> University of Tsukuba, Japan

miyamoto@risk.tsukuba.ac.jp

# Analysis of Trending Topics in Consumer Web Communication Data

Atsuo Nakayama<sup>1</sup>

In this study, we detected trending topics by classifying words into clusters based on the co-occurrence of words in web communications among consumers. We collected web communication data about certain specific theme based on their specific expressions of sentiment or interest. Because of the desire to identify market trends, the analysis of consumer web communication data has received much attention in Japan. To detect topics more easily, we tokenized each web communication data that was written in sentences or sets of words. However, one of the most difficult natural language-processing problems in Japanese is tokenization. This is referred to as the “wakachigaki” problem. In most Western languages, words are delimited by spaces and punctuation. In Japanese, words are not separated by spaces. We used morphological analyses such as tokenization, stemming, and part-of-speech tagging to separate the words. In our study, we used the Japanese morphological analyzer ChaSen to separate words in passages and to distinguish all nouns, verbs, and adjectives. ChaSen (<http://chasen.naist.jp/>) is a fast, customizable Japanese morphological analyzer that takes the form of a hierarchical structure. It is designed for generic use, and can be applied to a variety of language-processing tasks. The entry  $\times$  word matrix obtained from the web communication data was sparse and of high dimensionality, so it was necessary to perform a dimensionality reduction analysis. We classified the words extracted from web communication data using non-negative matrix factorization as a dimensionality reduction model (Lee & Seung, 2000).

**Acknowledgment:** This work was supported by a Grant-in-Aid for Scientific Research (C) (No. 16K00052) from the Japan Society for the Promotion of Science.

## References

- Lee, D.D., and Seung, H.S. (2000): Algorithms for Non-Negative Matrix Factorization. In: K. T. Leen, T. G. Dietterich, and V. Tresp, (Eds.) *Advances in Neural Information Processing Systems*. MIT Press, Cambridge, MA, vol. 13, pp. 556-562.

---

<sup>1</sup> Graduate School of Social Sciences, Tokyo Metropolitan University, 1-1 Minami-Ohsawa, Hachioji-shi, Tokyo 192-0397, Japan

atsuho@tmu.ac.jp



# Fully Bayesian Soft Impute for Matrix Completion

Fumitake Sakaori<sup>1</sup>, Hiroki Kurosawa<sup>1</sup>

Matrix completion methods for an incomplete low-rank matrix have been applied in many practical situation such as collaborative filtering, image processing and analysis of gene expression array. Accordingly, various matrix completion methods have been developed and investigated theoretically. The objective of these methods are to complete a incomplete observed matrix under the assumption that the observed matrix are represented as a sum of a low-rank matrix and a noise matrix.

Mazumder *et al.* (2010) proposed Soft-Impute, a sparse modeling for an incomplete matrix, where the sparsity of singular values of the low rank matrix is achieved by regularizing the nuclear norm of the low rank matrix. Todeschini *et al.* (2013) proposed an adaptive extension of the Soft-Impute, Hierarchical Adaptive Soft-Impute and gave their bayesian representation. The effectiveness of these methods heavily rely on the choice of tuning parameters.

In this study, we propose a fully bayesian modeling of Soft-Impute which can estimate the tuning parameters simultaneously.

## References

- 1 Mazumder, R., Hastie, T. & Tibshirani, R. (2010). Spectral Regularization Algorithms for Learning Large Incomplete Matrices, *Journal of Machine Learning Research*, **11**, 2287-2322.
- 2 Todeschini, A., Caron, F. & Chavent, M. (2013). Probabilistic low-rank matrix completion with adaptive spectral regularization algorithms, *Advances in Neural Information Processing Systems*, **26**, 845-853.

---

<sup>1</sup> Department of Mathematics, Chuo University. 1-13-27, Kasuga, Bunkyo-ku, Tokyo 112-8551 JAPAN

sakaori@math.chuo-u.ac.jp

# Tuning hierarchical clustering with domain knowledge

Johann M. Kraus<sup>1</sup> and Hans A. Kestler<sup>1</sup>

Cluster analysis presents a variety of tools from the explorative data analysis that are used for predicting an unknown structure hidden in data. Unsupervised cluster methods do not make use of domain knowledge about any possible grouping of the data. However, many partitional cluster algorithms were adapted to additionally make use of available background information either by constraining the search process or by modifying the underlying metric. Limitations in the reproducibility of clustering results in resampling experiments triggered the inclusion of domain knowledge into the hierarchical clustering process, too. Based on our previous work (Kestler et al. 2006, Kraus et al. 2007), we present a general framework for including domain knowledge into a hierarchical clustering process. Our new semi-supervised cluster strategy aims at assessing the reliability of hierarchical clustering. Reliable clusters now can be identified by searching for the most stable partitions under different clustering conditions in resampling experiments.

## References

- Kestler, H.A., Kraus, J.M., Palm, G., Schwenker, F. (2006), On the effects of constraints in semi-supervised hierarchical clustering. In: F. Schwenker, S. Marinai (Eds.): *Artificial neural networks in pattern recognition*. Springer, Berlin, 57–66.
- Kraus, J.M., Palm, G., Kestler, H.A. (2007), On the robustness of semi-supervised hierarchical graph clustering in functional genomics. *5th International Workshop on Mining and Learning with Graphs*. Florenz, 147–150.

---

<sup>1</sup> Institute of Medical Systems Biology, Ulm University, 89069 Ulm, Germany

# Classification on Large-Scale Data: Systematic Testing in the many-features case

*Claus Weihs<sup>1</sup>, Tobias Kassner<sup>1</sup>*

In this paper we develop a systematic approach for testing the performance of classical classification methods on Large-Scale Data in the case with many more features than observations. We examine, e.g, the influence of the distance of the classes, of the covariance matrix, of the balance between the classes, and of the true error rate on the performance of the classifiers.

---

<sup>1</sup> Computational Statistics, Faculty of Statistics, TU Dortmund, 44221 Dortmund, Germany

`weihs@statistik.uni-dortmund.de`

# Visualization of cross tabulation using association plot

*Yoshiro Yamamoto<sup>1</sup> and Sanetoshi Yamada<sup>2</sup>*

When comparing the response in a survey or medical treatments by some groups, we make the cross-tabulation tables then visualize them by such like bar plot or mosaic plot. For many questions or treatments, we want to find the item that the reaction of a particular group is different from the others. Association rule analysis are suitable for the kind of analysis. By using the coordinates by biplot of correspondence analysis it is possible to plot the relationship between items (treatments) and groups. In this visualization, correspondence analysis and association rules analysis are complement each other. Yamada and Yamamoto [1] introduce this kind of visualization with interactive by using Shiny.

For the medical treatments analysis, it is good to explain with odds ratio.

Then we introduce odds ratio criterion to our visualization method. To introduce this criterion, our interactive visualization has a meaningful reference plot.

## References

- 1 Yamamoto, Y. and Yamada, S. (2016) Visualization of cross tabulation by the Association rules by using the Correspondence analysis, COMPSTAT2016, to appear.
- 2 Yamada, S. and Yamamoto, Y. (2016) Visualization of the Questionnaire Result by the Association Rule Using the Correspondence Analysis, Bulletin of data analysis of Japanese Classification Society, 5(1) 3–15.

---

<sup>1</sup> Department of Mathematics, School of Science, Tokai University. 4-1-1 Kitakaname, Hiratsuka, Japan.

<sup>2</sup> Graduate School of Science and Technology, Tokai University. 4-1-1 Kitakaname, Hiratsuka, Japan.

yama@tokai-u.jp, S.Yamada@star.tokai-u.jp

# The imbalanced class problem revisited

*Adalbert F.X. Wilhelm*<sup>1</sup>

The last decades have seen the invention of a number of modern classification algorithms, such as CART, AdaBoost, support vector machines, and random forests which have shown to result in improved predictions for the expense of diminished interpretability. Imbalanced classes pose a challenge for these data-driven classifiers since the majority class provides more information to fine tune the classifier and equips even naive classifiers with good accuracy performance measures. However, the minority class is typical of higher importance and hence classifiers that yield improved predictions for the minority class are desired. In this presentation, we present an overview on various methods for classifying imbalanced data which each has its own advantages and disadvantages. The evaluation of the algorithms will be illustrated by examples predicting the occurrence of military conflicts.

---

<sup>1</sup> Jacobs University Bremen, Psychology and Methods, Campus Ring 1, 28759 Bremen, Germany

`a.wilhelm@jacobs-university.de`

# Analysis of environmental data with SDA and FDA

Masahiro Mizuta<sup>1</sup>

In our talk, we focus on an environmental dataset. Most of environmental data sets are constructed with location, time, and many factors. We utilize two approaches: Symbolic Data Analysis (SDA) [2] and Functional Data Analysis (FDA) [4]. FDA and SDA are both proposed around 1980's. They gave us new frameworks for complex structured data.

In FDA, the objects in the data sets are represented by functions or curves. When saying in other words, the individuals which were described as the points in  $p$  dimensional space in the frame work of the conventional multivariate analysis, are represented by a set of continuous functions. This representation is good for analysis of transition of the values depended on time.

In SDA, the targets objects are named as "concepts". The concepts have descriptions, e.g. interval values, distributional values, a set of them and so on. This means that SDA takes advantage of various kinds of descriptions and has the flexibility. We use the environmental data; air dose rates measured every 3 seconds by 32 route buses in Fukushima Prefecture. The records of the data include time, latitude, longitude, and air does rate. Total number of records is 13,508,200. This data set is a kind of big data. We analyze it using FDA, SDA and Mini Data approach [4].

## References

- 1 Japan Atomic Energy Agency (2014): Display radiation dose rates in Fukushima Prefecture at real time. TOPICS Fukushima 12 Dec 2014 No.56. <http://fukushima.jaea.go.jp/english/topics/pdf/topics-fukushima056e.pdf>
- 2 Billard, L. ,Diday E. (2006) Symbolic Data Analysis: Conceptual Statistics and Data Mining. John Wiley & Sons.
- 3 Mizuta, M. (2016). Post Big Data may be Mini Data. Proceedings of Hokkaido University and Korea University 4th Workshop in Statistics.
- 4 Ramsay, J. ,Silverman, B. W. (2005). Functional Data Analysis. 2nd edition, New York: Springer-Verlag.

---

<sup>1</sup> Advanced Data Science Lab., Information Initiative Center, Hokkaido University, N11, W6, Kita-ku, Sapporo 060-0811, Japan

`mizuta@iic.hokudai.ac.jp`

# Do Industrial End-Consumer Product-Configurators Use Rational Pricing? A Case Study.

*Andreas Geyer-Schulz<sup>1</sup>, Tino Fuhrmann<sup>1</sup>, Marvin Schweizer<sup>1</sup>*

In this paper we mine an anonymized open data set with car configuration data provided by TNS Infratest. We follow the a priori segmentation of car configurations by model line and engine type used by the car manufacturer. For all these segments we identify all sets of configurations with the same configuration price. For each segment with the same we compute the attribute lattices of car configurations with the same price. With the help of utility theory we formalize tests of rationality and we identify deviations from rationality in the data set. Finally, we discuss the role of rational pricing strategies in end-consumer car configurators and suggest ways of exploiting deviations from rationality in the customer purchase process.

---

<sup>1</sup> Karlsruhe Institute of Technology, Institute of Information Systems and Marketing, Department of Economics and Business Engineering, 76131 Karlsruhe, Germany

`andreas.geyer-schulz@kit.edu`

# Interaction-based co-clustering - but which interactions?

Hans-Hermann Bock<sup>1</sup>

We consider the problem of clustering simultaneously the rows and columns of a real-valued  $I \times J$  data matrix  $X = (x_{ij})$  by corresponding row and columns partitions  $\mathcal{A} = (A_1, \dots, A_m)$  and  $\mathcal{B} = (B_1, \dots, B_n)$ , respectively, with two given numbers  $m$  and  $n$  of classes (two-way clustering, bi-clustering, co-clustering). Rows and columns may correspond, e.g., to individuals and variables in a study, or to the categories of two qualitative predictor variables which may affect the target variable  $x_{ij}$ . Various methods have been proposed for solving such bi-clustering problems, e.g., by optimizing a clustering criterion, by probabilistic co-clustering modeling, by adapting classical  $k$ -means or EM algorithms, or by empirical hierarchical or patchwork approaches. In this paper we concentrate on methods that highlight the interaction structure within or between the 'block clusters'  $A_p \times B_q$  in the data matrix. When browsing through the literature, it appears that it is not at all clear which type of 'interaction' to use since there are various possible options, e.g.:

- (a) the individual interactions  $\gamma_{ij} := x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{\cdot\cdot}$ ,  
i.e., the individual overall deviations from additivity in  $X$
- (b) the blockwise interactions  $g_{pq} := \bar{x}_{A_p, B_q} - \bar{x}_{A_p, \cdot} - \bar{x}_{\cdot, B_q} + \bar{x}_{\cdot\cdot}$ ,  
i.e., the block-specific deviations from additivity in  $X$
- (c) the individual block-specific interactions  $s_{ij} := x_{ij} - \bar{x}_{i, B_q, \cdot} - \bar{x}_{A_p, j} + \bar{x}_{A_p, B_q}$  for  
 $i \in A_p, j \in B_q$ , i.e., the individual cluster-specific deviations from additivity  
in the submatrix  $X^{A_p \times B_q} = (x_{ij})_{i \in A_p, j \in B_q}$ .

This paper presents various co-clustering approaches dealing with these interactions, discusses the underlying models and bi-clustering structures, and cites possible application fields.

## References

- Bock, H.-H. (1980): *Simultaneous clustering of objects and variables*. In: Tomassone, R., Amirchahy, M., Néel, D. (eds.): *Analyse des données et informatique*. Cours de la Commission des Communautés Européennes à Fontainebleau, 19-30 Mars 1979, Institut National de Recherche en Informatique et en Automatique (INRIA), Le Chesnay, France, 187-203.
- I. Van Mechelen, Bock, H.-H., P. De Boeck (2004): *Two-mode clustering methods: a structured review*. *Statistical Methods in Medical Research* 13, 363-394.
- J. Schepers, H.-H. Bock, I. Van Mechelen (2016): *Maximal interaction two-mode clustering*. *J. of Classification* (accepted).

---

<sup>1</sup> Institute of Statistics, RWTH Aachen University, 52062 Aachen, Germany



# Tensor based relational learning for the author name disambiguation

Kei Kurakawa<sup>1</sup> and Yasumasa Baba<sup>2</sup>

Relational learning with tensor factorization has been attracting attention as a technique for data mining, data analysis, and data science (Kolda, 2009; Nickel, 2011). We apply tensor factorization techniques for the author name disambiguation that has been one of the most important issues of digital library research (Ferreira 2012).

In our approach, we represent author similarity matrices of bibliographic citation attributes on tensor slices, and apply tensor factorizations such as CP (CANDECOMP / PARAFAC) decomposition and Tucker decomposition to extract latent feature vectors as of author. Then, we cluster the authors represented by the feature vectors with a clustering technique, k-means to identify the group of citations for each author.

We conducted experiment of our approach with bibliographic citation data that we prepared for two author names. To process the data, we used a tensor library, *scikit-tensor* and a machine learning library, *scikit-learn* and compared effectiveness of our tensor based model to another model with latent variables, LDA.

## References

Ferreira, A. A., Gonçalves, M. A. and Laender, A. H. F. (2012): A brief survey of automatic methods for author name disambiguation. *ACM SIGMOD Record*, 41(2), 15. doi:10.1145/2350036.2350040

Kolda, T. G. and Bader, B. W. (2009). Tensor Decompositions and Applications. *SIAM Review*, 51(3), 455–500. doi:10.1137/07070111X

Nickel, M., Tresp, V. and Kriegel, H.-P. (2011). A Three-Way Model for Collective Learning on Multi-Relational Data. In: *28th International Conference on Machine Learning*. 809–816.

---

<sup>1</sup> National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, 101-8430, Japan

<sup>2</sup> The Institute of Statistical Mathematics, 10-3 Midoricho, Tachikawa, Tokyo, 190-8562, Japan

kurakawa@nii.ac.jp, baba@ism.ac.jp

# Majorization algorithm for dominance point model

*Jun Tsuchida<sup>1</sup>, Hiroshi Yadohisa<sup>2</sup>*

Asymmetric multidimensional scaling (AMDS) is a visualization method that can be applied to asymmetric (dis)similarity data. The dominance point model proposed by Okada and Imaizumi (2007) is an AMDS model. This model represents asymmetry between objects as the difference in the distances between objects and a dominance point. One of the advantages of this model is the easy interpretation of asymmetry because the coordinate vectors of dominance points constitute the parameters of asymmetry. One of the disadvantages of this model is that calculation by using this algorithm is time consuming because it is a non-linear optimization algorithm. In this presentation, we introduce a majorization-minimization algorithm for dominance point model. We decompose the objective function of the dominance point model into symmetric and asymmetric parts. We create majorizing functions of the symmetric and asymmetric parts. Then, we create the update formula. Although the majorization-minimization algorithm is an iterative algorithm, we obtain often simply and explicit update formula under each step. Because the update formula of the majorization-minimization algorithm is simple, we could also introduce a majorization-minimization algorithm for the two-mode three-way dominance point model proposed by Okada and Imaizumi(2015) and Okada and Imaizumi(2015).

## References

Imaizumi, T., and Okada, A. (2015). Dominance Ten Moderu No Ni Sou San Gen Hitaisyo (Hi)Ruizido Heno Kakutyō 2 [Extending Dominance Point Model to Two-Mode Three-Way Asymmetric (Dis)Similarities 2]. *Nihonkodokeiryogakkaihappyoronbunshu*, 43, 406–407. (In Japan.)

Okada, A., and Imaizumi, T. (2007). Multidimensional scaling of asymmetric proximities with a dominance point. In D. Baier, R. Decker, and H. -J. Lenz, (Eds.). *Advances in Data Analysis*. (pp.307–318). Heidelberg-Berlin, Germany: Springer-Verlag.

Okada, A., and Imaizumi, T. (2015). Dominance Ten Moderu No Ni Sou San Gen Hitaisyo (Hi)Ruizido Heno Kakutyō 1 [Extending Dominance Point Model to Two-Mode Three-Way Asymmetric (Dis)Similarities]. *Nihonkodokeiryogakkaihappyoronbunshu*, 43, 404–405. (In Japanese.)

---

<sup>1</sup> Graduate School of Culture and Information Science, Doshisha University, Tatara Miyakodani 1-3, Kyotanabe, Kyoto 610-0394, Japan

<sup>2</sup> Department of Culture and Information Science, Doshisha University, Tatara Miyakodani 1-3, Kyotanabe, Kyoto 610-0394, Japan

# Big Data Clustering: Is Subsampling Better than Fast Pre-clustering?

*Hans-Joachim Mucha*<sup>1</sup>

Often, Big Data is seen as a synonym for a massive dataset simply within the meaning of a very large number of observations (points)  $N$ . In this paper, the focus is on pre-clustering of such Big Data followed by a final clustering. Concretely, we propose a fast data pre-clustering by breaking a cloud of points successively into two clouds at its centroid by a hyperplane. The computational complexity of this iterative binary binning is linear proportional in  $N$ . Pre-clustering reduces the number of observations  $N$  to a much smaller number of micro-clusters  $M \ll N$ , say  $M = 1000$ . Alternatively, the well-known subsampling technique can be used to select  $M \ll N$  observations randomly. However, if  $N$  is very large then the subsampling rate  $M/N$  becomes much smaller than the usual rates of 0.5 to 0.95 known from simulation studies. In any case, a final clustering of the  $M$  micro-clusters or the  $M$  resampled observations, respectively, results in a partition into the desired  $K$  clusters, where usually  $K \ll M \ll N$ . By doing so, the computational complexity of the final clustering doesn't greatly matter with respect to  $N$ . Then the question arises: is our proposed fast pre-clustering better than subsampling? To answer this question the results of some simulation studies are presented.

---

<sup>1</sup> Weierstrass Institute for Applied Analysis and Stochastics (WIAS), 10117 Berlin, Germany



## 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. Ogigara, 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 Simultaneous 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 Vullings  
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 Independent 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 Vullings, Wolfram Schulte, Thilo Schwinn  
An Introduction to TkGofer



- 96-04 Thomas Beuter, Peter Dadam  
Anwendungsspezifische Anforderungen an Workflow-Management-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<sub>flex</sub>* - 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, 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- $P^c$  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 Flexible 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
- 2012-03 Felix Schüssel, Frank Honold, Michael Weber  
Influencing Factors on Multimodal Interaction at Selection Tasks
- 2012-04 Jens Kolb, Paul Hübner, Manfred Reichert  
Model-Driven User Interface Generation and Adaption in Process-Aware Information Systems
- 2012-05 Matthias Lohrmann, Manfred Reichert  
Formalizing Concepts for Efficacy-aware Business Process Modeling
- 2012-06 David Knuplesch, Rüdiger Pryss, Manfred Reichert  
A Formal Framework for Data-Aware Process Interaction Models
- 2012-07 Clara Ayora, Victoria Torres, Barbara Weber, Manfred Reichert, Vicente Pelechano  
Dealing with Variability in Process-Aware Information Systems: Language Requirements, Features, and Existing Proposals
- 2013-01 Frank Kargl  
Abstract Proceedings of the 7th Workshop on Wireless and Mobile Ad-Hoc Networks (WMAN 2013)
- 2013-02 Andreas Lanz, Manfred Reichert, Barbara Weber  
A Formal Semantics of Time Patterns for Process-aware Information Systems
- 2013-03 Matthias Lohrmann, Manfred Reichert  
Demonstrating the Effectiveness of Process Improvement Patterns with Mining Results
- 2013-04 Semra Catalkaya, David Knuplesch, Manfred Reichert  
Bringing More Semantics to XOR-Split Gateways in Business Process Models Based on Decision Rules
- 2013-05 David Knuplesch, Manfred Reichert, Linh Thao Ly, Akhil Kumar, Stefanie Rinderle-Ma  
On the Formal Semantics of the Extended Compliance Rule Graph
- 2013-06 Andreas Steigmiller, Birte Glimm, Thorsten Liebig  
Nominal Schema Absorption

- 2013-07 Hans A. Kestler, Matthias Schmid, Florian Schmid, Dr. Markus Maucher, Johann M. Kraus (eds)  
Statistical Computing 2013 - Abstracts der 45. Arbeitstagung
- 2013-08 Daniel Ott, Dr. Alexander Raschke  
Evaluating Benefits of Requirement Categorization in Natural Language Specifications for Review Improvements
- 2013-09 Philip Geiger, Rüdiger Pryss, Marc Schickler, Manfred Reichert  
Engineering an Advanced Location-Based Augmented Reality Engine for Smart Mobile Devices
- 2014-01 Andreas Lanz, Manfred Reichert  
Analyzing the Impact of Process Change Operations on Time-Aware Processes
- 2014-02 Andreas Steigmiller, Birte Glimm, and Thorsten Liebig  
Coupling Tableau Algorithms for the DL SROIQ with Completion-based Saturation Procedures
- 2014-03 Thomas Geier, Felix Richter, Susanne Biundo  
Conditioned Belief Propagation Revisited: Extended Version
- 2014-04 Hans A. Kestler, Matthias Schmid, Ludwig Lausser, Johann M. Kraus (eds)  
Statistical Computing 2014 - Abstracts der 46. Arbeitstagung
- 2014-05 Andreas Lanz, Roberto Posenato, Carlo Combi, Manfred Reichert  
Simple Temporal Networks with Partially Shrinkable Uncertainty (Extended Version)
- 2014-06 David Knuplesch, Manfred Reichert  
An Operational Semantics for the Extended Compliance Rule Graph Language
- 2015-01 Andreas Lanz, Roberto Posenato, Carlo Combi, Manfred Reichert  
Controlling Time-Awareness in Modularized Processes (Extended Version)
- 2015-03 Raphael Frank, Christoph Sommer, Frank Kargl, Stefan Dietzel, Rens W. van der Heijden  
Proceedings of the 3rd GI/ITG KuVS Fachgespräch Inter-Vehicle Communication (FG-IVC 2015)
- 2015-04 Axel Fürstberger, Ludwig Lausser, Johann M. Kraus, Matthias Schmid, Hans A. Kestler (eds)  
Statistical Computing 2015 - Abstracts der 47. Arbeitstagung
- 2016-03 Ping Gong, David Knuplesch, Manfred Reichert  
Rule-based Monitoring Framework for Business Process Compliance
- 2016-04 Axel Fürstberger, Ludwig Lausser, Johann M. Kraus, Matthias Schmid, Hans A. Kestler (eds)  
Statistical Computing 2016 - Abstracts der 48. Arbeitstagung
- 2016-05 Axel Fürstberger, Johann M. Kraus, Hans A. Kestler (eds)  
Classification 2016 - Abstracts of the 5th German-Japanese Symposium

**Ulmer Informatik-Berichte**  
**ISSN 0939-5091**

**Herausgeber:**  
**Universität Ulm**  
**Fakultät für Ingenieurwissenschaften und Informatik**  
**89069 Ulm**