



Vereniging voor Ordinatie en Classificatie / Dutch-Flemish Classification Society

Chairman: Mark de Rooij, Universiteit Leiden, Faculteit Sociale Wetenschappen, Instituut Psychologie, Methodologie & Statistiek, Postbus 9555, 2300 RB Leiden, Nederland (rooijm@fsw.leidenuniv.nl)

Secretary: Katrijn Van Deun, Universiteit van Tilburg, Faculteit Sociale Wetenschappen, Departement Methoden en Technieken van Onderzoek, Postbus 90153, 5000 LE Tilburg, Nederland (K.VanDeun@uvt.nl)

Treasurer: Tom Wilderjans, Universiteit Leiden, Faculteit der Sociale Wetenschappen, Instituut Psychologie, Methodologie & Statistiek, Postbus 9555, 2300 RB Leiden, Nederland (t.f.wilderjans@fsw.leidenuniv.nl)

Editor: Ralph Rippe, Universiteit Leiden, Faculteit der Sociale Wetenschappen, Instituut Pedagogische Wetenschappen, Algemene en Gezinspedagogiek, Postbus 9555, 2300 RB Leiden, Nederland (rippe@fsw.leidenuniv.nl)

VOC website: <http://www.voc.ac>

Postbankrekening (IBAN) NL86 INGB 0000 161723 t.n.v. Vereniging voor Ordinatie en Classificatie.

6th VOC Conference

May 19th, 2017 Leiden University

Faculty of Social and Behavioral Sciences
Wassenaarseweg 52, 2333 AK, Leiden, Room 1A01

10.00-10:30	Welcome and registration
10.30-11:30	Keynote address: Boudewijn Lelieveldt <i>Fast and scalable non-linear embedding techniques for high-dimensional data</i>
12.20-12:45	VOC Annual Members Meeting
12.45-13:45	Lunch
13.45-15:00	Submitted paper session 2
15.00-15:30	Coffee break
15:30-16:45	Submitted paper session 3
16.45-16:55	Announcement of the PhD Presenter Award Winner
16:55-...	Closing and Drinks

In this issue:

Short program 6 th VOC Conference (Leiden, May 19)	1
Registration details for the 6 th VOC Conference	1
From the President	2
Publications	2
Annual Report of the Secretary for the year 2016	4
Minutes of the Members Meeting 2016 (Tilburg)	4
Agenda Annual Members Meeting 2017 (Leiden)	5
Financial Report for 2016	6
Book of Abstract 6th VOC Conference	7
Route description	18

Registration details for the 6th VOC Conference

Those who would like to join the 6th VOC Conference are welcome and are kindly requested to register through our website <http://www.voc.ac> (go to 'meeting'). Details are provided through the website.

From the President

During the last VOC conference I was elected as the new president of the VOC. My name is Mark de Rooij and I am a professor of Methodology and Statistics in Psychology at Leiden University. I started as a PhD student in Leiden in 1996 under the supervision of one of the founding fathers of our society, Willem Heiser. My research focusses on visualization techniques in models for categorical data. I have been a member of the board of the VOC from 2002 till 2011 and am very happy to return to the board of the society and to cooperate with the other board members. With the rise of data in all parts of society both visualization and classification are among the most important tools to obtain knowledge from these data and I will try to spread this message as often and far as I can.

With a new president we also say goodbye to the previous president, Jeroen Vermunt. I would like to thank Jeroen for all his important work for the society. He introduced the yearly COC conferences. Moreover, he was the initiator for bringing the IFCS conference to Tilburg. Also Ralph Rippe will leave the board of the VOC. Ralph has been active in the board for 6 years and served as the newsletter editor. I would like to thank Ralph for his contributions. In the annual member meeting we will nominate two new board members: Hilde Tobi from Wageningen University and Pieter Schonees from the Erasmus University Rotterdam.

On Friday May 19th we organize the VOC conference in Leiden. It promises to be a very interesting meeting where the keynote speaker will be Boudewijn Lelieveldt who will talk about multidimensional scaling, or ordination, techniques for high dimensional data. Professor Lelieveldt is a professor at the Radiology department of the Leiden University Medical Center and at Delft University. Besides the keynote speaker we have a really interesting set of presentation from both senior and junior members of our society. More details can be found in this newsletter.

In 2017 there is also an IFCS conference, this year in Tokyo. On behalf of the VOC Tom Wilderjans organizes a session "Advanced Techniques for Analyzing (Big) Multi-set/Multi-subject Data" and one of our members and previous Board Member, Marieke Timmerman will give an invited presentation.

I am looking forward to see you all, either in Leiden or in Tokyo!

Mark de Rooij

Publications

Albers, C.J., Gower, J.C., Kiers, H.A.L. (2017). Rank properties for centred three-way arrays. F. Mola, C. Conversano, M. Vichi (eds), Classification, (Big) Data Analysis and Statistical Learning, Studies in Classification, Data Analysis, and Knowledge Organization, Springer, (accepted for publication, preprint available from <http://casperalbers.nl/publications/rankproperties.pdf>).

Ernst, A.F. & Albers, C.J. (2017). Regression assumptions in clinical psychology research practice - A systematic review of common misconceptions. PeerJ (Accepted for publication).

Van der Gaag, M., Kunnen, S., Albers, C.J. (2017). Micro-level mechanisms of identity development: the role of emotional experiences in commitment development. Developmental Psychology, (accepted for publication).

Meijer, R.R., Boevé, A.J., Tendeiro, J.N., Bosker, R.J., Albers, C.J. (2017). The Use of Subscores in Higher Education: When Is This Useful?. Frontiers in Psychology: Educational Psychology, 8(305)

Boevé, A.J., Meijer, R.R., Bosker, R.J., Vugteveen, J., Hoekstra, R., Albers, C.J. (2016).

Implementing the flipped classroom: An exploration of study behaviour and student performance. *Higher Education*, doi:10.1007/s10734-016-0104-y.

Krone, T., Albers, C.J., Timmerman, M.E. (2016). Bayesian dynamic modelling to assess differential treatment effects on panic attack frequencies. *Statistical Modelling*, 16(5): 343-359.

Krone, T., Albers, C.J., Timmerman, M.E. (2016). Comparison of Estimation Procedures for Multilevel AR(1) Models. *Frontiers in Psychology*, 7(486).

Hoendervanger, J.G., De Been, I., Van Yperen, N., Mobach, M., Albers, C.J. (2016). Flexibility in use: Switching behaviour and satisfaction in activity-based work environments. *Journal of Corporate Real Estate*, 18(1), 48-62.

Nierop AFM, Niklasson A, Holmgren A, Glander L, Rosberg S, Albertsson-Wikland K: Modelling individual longitudinal human growth from fetal to adult life - QEPS I. *Journal of theoretical biology* 2016, 406:143-165. doi: 10.1016/j.jtbi.2016.06.007

Holmgren A, Niklasson A, Nierop AFM, A.Aronson S, Sjöberg A, Lissner L, Albertsson-Wikland K. Pubertal height gain is inversely related to peak BMI in childhood. *Pediatric Research* 2017, 81:448-454. doi:10.1038/pr.2016.253

Delaney A, Tamás PA, Tobi, H. Which standards from which disciplines? A test of systematic review for designing interdisciplinary evaluations. *Journal of Development Effectiveness* 2017,9: 82-100. DOI: 10.1080/19439342.2016.1160419

Meijering JV, Tobi H. The effect of controlled opinion feedback on Delphi features: mixed messages from a real-world Delphi experiment. *Technological Forecasting & Social Change* 2016, 103: 166-173. DOI: 10.1016/j.techfore.2015.11.008

Van den Brink A, Bruns D, Tobi H & Bell S. (Eds., 2017) *Research in Landscape Architecture: methods and methodology*. New York: Routledge.

Scheepers P, Tobi H, Boeijen H. (Eds., 2016). *Onderzoeksmethoden [Research methods](9th Ed.)* Amsterdam: Boom Uitgevers.

Windhorst, D. A., Rippe, R. C., Mileva-Seitz, V. R., Verhulst, F. C., Jaddoe, V. W., Noppe, G., ... & Bakermans-Kranenburg, M. J. MILD PERINATAL ADVERSITIES MODERATE THE ASSOCIATION BETWEEN MATERNAL HARSH PARENTING AND HAIR CORTISOL. DIFFERENTIAL SUSCEPTIBILITY TO PARENTING, 77.

Rippe, R. C., Noppe, G., Windhorst, D. A., Tiemeier, H., van Rossum, E. F., Jaddoe, V. W., ... & van den Akker, E. L. (2016). Splitting hair for cortisol? Associations of socio-economic status, ethnicity, hair color, gender and other child characteristics with hair cortisol and cortisone. *Psychoneuroendocrinology*, 66, 56-64.

Rijlaarsdam, J., Pappa, I., Walton, E., Bakermans-Kranenburg, M. J., Mileva-Seitz, V. R., Rippe, R. C., ... & Cecil, C. A. (2016). An epigenome-wide association meta-analysis of prenatal maternal stress in neonates: A model approach for replication. *Epigenetics*, 11(2), 140-149.

Windhorst, D. A., Mileva-Seitz, V. R., Rippe, R. C., Tiemeier, H., Jaddoe, V. W., Verhulst, F. C., ... & Bakermans-Kranenburg, M. J. (2016). Beyond main effects of gene-sets: harsh parenting moderates the association between a dopamine gene-set and child externalizing behavior. *Brain and behavior*, 6(8).

Camfferman, R., Jansen, P. W., Rippe, R. C., Mesman, J., Derks, I. P., Tiemeier, H., ... & van der Veek, S. M. (2016). The association between overweight and internalizing and externalizing behavior in early childhood. *Social Science & Medicine*, 168, 35-42.

Annual Report of the Secretary for the year 2016

1. Number of members

The VOC started end 2015 with 118 members and counted 120 members at the end of 2016. Thirteen memberships were terminated and there were fifteen new members registered. In 2016, 44 members paid contribution.

2. Board

The Board of the VOC was composed as follows in 2016:

Mark de Rooij	President
Katrijn Van Deun	Secretary
Tom Wilderjans	Treasurer
Ralph Rippe	Newsletter Ed. + Webmaster
Jeroen Jansen	Member

The Board met once in 2016. The main topic was the organization of the 5th VOC conference.

3. Activities

The main activity of the VOC was the fifth VOC conference.

The fifth VOC Conference took place at Tilburg University (The Netherlands) on the 27th of May 2016 with a full day program, including twelve contributions by VOC members on a range of topics. A key note contribution was given by Eric Postma (Tilburg University, The Netherlands) on 'The return of Neural Networks'. At this occasion, a best student presentation award was given to Andrei Barcaru (University of Amsterdam) who presented a paper on 'Towards Bayesian analysis of GCxGC and GCxGC-MS data'. The conference had approximately 35 participants.

4. Publicity

The newsletter appeared once. The VOC conference was also announced to non-VOC members, using the IFCS newsletter, the VVS-site and mailing lists (IOPS, BBC, ICS).

Minutes of the VOC Annual Members Meeting (May 27th, 2016, Tilburg)

1. Opening of the meeting

The president, Jeroen Vermunt, opens the meeting.

2. Minutes of the previous meeting

The minutes of the previous meeting were approved.

3. Annual Report

- Members

The number of VOC members

-Miscellaneous

It is possible to do VOC payments through an app for mobile phones. Also online payments can be done.

4. Financial Report

The treasurer, Tom Wilderjans, has eliminated the backlog in contributions. 780 euros of overdue contributions have been paid. So far, 34 members paid of the 50 that are expected to pay; there are about 25 student members (these do not have to pay) and about 25 people who are registered as a non-student member but who did not pay their contributions during the last years. By next year it should be clear whether these people wish to remain a VOC member or not.

Equity is expected once the membership fees have been paid; this is not a lot of money but the VOC has only limited expenses.

5. Cash Committee

Elise Dusseldorp has controlled the accounts together with Lianne Ippel: The accounts are in order. The cash committee commented that the overview was now much clearer than last year. There were no discrepancies between receipts and payments; all is in order.

Marieke Timmerman and Jeroen Vermunt are appointed as the cash committee for the next two years.

6. Composition of the board

The term of the President, Jeroen Vermunt, ends. Mark de Rooij is elected as the new President of the VOC. Ralph Rippe will remain on the board as newsletter editor and webmaster as long as no other board members join the VOC. Because many of the VOC board members are from Leiden University, ideally new board

members are attracted from other organizations. The board is still looking for new board members.

7. Miscellaneous

It is proposed to raise the member fee from 20 to 25 euros. This is approved by the VOC members present.

The issue of having a VOC group through use of social media (LinkedIn, Facebook) is raised. There is no agreement whether this would be beneficial or not.

The VVS is open to collaboration with the VOC or even to let the VOC become a subsection from the VVS. The board is not in favor of becoming a part of the VVS.

8. Closure of the meeting

Mark de Rooij closes the meeting.

Agenda for the VOC Annual Members Meeting (May 19, 2017, Leiden)

1. Opening of the Members Meeting

2. Minutes of the Members Meeting 27th May 2016

The Minutes of this Meeting are included in this Newsletter (see p. 8).

3. Annual Report of the Secretary on the year 2016

The Annual Report is included in this Newsletter (see p. 8).

4. Financial report of the treasurer on the year 2016

The Financial Report is included in this Newsletter (see p. 7).

Report of the cash committee

Marieke Timmerman and Jeroen Vermunt have controlled the accounts. The accounts are in proper order. The cash committee was pleased with the well-organized administration and the adverse balance of about 5000 euros. Further, the measures taken to facilitate the membership payments have had their positive effects. The cash committee would like to compliment the treasurer Tom Wilderjans.

5. Composition of the Board

The board is composed of the following members (with their remaining term, in years, between brackets):

- Mark de Rooij, President (2)
- Katrijn Van Deun, Secretary (0)
- Tom Wilderjans, Treasurer (1)
- Ralph Rippe, Newsletter Editor (0)
- Jeroen Jansen, board member (1)

The terms of Ralph Rippe and Katrijn Van Deun end. Candidates for these positions can register up to 24 hours before the meeting with the President.

The board is looking for additional members. Candidates may contact the President before the meeting. The members may give their opinion during the meeting.

6. Miscellaneous

7. Questions before closure of the meeting

8. Closure of the Members Meeting

Financial Report for 2016

Revenue		Expenditure	
membership fees (44 paying members)	880	transaction costs ING	106,38
overdue membership fees	800	hosting website	121
ADAC subscriptions	124,5	ADAC subscription 2016	41,34
interest savings account	14,07	Gift keynote VOC conference	31,17
Total	1818,57	Total	299,89

Debet		Credit	
Balance ING account	2577,37	Accounts payable	0
Balance savings account	2854,97	Equity	5432,34
Total	5432,34	Total	5432,34

Notes to the balance sheet

- (1) Contributions from 44 members have been collected in 2016
- (2) Compared to their previous year, there is a substantial increase in the equity
- (3) An overview of the evolution of the equity

2016	€ 5.432,34
2015	€ 3.913,66
2014	€ 4.019,92
2013	€ 5.444,46
2012	€ 5.524,70
2011	€ 6.194
2010	€ 7.621
2009	€ 8.189
2008	€ 6.248
2007	€ 5.914
2006	€ 6.869
2005	€ 6.057
2004	€ 5.019
2003	€ 6.795
2002	€ 6.408
2001	€ 5.898
2000	€ 5.731
1999	€ 4.871
1998	€ 5.100



6th VOC Conference
May 19, 2017
Leiden University, The Netherlands
Pieter de la Court building, Room 1A01

Book of Abstracts

Scope

The Dutch/Flemish Classification Society, VOC, aims at communicating scientific principles, methods, and applications of ordination and classification. The VOC is a member of the International Federation of Classification Societies (IFCS).

Program

10:00 Welcome and registration

10:30-11:30 *Keynote address*

10:30 **Boudewijn Lelieveldt** Fast and scalable non-linear embedding techniques
for high-dimensional data

11:30-12:20 *Submitted paper session 1*

11:30 J. Durieux Detecting disease subtypes by means of Cluster Independent
Component Analysis (C-ICA) of multi-subject brain data

11:55 M. Warrens External validity indices for individual clusters

12:20 *VOC Annual Members Meeting 2017*

12:45 Lunch

13:45-15:00 *Submitted paper session 2*

13:45 Y. Han Mechanisms of the transition to adulthood in cross-national
comparison: an application of Hidden Markov Models

14:10 H. Kelderman Latent variable models whose parameters are functions of a
continuous variable

14:35 M. Koeman Comparing dimension reduction and variable selection-based
methods for Fault Diagnosis in High Dimensional Data

15:00 *Coffee and tea break*

15:30-16.45 *Submitted paper session 3*

15:30 Z. Bakk Two-step estimation of models between latent variables

15:55 H. van der Hoef Decomposing information-theoretic validity indices

16:20 X. Li Meta-CART: a flexible approach to identify interaction between
moderators in meta-analysis

16.45 *Announcement of the PhD Presentation Award Winner*

16:55 *Closing and drinks*

KEYNOTE**Fast and scalable non-linear embedding techniques for high-dimensional data****Boudewijn Lelieveldt***Department of Radiology,
Leiden University Medical Center,
Leiden, the Netherlands*

Since 2010, a multi-disciplinary research team at Leiden University Medical Center and Delft University of Technology has been working on a number of novel data analytics techniques that address the analysis and visualization challenges inherent to high-dimensional data. We focused on the non-linear similarity embedding technique tSNE (t-distributed stochastic neighbor embedding): tSNE builds a 2D scatter plot, in which points represent high-dimensional data vectors. These are grouped together in the plot based on their feature profile similarity, while trying to preserve the local neighborhood structure of the high-dimensional data points. tSNE therefore focuses on preserving small differences, while it discards large differences (contrary to PCA).

Expanding on the basic tSNE algorithm, we developed Dual tSNE and linked-view tSNE to enable fast and interactive comparison of multiple networks. Moreover, we developed Approximating tSNE and Hierarchical tSNE to remove the speed and scale limits respectively of tSNE-based approaches. A number of application examples will be discussed, in relation to imaging data, and to –omics data. We developed the publically accessible web portal for mining gene expression in the adult and developmental human brain: the BrainScope.nl portal. Its most prominent feature is the linked, all-in-one visualization of genes and samples across the whole brain and genome, and across development. In addition, we developed Cytosplore, a PC application for fast and interactive cell phenotyping from massive, high-dimensional single cell datasets. Finally, a number of applications in hyperspectral imaging will be discussed.

Detecting disease subtypes by means of Cluster Independent Component Analysis (C-ICA) of multi-subject brain data

Jeffrey Durieux

*Institute of Psychology,
Methodology & Statistics,
Leiden University,
Leiden, the Netherlands*

An emerging challenge in the study of brain diseases and mental disorders, like dementia and depression, consists of revealing systematic differences and similarities between subgroups of patients in functional connectivity patterns (FCPs), that is, coordinated activity across brain regions. As such, existing subtypes of the disease may be characterized in terms of FCPs and disease subtypes may get detected which transcend the current diagnostic boundaries and which show a differential development and prognosis of the pathology.

In order to obtain FCP's, researchers often collect resting-state functional Magnetic Resonance Imaging (rs-fMRI) data and analyze this data with Independent Component Analysis (ICA). ICA is a technique that decomposes a multivariate observed signal into a set of underlying independent source signals and a mixing matrix. In an fMRI context, the sources represent spatial maps, which corresponds to FCPs, and the mixing matrix contains the associated time courses.

Analyzing the brain data of each patient separately with ICA has as major drawback that each patient will be characterized by different FCPs, which makes it difficult to detect the systematic differences and similarities in FCPs between (groups of) patients. Therefore, we propose Cluster Independent Component Analysis (C-ICA). The goal of this method is to cluster the patients into homogenous groups based on the similarities and differences in their FCPs. As such, patients allocated to the same cluster are assumed to have similar connectivity patterns, whereas patients belonging to different clusters will be described by different FCPs. This allows a data-driven detection of disease/disorder subtypes based on different FCPs'.

In this presentation, the C-ICA model is proposed, along with an alternating least squares type of algorithm to estimate its parameters. Further, the results of a simulation study to evaluate the performance of the novel C-ICA method are presented.

External validity indices for individual clusters

Matthijs Warrens

*University of Groningen,
the Netherlands*

Clustering methods are used in many different disciplines to assign observations to meaningful groups. Different clustering methods perform well in different situations, and no clustering method has been shown to dominate other methods across all application domains. In choosing a clustering method it is important that the characteristics of a method are well understood. Therefore, an important and fundamental topic in cluster analysis research is the validation of the cluster results. To evaluate the performance of a clustering method researchers typically assess the agreement between a reference standard partition that purports to represent the true cluster structure of the objects, and a trial partition produced by the method that is being evaluated. High agreement between the two partitions indicates good recovery of the true cluster structure.

Agreement between a reference and a trial partition can be assessed with so-called external validity indices. Commonly used examples from the cluster analysis community are the Rand index and the Hubert-Arabie adjusted Rand index. In the machine learning community, measures from information retrieval, like recall, precision and the F-measure, are commonly used.

If a reference partition has three or more clusters it is usually of interest to assess which clusters were and which were not recovered correctly by the clustering method that is being evaluated. This knowledge provides insights into characteristics of both the clustering method and the data. Furthermore, understanding which clusters were and which were not recovered correctly seems instrumental for understanding the characteristics of clustering methods.

However, in clustering validation studies researchers tend to use overall measures, e.g. the Rand indices or the overall F-measure. Overall measures quantify agreement between two partitions for all clusters simultaneously, and thus, only give a general notion of what is going on. In this presentation we show that overall measures based on counting pairs tend to reflect how well the larger clusters are recovered. They provide less information on the recovery of smaller clusters.

To evaluate the recovery of individual clusters, researchers may use recall (sensitivity, classification rate), precision (positive predicted value) and the F-measure. However, to calculate these indices researchers have to match the true clusters with the found groupings. This arbitrariness of matching results in several problems: the group assignment can be manipulated to generate either more or less favorable classification rates, and partitions can be compared only if they have the same number of clusters.

In this presentation we present four new external validity indices that can be used to assess the recovery of individual clusters. Two indices are alternatives for recall (sensitivity, classification rate). Like the Rand indices, the new indices are based on counting pairs of objects. Furthermore, unlike recall, precision and the F-measure, they do not require arbitrary matching of clusters.

Mechanisms of the transition to adulthood in cross-national comparison: an application of Hidden Markov Models

Sapphire Yu Han

NiDi,

the Netherlands

Recent theories about social and demographic change, such as individualization and the second Demographic Transition (SDT), suggest a type of late, protracted and complex pathway to adulthood. Our previous work demonstrates the application of a first order Hidden Markov model to uncover the mechanisms of transition to adulthood and the roles played by gender and education level of the birth cohort between 1956 and 1965 in France. Methodologically, the Hidden Markov model largely reduces the complex sequence data into life state (hidden state) based transition sequences. Substantively, our result suggests a fertility and partnership driven pathway of transition to adulthood, while covariates played different roles in each of the life states. To further test the applicability of Hidden Markov models and to deepen our understanding of the transition differences between Western countries, we expand the Hidden Markov modeling to a cross-national comparison context. Theoretically, we argue that different Western countries are at different stages of SDT at a given cohort and Hidden Markov models can detect these differences. Therefore, this study adopts a life course approach using Hidden Markov models to quantify the transition to adulthood in a range of European countries representing different welfare regimes. We will test hypothesis on social class- (parental SES, education, etc.) and gender related background variables in state transitions using respondents born between year 1961 and 1970 in Generations and Gender Survey (GGS), which consists full annual monthly life course sequence data of leaving parental home, partnership history and fertility history between age 15 to 35.

Latent variable models whose parameters are functions of a continuous variable

Henk Kelderman

*Institute of Psychology,
Methodology & Statistics,
Leiden University,
Leiden, the Netherlands*

The parameters of a latent variable model may not be invariant through time. If the parameters of the measurement model are not invariant, one or more observed indicators of latent variable(s) may suffer from response shift, which may make the comparisons of latent variable scores through time invalid. If parameters in the population model are not invariant population properties may change, for example factors may merge or diverge or the latent variable means or latent variable variance may decrease or over time. The former may be of methodological interest and the latter may be of substantive interest to the researcher. If item responses are administered at a smallish number of discrete time points these phenomena can be studied with multiple-group latent variable models. In this paper we study the case where time is assumed metric. We present a model that yields smooth functions of model parameters through time. The model is illustrated on a set of Big-Five personality data administered over a time period of 21 years.

Comparing dimension reduction and variable selection-based methods for Fault Diagnosis in High Dimensional Data

Mike Koeman

Radboud University,

Nijmegen, the Netherlands

Identification of abnormal variables in a single sample, i.e. fault diagnosis, is a crucial step in statistical process control (SPC) to inform the researcher about the root cause of a fault [1]. Similarly, in (personalized) health care the aim is to identify abnormal patterns in e.g. metabolomics data of a single patient to diagnose a disease [2].

For fault identification in high-dimensional data some form of feature reduction has to be applied, which typically is a dimension reduction using Principal Component Analysis (PCA). Subsequently, contribution plots based around Hotelling t^2 and the Q-statistic are used to diagnose the fault. It is well-known, however, that reliable identification of the variables primarily associated to the fault is hampered by the so-called smearing effect, which is a result of the dimension reduction step [3].

Recently, several variable selection-based fault diagnosis approaches have been proposed, where the abnormal variables correspond to the first selected variables [4, 5]. The application of these approaches to high-dimensional data does not require dimension reduction. This way, the smearing effect is circumvented, which should result in more reliable fault diagnosis. However, these approaches have their own limitations when it comes to fault diagnosis. For example, in the case of highly correlated abnormal variables it may not be guaranteed that both are selected. The aim of the present work is to compare methods based on dimension reduction to methods based on variable selection for fault diagnosis in high-dimensional data. Simulated data sets are used to highlight the strengths and weaknesses of both approaches.

[1] MacGregor, John F., and Theodora Kourti. "Statistical process control of multivariate processes." *Control Engineering Practice* 3.3 (1995): 403-414.

[2] Engel, Jasper, et al. "Towards the disease biomarker in an individual patient using statistical health monitoring." *PloS one* 9.4 (2014): e92452.

[3] Van den Kerkhof, Pieter, et al. "Analysis of smearing-out in contribution plot based fault isolation for Statistical Process Control." *Chemical Engineering Science* 104 (2013): 285-293.

[4] Wang, Kaibo, and Wei Jiang. "High-dimensional process monitoring and fault isolation via variable selection." *Journal of Quality Technology* 41.3 (2009): 247.

[5] Zou, Changliang, and Peihua Qiu. "Multivariate statistical process control using LASSO." *Journal of the American Statistical Association* 104.488 (2009): 1586-1596.

Two-step estimation of models between latent variables

Zsuzsa Bakk

*Institute of Psychology,
Methodology & Statistics,
Leiden University,
Leiden, the Netherlands*

"We consider models which combine latent class measurement models for categorical latent variables with structural regression models for the relationships between the latent classes and observed explanatory and response variables. We propose a new two-step method of estimating such models. In its first step the measurement model is estimated alone, and in the second step this measurement model is held fixed when the structural model is estimated. Simulation studies and applied examples suggest that the two-step approach is an attractive alternative to existing one-step and three-step methods. We derive variance estimates for two-step estimates of the structural model which account for the uncertainty from both steps of the estimation, and show how the method can be implemented in standard software.

Key words: Latent variables; Mixture models; Structural equation models; Pseudo maximum likelihood estimation"

Decomposing information-theoretic validity indices

Hanneke van der Hoef

*Heymans Institute for Psychological Research,
University of Groningen,
Groningen, the Netherlands*

In (semi-)supervised clustering, external validity indices are used to quantify how well a partition matches a true or 'golden' standard. Over the past years, many external validity indices have been developed, which can be categorized into three approaches: pair-counting, information theoretic and set-matching measures. While the aim of external validity indices is to quantify how well a partition matches a (golden) standard, most external validity indices are overall measures which only provide a general overview of the recovery of clusters. Little information is provided on the recovery of individual clusters.

By decomposing overall measures into information chunks corresponding to individual clusters, more insight can be provided into the recovery of individual clusters. While the decomposition of overall indices can be applied to all three approaches of external validity, differences exist in the weighting of overall measures between the different approaches.

In this presentation, information-theoretic indices will be discussed, in particular two asymmetric versions of the Normalized Mutual Information (NMI), which form the 'building-blocks' for several other information-theoretic indices. While in information-theory normalization is a commonly agreed property to take into account the effect of cluster size, I will show that even these normalized indices still are affected by cluster size imbalance. More specifically, I will show that information-theoretic indices are weighted means, which can be determined using a function of the logarithm of the relative cluster size. Hence, weights depend on the relative size of clusters. This will be shown using several illustrative as well as real-data examples.

Meta-CART: a flexible approach to identify interaction between moderators in meta-analysis

Xinru Li

*Mathematical Institute,
Leiden University,
Leiden, the Netherlands*

Background: Meta-analysis is a valuable tool to quantitatively synthesize findings from multiple studies in a systematic way. It can be used to evaluate the overall outcome (i.e., effect size), and estimate the relationship between study-level covariates (i.e., moderators) and the effect sizes. In many areas, there are often multiple moderators available (e.g., patient characteristics). In such cases, traditional meta-analysis methods often lack sufficient power to investigate interaction effects between moderators, especially high-order interactions. To solve this problem, meta-CART was proposed by integrating Classification and Regression Trees (CART) into meta-analysis. In this study we improved the existing version of meta-CART upon two aspects: 1) to integrate the two steps of the approach into one; 2) to consistently take into account the fixed-effect or random-effects assumption in both the splitting and the interaction detection process.

Method: For fixed effect meta-CART, weights were applied and subgroup analysis was adapted. For random effect meta-CART, a sequential partitioning algorithm was developed. The performance of the improved meta-CART was investigated via an extensive simulation study on different types of moderator variables (i.e., dichotomous, ordinal, and multinomial variables), and via an application study.

Results: The simulation results show that the new methods can achieve good control of Type I error (< 0.05) and power (> 0.80) in general. To achieve good recovery rates of moderators (> 0.80), the number of studies needs to be larger than 40 to identify simple interaction effect, and large than 80 to identify complex interaction effects.

Discussion: The improved version of meta-CART applies the fixed- or random-effects assumption consistently in both detection and test procedure. Researchers may choose between fixed- or random-effects model based on their research question and the assumption of residual heterogeneity. The application example shows that meta-CART is able to identify interaction between moderators and provide interpretable results.

Route description

Faculty of Social and Behavioural Sciences
 Pieter de la Court Building
 Wassenaarseweg 52
 2333 AK Leiden, The Netherlands

1 By car

Route from the A44:

Leave the A44 at exit 8 (exit: Leiden-Valkenburg-Katwijk-Noordwijk from the direction of The Hague, exit: Leiden-Utrecht from the direction of Amsterdam). Take the direction to Leiden centre/Naturalis via the Plesmanlaan.

Route from the A4:

Leave the A4 at exit 7 (Zoeterwoude dorp) en continue on the N206 towards Katwijk. At the large T-crossing with the Plesmanlaan take a right turn towards Leiden centre/Naturalis.

On the Plesmanlaan take the first left turn possible, you are now on the Einsteinweg. Continue this road and turn right at the roundabout, onto the Max Planckweg. Follow this road until the Wassenaarseweg, take a right turn. Drive up to the Pieter de la Court Building of the Faculty of Social and Behavioural Sciences (big square yellow building), before the roundabout.

Paid parking facilities

The Rijnveste car park is to the right of the Pieter de la Court building.

2 By public transport

When you arrive by train take the back entrance/exit in the direction of the the Leiden University Medical Centre (LUMC). At the next roundabout take the first turn to the right (see map).

