



International Federation of Classification Societies Newsletter

President: David Hand

Number 34

November 2007

Editor: André Hardy

The International Federation of Classification Societies, founded in 1985, is composed of:

Associação Portuguesa de Classificação e Análise de Dados, British Classification Society, Central American and Caribbean Society of Classification and Data Analysis, Classification Society of North America, Gesellschaft für Klassifikation, Irish Pattern Recognition and Classification Society, Japanese Classification Society, Korean Classification Society, Société Francophone de Classification, Società Italiana di Statistica, Statistical Society of Slovenia, Vereniging voor Ordinate en Classificatie and Section of Classification and Data Analysis of the Polish Statistical Society.

The IFCS is a non-profit, non-political scientific organization, the aims of which are to further classification research. Among other activities, the IFCS organises a biennial conference, and supports the "Journal of Classification" and the journal "Advances in Data Analysis and Classification (ADAC)".

IFCS Homepage:

<http://www.classification-society.org/>

Designed and maintained by David Dubin.

The website contains among others: the IFCS Constitution and By-Laws, the IFCS newsletters, and pointers to the websites of the member societies.

Contents

Words from the IFCS President
News from the IFCS Scientific Secretary
News from the CLAD
News from the GfKI
News from the VOC
News from the JCS
News from CSNA
News from the SFC
EGC 2008
The Isaac Newton Institute
Journal "Advances in Data Analysis and Classification"
Short book review
Workshop in honour of Edwin Diday

Words from the IFCS President November 2007

This will be my final contribution to the Newsletter as President. It is true to say that the two years have flown by.

Classification is a very broad subject. Even restricting ourselves to the study of data analytic classification methodology, it is apparent that it has many interpretations. At one extreme, some take it as being primarily about methods for discovering natural classes of objects, or for finding convenient ways of partitioning sets of objects. Others (including me) broaden this to include diagnostic and assignment methods. And others broaden this yet further to include more general data analytic tools and methods.

But there is also a whole world of classification beyond the data analytic. For example, National Statistical Offices are concerned with classification of industries, of employers, of health care providers and so on, so that accurate national statistics can be produced. This is clearly vital if a government is going to steer a country and its economy and for such things as bank scorecards for calculating default risk on loans.

Libraries are concerned with systems for classifying books, and have developed approaches such as the Dewey decimal system. The reason for imposing classes here is to permit ready retrieval of relevant books. (Though my personal experience shows that sometimes literary classification schemes break down. I once found a copy of my first book (on statistical classification methods, of course, and entitled *Discrimination and Classification*) in the race-relations section of a bookshop. I suppose I should have been glad that the shop was stocking it at all, but I think anyone who bought it from that section of the shop would have been disappointed when they discovered its content.)

Another organisation which is concerned with classifying information is the World Intellectual Property Organisation (WIPO): “Four WIPO treaties ... created classification systems which organize information concerning inventions, trademarks, and industrial designs into indexed, manageable structures for easy retrieval”. WIPO’s website

(<http://www.wipo.int/classifications/en/>) describes how “WIPO administers four treaties establishing international classifications, the texts of which are also available:

- the Strasbourg Agreement concerning the International Patent Classification (IPC);
- the Nice Agreement concerning the International Classification of Goods and Services for the Purposes of the Registration of Marks;
- the Locarno Agreement establishing the International Classification for Industrial Designs;
- the Vienna Agreement establishing the International Classification of the Figurative Elements of Marks.”

Other important classifications are the International Classification of Diseases, drug and medical device classifications, and many others. Such systems go from familiar systems such as biological taxonomies and the periodic table of the elements to more exotic classifications such as security classifications and film censors’ classifications.

Entering the single word ‘classification’ into Google produced 166 million hits. Reading some of the entries leads one to suspect that the whole of human knowledge is based on classification. And that, of course, is not far from the truth. No wonder, then, that classification is such a fascinating topic for us data analysts.

And then, perhaps unsuspected by many of you, there’s the (perfectly genuine) International Association of Classification Societies - the IASC (<http://www.iacs.org.uk/>). When I first heard of this I wondered if I had slipped into a strange parallel world. Would I find that the Association’s member organisations consisted of the UK Classification Society, the Classification Society of the United States, the Korean Classification Association, and other parallels to the national societies with which I was familiar? In case you are just as puzzled as I was, here is an extract from the Wikipedia entry on classification societies: “Classification societies set technical rules, confirm that designs and calculations meet these rules, survey ships and structures during the process of construction and commissioning, and periodically survey vessels to ensure that they continue to meet the rules. Classification societies are also responsible for classing oil platforms, other offshore structures, and submarines. Included in the survey process is the survey and certification of diesel engines, large or critical pumps such as fire or main bilge pumps, and other machinery vital to the function of the ship. This is frequently done at the manufacturer’s plant, which may be hundreds of miles from the shipyard or drydock.”

Now I know where I have been going wrong - and this might also explain the presence of the bemused marine engineer I met at the last IFCS conference in Ljubljana.

The next IFCS conference will take place in Dresden in 2009, and, if I do not see you before, then I look forward to seeing you there. Meanwhile, from December I will be leaving you in the very capable hands of Buck McMorris.

Best wishes to you all.

David J. Hand
IFCS President

News from the IFCS Scientific Secretary – November 2007

Dear Colleagues and Friends of the Classification,

The IFCS Executive Committee (EC) members met at the beginning of June in Brussels. Among other administrative items, the following main topics were discussed during the meeting:

1. **IFCS 2009:** The EC went through the IFCS Conferences Guidelines with representatives of the GfKI (Claus Weihs) and the Dresden group (Werner Esswein) in charge of the IFCS 2009 organization (www.ifcs2009.de - 13-18 March 2009). Both organizational, scientific and financial aspects were discussed. Such an organization looks already very promising and the EC wishes to thank once again the GfKI and the Dresden group for endeavoring at best to the benefit of IFCS.
2. **IFCS new member societies:** The EC members are still pursuing previously established contacts with groups from China, Spain and Greece who seem to be interested in submitting applications to eventually become new IFCS member societies. It was pointed out that, in the near future, the IFCS shall have the objective to broaden its audience and interests by addressing novel domains and groups of classification researchers outside the classical IFCS community.

3. **Chikio Hayashi Awards Program:** The Guidelines and the Application Form for the Chikio Hayashi Awards Program were updated so as to characterize such an important Award as a real scientific award provided by the IFCS to young researchers. Among the new features of this award: candidates need to be authors of a paper to be presented at the IFCS Conference; it is compulsory that a candidate who has been selected for an award attends the IFCS conference for at least three days and presents a paper in a scientific session; the presentations of the CHA winners will be marked in the scientific program of the IFCS Conference as “Winner of the Chikio Hayashi Scientific Award”.
4. **Financial Documents:** The financial report 2006 and the budget proposal 2007 were prepared by Iven Van Mechelen and sent to the Executive Committee members who approved unanimously. These documents were then sent to the Finance Committee for audit and final approval.

Following the EC meeting in Brussels, the IFCS was called to express three formal votes on administrative issues. The following results were officially delivered on July 21st:

- On a motion put by the IFCS Executive Committee, the IFCS Council has dissolved the IFCS Group-at-Large;
- The IFCS Council has approved Patrick Groenen (term till spring 2011) as a new member of the Election Committee. The Election Committee is now made of:
 - Helena Bacelar-Nicolau (CLAD) – Committee Chair
 - F. James Rohlf (CSNA)
 - Tae Rim Lee (KCS)
 - Partrick Groenen (VOC)
- The IFCS Council has approved Akinori Okada (term till spring 2010) as a new member of the Finance

Committee. The Finance Committee is now made of:

- Reinhold Decker (GfKI)
- Akinori Okada (JCS)
- F. James Rohlf (CSNA)
- Patrick Groenen (VOC)

At present, the IFCS Election committee, chaired by Helena Bacelar-Nicolau, is getting ready to run the election for the next President Elect, Scientific Secretary and Treasurer of the IFCS as the terms of the current officers will be all expiring at the beginning of year 2008. In accordance with the IFCS election rules, voting forms with candidate lists will be sent to the Council members by 20th November, with a deadline of three weeks for answering. The results of this election will be most likely known in the mid of January 2008.

Taking into account the communications I received officially from member societies in the past semester, the current list of IFCS member societies representatives in the IFCS Council is as follows (in alphabetical order by the acronyms of member societies):

- Boris Mirkin (BCS)
- Helena Bacelar-Nicolau (CLAD)
- Maurizio Vichi (CLADAG-SIS)
- Mel Janowitz and F. James Rohlf (CSNA)
- Berthold Lausen and Claus Weihs (GfKI)
- Rozenn Dahyot (IPRCS)
- Yoshiharu Sato and Keiji Jajima (JCS)
- Tae Rim Lee (KCS)
- Patrice Bertrand (SFC)
- Krzysztof Jajuga and Andrzej Sokolowski (SKAD)
- Javier Trejos (SoCCCAD)
- Anuska Ferligoj (SSS)
- Ron Wehrens (VOC)

Finally, please have a look at the IFCS website (www.classification-society.org) for the latest news on member societies, the journals officially supported by the IFCS (Journal of Classification – JoC, and Advances in Data Analysis and Classification – ADAC) and up-to-date information on the IFCS Council composition with contact details of member societies.

I take this opportunity to wish you and your beloved ones all the best for a Merry Christmas and a Happy New Year!

Vincenzo Esposito Vinzi
IFCS Scientific Secretary

News from CLAD

1. Call for Participation

The XV Annual Meeting, JOCLAD 2008, of the *Associação Portuguesa de Classificação e Análise de Dados CLAD* will be held on March 27-29, 2008, at the High School of Enterprise Sciences in the Polytechnic Institute of Setúbal. The official web site is <http://www.esce.ips.pt/projectos/joclad08/>.

2. CLAD participation in ISI2007, ISBIS2007 and ICSMRA2007 International Meetings

As announced before, CLAD co-sponsored the ISI2007 Meeting of the International Statistical Institute, held in the Lisbon Conference Centre (<http://www.isi2007.com.pt/>), and supported the satellite conferences ISBIS2007 (International Symposium on Business and Industrial Statistics) held at the Department of Mathematics of the University of Azores (<http://www.isbis2007.uac.pt/intro.php>) and ICSMRA2007 (International Conference on Statistical Methods for Risk Analysis) held at the Faculty of Medicine of the University of Lisbon (<http://www.fm.ul.pt/ICSMRA2007>).

In the ISI2007 Newsletters the following short CLAD's chronicle was presented:

“Founded June 1994 and Member Society of the International Federation of Classification Societies-*IFCS* since March 1995, the Portuguese Association of Classification and Data Analysis-*CLAD* main goals are the promotion and coordination of the investigation and practice of classification and data analysis. *CLAD* is the privileged national institution to promote the cooperation and to give assistance in fields such as organisation, management and improvement of services and products related to its social purpose, namely in what concerns:

- publications, seminars, applications and national and international meetings;
- representation of the users before the national and international organisations.

CLAD has developed privileged relationships with other Classification and Statistical Societies, and signed cooperation protocols of associate membership with the *Société Francophone de Classification*, the *German GfKl* and the *Associação Brasileira de Estatística*. A new cooperation protocol will be signed this year, hopefully during the *ISI2007* meeting, between *CLAD* and *SPE*, *Sociedade Portuguesa de Estatística*. Besides several *IPM*'s have regularly been organized on behalf of *CLAD* "Sister Society" in the *ISI* meetings since 1997, where the organizer and each invited speaker or discussant belong both to *CLAD* and to a *ISI* and/or *IFCS* member society.

Hence *CLAD* naturally sponsors the *ISI2007* meeting in Lisbon as well as the satellite meetings *ISBIS2007* at the University of Azores and *ICSMRA2007* at the University of Lisbon. Also naturally an *IPM* will be again held on behalf of *CLAD* during the *ISI2007* meeting, this time on the topic *Analysis of Complex Data in Health Sciences*. *CLAD* president Fernando C. Nicolau (2004-present) from New University of Lisbon, past-president Paulo Gomes (2000-2004) from National Statistics Institute-*INE*, and first president Helena Bacelar-Nicolau (1994-2000) from the University of Lisbon will participate in this *IPM*, as well as *IFCS* past president Henk Kiers, Leon Bobrowsky, Dean of the Faculty of Computer Science, Bialystok University of Technology in Warsaw, and Franz Halberg, founder of the Laboratories of Chronobiology at the University of Minnesota".

2. *CLAD* new cooperation protocol

A cooperation protocol of associate membership between *CLAD* and *SPE*, the *Sociedade Portuguesa de Estatística*, member of the International Statistical Institute, was signed in September.

Helena Bacelar-Nicolau

hbacelar@fpce.ul.pt

cladlead@fpce.ul.pt

News from GfKL

39th Meeting „Statistical Computing 2007“ July 1-4, 2007 Schloss Reisenburg , Germany

Organized by the working groups Statistical Computing (GMDS/IBS-DR), Klassifikation und Datenanalyse in den Biowissenschaften (GfKL) and the Working Group Computationale Statistik of the Österreichische Statistische Gesellschaft. The invited Sunday evening lecture was given by Antony Unwin (Augsburg): Data Visualisation of LARGE data sets (Or do we mean large?).

The scientific program of the meeting:

Monday, July 2nd: Korbinian Strimmer (Leipzig) A framework for the simultaneous estimation of tail-area-based and local false discovery rates from arbitrary test statistics; Wiebke Werft (Heidelberg) An extension to the Multtest R Package – multiple testing procedures using logistic regression ; Sabine Kern et al. (Dummerstorf) Gene expression in heterogeneous tissues: Is deconfounding possible?; Natalia Becker et al. (Heidelberg) CrossChip: Correlation Analysis of Gene Expression (Expression Profiling) and DNA Copy Number (arrayCGH) Data; Oscar M. Rueda & Ramon Diaz-Uriarte (Madrid) A method for the analysis of CGH arrays and the computation of probabilistic minimal common regions; Thomas Kneib (München) BayesX - Software for Bayesian Inference in Structured Additive Regression; Julia Braun & Leonhard Held (Zürich) Evaluation and criticism of probabilistic forecasts; Andrea Riebler et al. (Zürich) Bayesian Variable Selection for Detecting Adaptive Genomic Differences among Populations; Robin Nunkesser et al. (Dortmund) Evolutionary Computation for Problems in Computational Statistics; Eugen Grycko (Hagen) On the viscosity of a model fluid; Hadley Wickham (Ames, USA) An implementation of the grammar of graphics in R: ggplot; Achim Zeileis & Kurt Hornik (Wien) Choosing Color Palettes for Statistical Graphics; Karl Moder (Wien) Optimal species distinction by discriminant analysis.

Tuesday, July 3rd: Anthony Rossini (Basel) The Statistical Reality Engine: A platform for exploring statistical graphics in virtual reality; Werner Adler & Berthold Lausen (Erlangen) Bias corrected ROC analysis for visualization of classifier performance; Michael Schomaker et al.

(München) New approaches for Model-selection under missing data; Harald Binder et al. (Freiburg) Penalty selection for optimal prediction performance of sparse microarray survival models; Wolfgang Hartmann (Heidelberg) Variable Selection with Support Vector Machine Algorithms; Holger Fröhlich et al. (Heidelberg) A Bayesian Approach for Estimating Signaling Networks through Nested Effects Models; Lars Kaderali et al. (Heidelberg) Gene Network Inference using a Stochastic Differential Equations Model; Achim Tresch (Mainz) Analysis of Yeast Signalling Pathways by Nested Effects Models applied to Gene Knockout Microarray Data; Ramon Diaz-Uriarte et al. (Madrid) Developing web-based and parallelized biostatistics/bioinformatics applications: ADaCGH as a case example; Claudio Lottaz (Regensburg) Automated Microarray Analysis in Clinical Gene Expression Studies; Uwe Ziegenhagen et al. (Berlin) Yxilon – A Client/Server Based Statistical Environment; Taleb Ahmad et al. (Berlin) Using Wiki to build an e-learning system in statistics in Arabic language.

Wednesday, July 4th: Markus Franke (Karlsruhe) Updating Restricted Random Walk Clusters; Theresa Scharl & Friedrich Leisch (Wien) Quality-Based Clustering of Functional Data: Applications to Time-Course Gene Expression Data; Sergej Potapov et al. (Erlangen) TWIX - Classification trees with extra splits; Raluca Ilinca Schmitt et al. (Dortmund) Logic Regression, a Suitable Method for Diagnostic Classifications?; Alexander Seewald (Wien) Intelligent Data Analysis in Bioinformatics.

Abstracts of the talks and further information under: www.statistical-computing.de/Reisensburg2007

Berthold Lausen

Fall Meeting 2007 of the AG-DANK.

This year, the annual Fall Meeting (Herbsttagung) of the Working Group for Data Analysis and Numerical Classification (AG-DANK) of the German Classification Society was being held at the University of Dresden. It was hosted by Professor Locarek-Junge. The focal topic was "discrete, economic, and financial data sets."

Professor Huschens, also Dresden, contributed an interesting data set for analysis. It pertained to creditworthiness and was analysed by four participants with different methods.

The detailed program was as follows:

Friday, November 16, 2007

14:00--14:10 Hermann Locarek-Junge, Gunter Ritter: Welcome
14:10--14:40 Steffi Höse, TU Dresden: Modeling and estimation of multivariate dependencies in panel data on credit-worthiness
15:10--15:40 Karin Sahmer, IUT Caen: A clustering algorithm to analyse creditworthiness
15:40--16:20 Coffee break
16:20--16:50 Magnus Ekdahl: Uni Linköping: The effect of dependence approximations for (model based) discrete classification
16:50 Analyses of the data set (Stefan Huschens, Gerhard Pöppel, Gunter Ritter, Karin Sahmer)

Saturday, November 17, 2007

9:00 -- 9:30 Achim Mucha, WIAS Berlin: On the cluster analysis of contingency tables
9:30 -- 10:30 Hans-Hermann Bock, RWTH Aachen: Probabilistic models and statistical methods for multivariate interval data
10:30 Farewell

Friday night was reserved for a dinner at the "Homage" where we were discussing private matters and talking shop.

Gunter Ritter

gfk12008: Advances in Data Analysis, Data Handling, and Business Intelligence July 16 – 18, 2008 in Hamburg, Germany

The 32nd Annual Conference of the German Classification Society – Gesellschaft für Klassifikation (GfKl), a member of the International Federation of Classification Societies (IFCS), focuses on data analysis, data handling, and business intelligence. The conference is organized in cooperation with the British Classification Society (BCS) and the Dutch/Flemish Classification Society (VOC). The scientific program will include sessions with invited talks and contributed presentations from a

broad range of topics. Special emphasis will be laid on interdisciplinary research and the interaction between theory and practice.

GfKI 2008 will have an overall best paper award as well as a special best paper award for young researchers. Conference languages are English and German. Participants intending to present a talk are invited to submit an abstract by e-mail. All abstracts will undergo a reviewing process. Please see the conference website for details regarding the required structure of the abstracts. Accepted abstracts will be distributed to the conference participants. The post-conference proceedings will be published in the Springer Series 'Studies in Classification, Data Analysis, and Knowledge Organization'. Detailed information about the submission and refereeing procedure as well as formatting instructions can be found on the conference website <http://gfk12008.hsu-hh.de>

Important Dates:

| | |
|-------------------|------------------------------------|
| Feb. 11, 2008: | Deadline for abstract submission |
| April 14, 2008: | Notification of acceptance |
| July 16–18, 2008: | 32nd GfKI Conference |
| July 31, 2008: | Deadline for full paper submission |
| Oct. 15, 2008 | Notification of acceptance |
| Nov. 10, 2008 | Deadline for camera ready papers |

The scientific program will include sessions with invited talks and contributed presentations from a broad range of topics. Special emphasis will be laid on interdisciplinary research and the interaction between theory and practice. The local organizers of the conference are Wilfried Seidel, Claudia Fantapié Altobelli, Andreas Fink, Hartmut Hebbel, Detlef Steuer and Ulrich Tüshaus of the Faculty of Economics and Social Sciences, Helmut-Schmidt-University Hamburg.

The broad range of relevant topics is illustrated by the following list of intended sections: **Theory and Methods:** Exploratory Data Analysis; Clustering and Classification; Pattern Recognition and Machine Learning; Forecasting; Visualization and Scaling Methods; Mixture Analysis; Bayesian, Neural, and Fuzzy Clustering; Graphs, Trees, and Hierarchies; Evaluation of Clustering Algorithms and Data Structures; Data Science and Innovative Tools;

Data Cleaning and Pre-Processing; Data, Text, and Web Mining; Computational Intelligence; Information Extraction and Retrieval; Personalization and Intelligent Agents; Tools for Intelligent Data Analysis; Knowledge Representation and Discovery.

Applications: Subject Indexing and Library Science; Marketing and Management Science; e-commerce, Recommender Systems, and Business Intelligence; Targeting in Social Sciences; Banking and Finance; Production, Controlling, and OR; Biostatistics and Bioinformatics; Genome and DNA Analysis; Medical and Health Sciences; Archaeology and Geography; Engineering and Environment; Image and Signal Processing; Administrative Record Census; Linguistics and Statistical Musicology; Education; Psychology.

Invited speakers include:

Adi Ben-Israel, Rutgers University, USA
Gilles Bisson, La Tronche, France
Gilles Celeux, INRIA, France
Patrick Groenen, University Rotterdam, Netherlands
Tagashi Imaizumi, Tama University Tokyo, Japan
Sabine Krolak-Schwerdt, Univ. Wuppertal, Germany
Geoff McLachlan, University of Brisbane, Australia
Frank R. McMorris, University of Wisconsin-Illinois, USA
Francesco Palumbo, University of Macerata, Italy
Walter Radermacher, Präsident Statistisches Bundesamt, Wiesbaden, Germany
Erko Stackebrandt, DSMZ, Braunschweig, Germany
Raimund Wildner, GfK, Nürnberg, Germany

The Annual GfKI Conference 2008 will be hosted by the Helmut-Schmidt-University, Hamburg. Hamburg is the second largest city of Germany and, according to its inhabitants, one of the most beautiful cities in the world. Hamburg's geographical location is so favourable that it is within easy reach for the rest of the world. The German research council (DFG) supports the gfk1-2008.

Erlangen, Hamburg and Marburg
Berthold Lausen, Wilfried Seidel
and Alfred Ultsch

News from the VOC

The autumn meeting of the VOC on November 17, 2007 at Leiden University was a successful event with about 30 attendants. The theme 'Biostatistical Applications' was discussed from different perspectives.

Jos Twisk (Free University of Amsterdam) opened the meeting with an overview of different strategies to analyse recurrent event data. He nicely showed that different approaches may yield different solutions and conclusions. Caspar Looman (Department of Public Health, Erasmus MC Rotterdam) discussed a method to quantify the attribution of diseases to the prevalence of disability, with a nice factorial based approach to parsimoniously cover various interactions. Frans Oort (University of Amsterdam) thoroughly introduced various aspects of the issue of response shift, and discussed how response shift can be formally defined in terms of measurement bias. Hein Putter (Leiden University Medical Center) discussed a latent class approach to analyzing longitudinally measured outcomes. The thus obtained longitudinal profiles could be nicely related to predictors. Luc Bonneux (NIDI) gave a lively and highly interesting overview on various form of bias. Richard Gill (Leiden University) closed our meeting with a discussion on the case

Lucia de B., who was sentenced for murder, and where statistic based arguments were used in the process. Richard Gill convincingly argued that the arguments used were absolutely unconvincing.

The up-to-date information on the activities of the VOC is available via

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

Marieke Timmerman

News from the JCS

(I) Reporting Result of Election for the Executive Members of JCS

The new staff of the Japanese Classification Society was elected as the following list according to the articles of the JCS for electing the Executive Members. The tenure is for *April 1, 2007 - March 31, 2009*.

PRESIDENT:

Akinori Okada
Rikkyo (St. Paul's) University
3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo
171-8501, Japan
Tel: +81-3-3985-2319
Fax: +81-3-3985-2833
okada@rikkyo.ac.jp

SECRETARIAT MEMBERS:

Secretary

Kazunori Yamaguchi
Rikkyo (St. Paul's) University
3-34-1, Nishi-Ikebukuro, Toshima-ku, Tokyo
171-8501, Japan
Tel: +81-3-3985-2325
Fax: +81-3-3985-2833
kyamagu@rikkyo.ac.jp

General affairs

Treasurer

Yuki Toyoda
Tama University
4-1-1 Hijirigaoka, Tama, Tokyo 206-0002
Tel: +81-42-337-7265
Fax: +81-42-337-7103
toyoda@tama.ac.jp

Public Relations

Yoshiro Yamamoto
Tokai University
1117, Kita-kaname, Hiratsuka 259-1292,

Japan

yamamoto@sm.u-tokai.ac.jp
Takahiro Tsuchiya
The Institute of Statistical Mathematics
taka@ism.ac.jp
Osamu Yoshimura
Nagasaki University
osamu@net.nagasaki-u.ac.jp

Auditors

Atsuhiko Hayashi
The National Center for University Entrance
Examinations
hayashi@rd.dnc.ac.jp
Akinobu Takeuchi
Jissen Women's University
akitake@univ.jissen.ac.jp

THE EXECUTIVE COMMITTEE MEMBERS:

Tadashi Imaizumi, Noboru Ohsumi, Tatsuo Ohtsu, Yutaka Kano, Koji Kurihara, Yoshiharu Sato, Kazuo Shigemasu, Takahiro Tsuchiya, Yuki Toyoda, Yasumasa Baba,

Fumi Hayashi, Masahiro Mizuta, Masakatsu Murakami, Keiji Yajima, Haruo Yanai, Kazue Yamaoka, Kazunori Yamaguchi, Yoshiro Yamamoto, Ryozo Yoshino

(II) The JCS Symposium

The Japanese Classification Society will be held the symposium on December 22, 2007 at the RENAISSANCE CENTER of Tama University.

The official Web site is <http://wwwsoc.nii.ac.jp/jcs/>

(III) Information about 23rd Annual Research Meeting of the Japanese Classification Society 1

The 23rd Annual Research Meeting is in the planning stage to be held on March, 2008. The detailed announcement will be appeared on the JCS Web-page as <http://wwwsoc.nii.ac.jp/jcs/> in the future.

Yoshiro Yamamoto

Classification Society of North America (CSNA)

CSNA Annual Conference Information

The 2007 annual meeting was held at the University of Illinois at Urbana/Champaign, June 7-10, overlapping with the meeting of the Digital Humanities group. The meeting was sponsored by the National Center for Supercomputing Applications, the UIUC College of Education and the Graduate School of Library and Information Science. The organizers of the meeting were David Dubin and Carolyn Anderson. The keynote speaker was Dr. Michael J. Kurtz of the Smithsonian Astrophysical Observatory. Dr. Kurtz spoke on The Astronomical Information Network. Further information on the annual meeting is available at <http://www.classification-society.org/csna/csna07.html>.

The 2008 Annual Meeting of the Classification Society is being organized by Doug Steinley (Univ. of Missouri) and Bill Shannon (Washington Univ.) to be held at the Washington University campus on June 5-7. The meeting will address various timely clustering and classification problems (many

focused on massively large datasets). Please do not hesitate to contact the organizers Doug Steinley at steinleyd@missouri.edu; Bill Shannon at wshannon@wustl.edu.

Journal of Classification

Issue No. 1, Vol. 24 (2007) of the Journal of Classification was mailed to subscribers in the spring, and Issue No. 2 will go out within the next several weeks. The editor of the Journal is Willem Heiser of Leiden University. Issue No. 1 was mailed with the bibliographic CD, Classification Literature Automated Search Service, now containing several out-of-print books on classification and clustering. This bibliographic service has been edited by Fionn Murtagh. Murtagh will become President of CSNA (and of the British Classification Society) in January, whereupon Michael J. Kurtz of the Smithsonian Astrophysical Observatory will become editor of the bibliographic service.

The Classification Society Distinguished Dissertation Award

The Classification Society Distinguished Dissertation Award, supported by Chapman and Hall/CRC, will be for the best PhD (or approximately equivalent doctoral) dissertation nominated by an annual deadline. The theme is clustering, classification, related areas of data analysis, encompassing both associated theory and/or applications.

The Award is administered by CSNA. An evaluation committee is set up by CSNA to evaluate the nominated theses. This committee should have between 3 and 5 members. The chair of the committee is appointed by the CSNA President. The committee is selected by the committee chair, in conjunction with the CSNA President, and is approved by the CSNA Board.

In 2007-2008, the evaluation committee is: Fionn Murtagh (chair, Dept. of Computer Science, Royal Holloway, University of London), Samantha Prins (Dept. of Mathematics & Statistics, James Madison University), Douglas Steinley (Dept. of Psychological Sciences, University of Missouri-Columbia) and Willem Heiser (Dept. of Psychology, University of Leiden).

Members of the evaluation committee must declare any potential conflict of interest. A member of the evaluation committee cannot play any role in regard to evaluating his/her own PhDs, nor also in regard to nominations where there has been a direct collaborative link (e.g. through joint publications).

Nominations will be through the evaluation committee chair. It may be useful to have printed copies sent to committee members, but in any case web-accessible versions have to be made available. The evaluation panel may seek one or two external referee reports for each nominated dissertation.

Typical criteria for nomination for the Award include: (i) the most innovative or impressive work in theory/methodology, or the most innovative or well developed application(s); and (ii) the literature review has to be thorough.

Nominations are to be received by 1 February each year, from the author (PhD), their advisor/supervisor, or other related person. Nomination includes the name and contact points, an online copy of the dissertation, and a short description of why the dissertation merits the award (see above). A decision is made by the evaluation committee by 1 May. The evaluation committee can recommend one award; an award and a runner-up; two joint winners; or other combinations if they think it appropriate.

For the next submission deadline, 1 February 2008, nominated PhDs have to have been successfully completed in the calendar years 2005, 2006 or 2007.

Awards are ratified by the CSNA Board. Publicity is made in the CSNA and other newsletters, and on relevant websites. The short-list of nominated dissertations will also be openly published.

For non-English language theses, it is requested to provide an extended abstract, and (where relevant) pointers to papers associated with the thesis that have been published.

In 2008 the Award totals US\$ 500 in book vouchers from Chapman and Hall/CRC. The winner will be invited to make a presentation in a special plenary session at the CSNA Annual Meeting. Up to \$500 of the Award winner's travel expenses to the Annual Meeting will be covered by CSNA.

Contact for information and nominations:
fmurtagh at acm dot org

<http://www.classification-society.org/csna/dissaward.html>

The WWW and ascii version of the CSNA Newsletter is made available as a service of the Classification Society of North America (web site: <http://www.classification-society.org/csna/csna.html>). Information on becoming a member of CSNA is available at the CSNA website.

Stanley L. Sclove, Secretary

News from the SFC

Election

Rosanna Verde, Seconda Università degli Studi di Napoli, has been reelected as Vice-President of the SFC.

Gilles Venturini, University François Rabelais, Tours, has been elected as Vice-President of the SFC.



The first joint meeting of the Société Francophone de Classification and the Classification and Data Analysis Group will take place in Caserta, on June, 11th-13th, 2008. The meeting is locally organized by “Facoltà di Studi Politici e per l’Alta Formazione Europea e Mediterranea Jean Monnet - SUN”. The Meeting program includes keynote speakers’ presentations, invited sessions, contributed sessions, poster sessions.

The conference will focus on the following areas of interest:

Classification and Clustering

- Non parametric classification
- Clustering and Dissimilarities

Multivariate statistics and Data Analysis

Applied Multivariate Statistics

Computational Methods

These include the following main topics:

- Association Rules
- Business Intelligence
- Categorical Data Analysis
- Classification and Discrimination
- Clustering
- Computational Bayesian Methods
- Data Mining
- Data Streams and Massive data
- Design of Experiments
- Dimensionality Reduction
- Econometrics and Statistical Finance
- Functional Data Analysis
- Genomics and Microarray Data Analysis
- Graphics and Data Visualization
- Imprecise Data and Fuzzy Methods
- Machine Learning
- Matrix Computations and Statistics
- Metadata and Data Representation
- Multivariate Data Analysis
- Neural Networks and Genetic Algorithms
- Nonparametric Statistics and Smoothing
- Optimization Algorithms
- Partial Least Squares
- Pattern Recognition
- Resampling Methods
- Risk Analysis and Scoring
- Robustness
- Sensometrics
- Spatial Statistics
- Structural Equations Models
-

Important dates:

- Title submission: Dec, 9th, 2007
- Extended abstract submission (2-4 pages): Feb, 20th, 2008
- Notification to Authors: Apr, 7th, 2008
- Revised papers submission: Apr, 27th, 2008
- SFC - CLADAG meeting: June, 11-13th, 2008

- Registration deadline: Apr, 15th, 2008
- Late registration deadline: after Apr, 15th, 2008

Further information are available on the conference website,

www.sfc-cladag.unina2.it.

Vincenzo Esposito Vinzi

SFC 2009

The 16th Francophone Classification Society SFC will be held at the University of Joseph Fourier (<http://www.ujf-grenoble.fr/>) at Grenoble (France), on September 9-11, 2009. The SFC 2009 will be hosted by the TIMC-IMAG laboratory (<http://www-timc.imag.fr/>) of the Joseph Fourier University. The scientific program will include sessions with invited talks and contributed presentations from a broad range of topics. Special emphasis will be laid on interdisciplinary research, and the interaction between theory and practice. The local organizers of the conference are Ahlame Douzal, Gilles Bisson, Anne Guérin, Eric Gaussier, Cécile Amblard and Jérôme Gensel. The first call for the conference will be published in September 2008 (see <http://sfc.enst-bretagne.fr/>).

Ahlame Douzal on behalf of the organizers

Extraction et Gestion des Connaissances, EGC'2008

(Knowledge Discovery and Management)

January 29 - February 1, 2008, Sophia Antipolis,
France

<http://www-sop.inria.fr/axis/egc08>

Honorary chair : Osmar Zaiane, Univ. of Alberta,
Canada

Call for participation

We are pleased to announce that the 8th French Conference "Extraction et Gestion des Connaissances" (EGC'2008) will take place in the INRIA Sophia Antipolis, close to Nice, France, from January 29 to February 1, 2008.

This conference is devoted to theories, methods and applications of Data Mining, Machine Learning, Knowledge Discovery in Data Bases, Knowledge Management and engineering. Its aim is to gather all researchers and company managers in this area in order to present work of high quality and exchange new ideas.

EGC'2008 includes research contributions in several ways: tutorials, workshops, oral presentations, posters, and software demonstrations.

We are also pleased to welcome Invited talks :

- Osmar Zaïane, honorary chair of EGC'2008, Univ. of Alberta, Canada.
- Francoise Soulie Fogelman, KXen, France

- Usama Fayyad, Yahoo! Inc., USA

The proceedings of EGC'2008 will appear in a special issue of the French journal RNTI (Cépaduès publisher). Each paper was read by three reviewers (see the program committee on our web site).

Two special scientific prizes of 1500 euros each will be awarded to the best academic and industrial papers.

Registration will be available on-line (<http://www-sop.inria.fr/axis/egc08>).

Fabrice Guillet and Brigitte Trousse, (Program chair and local chair).

The Isaac Newton Institute

IFCS members will be interested to know that the Isaac Newton Institute at Cambridge, U.K. is sponsoring a research program on “Statistical Methods for Complex, High-Dimensional Data” from January 7 through June 27, 2008. The organizers are David Banks, Peter Bickel, Iain Johnstone, and Mike Titterton.

The opening workshop will be held from January 7-11, and the speakers and schedule are posted on the INI website (www.newton.cam.ac.uk/programmes/SCH).

There will be a second workshop, focusing on bioinformatics, on March 31 to April 4, and a closing workshop on June 23-27. Additionally, there will be a satellite workshop on Bayesian methods for high-dimensional data from April 14-16 that is being organized at the University of Warwick by Mark Steel and Fabio Rigat.

Participation in these workshops is by invitation, due to severe space limitations. The first workshop is full, but the subsequent workshops may have some space available, especially for new researchers. A number of IFCS members are already scheduled to participate, and interested persons are encouraged to contact David Banks (banks@stat.duke.edu).

ADAC - Advances in Data Analysis and Classification

In December, Springer-Verlag will publish issue 3/2007 of the journal ADAC with the following scientific articles:

- Anuradha Roy, Ricardo Leiva: *Discrimination with jointly equicorrelated multi-level multivariate data*
- Nicolas Paul, Michel Terre, Luc Fety: *A global algorithm to estimate the expectations of the components of an observed univariate mixture*
- Alexander K. Seewald, Florian Kleedorfer: *Lambda pruning: an approximation of the string subsequence kernel for practical SVM classification and redundancy clustering*
- Wojtek J. Krzanowski, David J. Hand: *A recursive partitioning tool for interval prediction*
- Claus Weihs, Uwe Ligges, Fabian Mörchen, Daniel Müllensiefen: *Classification in music research.*

The contents of former issues of ADAC can be obtained from the journal's website www.springer.com/11634 where general information and author instructions are provided as well. In 2008 ADAC will publish a Special Issue on ‘Optimisation and Nonconvex Programming in Data Mining’ (see www.gfkl.de) where, e.g., optimisation problems in classification are considered. Guest editors are Le Thi Hoai An (Metz/France), Pham Dinh Tao (Rouen/France), and Gunter Ritter (Passau/Germany). Manuscripts should be sent to

lethi@univ-metz.fr or ritter@fmi.uni-passau.de
before January 31, 2007.

ADAC is supported by IFCS and several IFCS Member societies. Members of all IFCS Member Societies can order the journal for a reduced price if they order it via the Treasurer of their Classification Society (about 39 Euros per year plus VAT).

Hans-Hermann Bock

Short book review

MINING GRAPH DATA. (2007) D.J.Cook and L..Holder (eds.). Hoboken, New Jersey. John Wiley and Sons. ISBN-13 978-0-471-73190-0. ISBN-10 0-471-73190-0.

Data mining is now properly established as a legitimate data analytic discipline - that concerned with extracting interesting or useful structures from large data sets. Indeed such has been the growth of the subject and its application areas that several specialist sub-disciplines of data mining have grown up, including, for example, text mining and image mining, not to mention specialist areas of application, such as mining clinical trial reports.

The book under review is concerned with a relative new area of data mining, but one which is rapidly becoming more important and which will clearly be very important in the future: mining data that are represented as a graph.

After the introductory chapter, the book is divided into three parts.

Part I provides an introduction to graph concepts, terminology, and tools. The first chapter in this part, Chapter 2 (Bunke and Neuhaus), looks at graph matching algorithms and ways of defining distances between graphs. The chapter is brief, so is certainly not a comprehensive review of such tools. Chapter 3 (Didimo and Liotta) describes and illustrates many different ways of displaying graph structures, each useful for bringing out different aspects of the structures. Chapter 4 (Chakrabarti and Faloutsos) describes methods for generating random graphs with given properties. The ability to do this is useful for a

wide variety of reasons (think of the number of ways random number generators are used), including the testing of new algorithms and the generation of distributions under a null hypothesis for testing purposes. I thought that the collection of definitions of different aspects of graphs in this chapter was particularly helpful

Part II looks at computational methods for finding patterns in graph data. Chapters 5 (Yan and Han) and 6 (Kuramochi and Karypis) are concerned with finding frequent substructures - c.f. association rule analysis. The a priori rule is, of course, a basic element in such searches, and it is described here, but the material, especially in Chapter 6, goes much further. This seemed to me to be one of the more comprehensive chapters in the book. Chapter 7 (Cook, Holder, and Ketkar) describes methods of supervised and unsupervised learning of graph structures, and some experimental comparisons between the *Subdue* approach (based on compressing the input graph), the frequent substructure approach, and inductive logic programming. Chapter 8 (Jonyer) describes extracting simple production rules for graph grammars. Chapter 9 (Ohara, Nguyen, Mogi, Motoda, and Washio) describes two algorithms for constructing decision trees from graph data. Chapter 10 (Liquière) describes a formal model for learning from labelled graph examples. Kernel methods are becoming ubiquitous these days, and this book does not disappoint: Chapter 11 (Gärtner, Horváth, Le, Smola, and Wrobel) describe kernel methods for supervised learning on graphs. Chapter 12 (Shimbo and Ito) describes a kernel approach to modelling both importance and the similarity of documents, in the context of bibliometric graphs. Finally in this section, Chapter 13 (Bhattacharya and Getoor) addresses a fundamental problem which arises in many problems representable as graphs. This is the problem of entity resolution, or matching nodes which describe the same entity.

Part III opens with a Chapter 14 (Okada) describing the problems of chemical graphs. It seems to me that Chapters 15 (Zaki) and 16 (Gibson, Kumar, McCurley, and Tomkins) would have sat better in Part II. Chapter 15 is about mining frequent subtrees, and briefly describes bioinformatics applications, and Chapter 16 is concerned with finding dense subgraphs. Finally, Chapter 17 brings us back to applications,

discussing the currently very hot topic of social network analysis.

Some familiarity with basic computer science and data mining is assumed, although no prior knowledge of graph theory is necessary (indeed, very little formal graph theory is used). There is an associated web site at <http://www.eecs.wsu.edu/mgd/> (note - there is a typo in the web address given in the preface of the book itself).

Perhaps inevitably in a collection such as this, there was some overlap, especially in the introductory parts of the chapters, as each author set the scene. Some of this could have been removed, and put into a single section in the opening chapter. Also, several of the chapters describe the authors own, highly specific work. This is understandable, but detracts from the value of the book as an overview and introduction to the area. A broader brush in these chapters, setting the work into context, would have been more useful. Having said that, and with some variation, overall I found most of the chapters excellent. Indeed, my main criticism would be that space constraints mean that many of them necessarily left many important things unsaid. This is hardly the fault of the authors or editors!

Overall the book is very nicely produced and would provide an excellent broad initial introduction to the area and its problems (and indeed opportunities) for a new graduate student entering the area.

David J. Hand
Imperial College, London

Workshop on Data Analysis and Classification in honour of Edwin DIDAY

Edwin Diday's contribution to Classification and Data Analysis, where he often leads the way opening new fields of research, is in many aspects outstanding. To honour his scientific achievements a "Workshop on Data Analysis and Classification", organised by friends and ancient students as a Tribute Session, took place last September 4th in Paris, at CNAM (Conservatoire

National des Arts et Métiers). The event, which counted more than 80 participants coming from 16 different countries, was an opportunity for research friends and ancient students to congratulate Edwin and rejoice with him. Pierre Cazes (Université Paris Dauphine, France) made a touching introduction and summarized Edwin Diday's career and scientific contributions. Fellow distinguished scientists made interesting invited talks; these were: Lynne Billard (University of Georgia, Athens, USA) - "Dependencies and variation components of symbolic interval-value data"; Hans Hermann Bock (RWTH Aachen, Germany) - "A history of the /k/-means algorithm"; David Hand (Imperial College, London, England) - "Data analytic tools for mining personal banking data to detect fraud"; Melvin F. Janowitz (DIMACS Center, Rutgers University, USA) - "Classify then summarize, or summarize then classify"; Ludovic Lebart (CNRS/ENST, Paris, France) - "Stability of principal axes"; Fred R. McMorris (Illinois Institute of Technology, Chicago, USA) - "Consensus Pyramids"; Alfredo Rizzi (Università La Sapienza, Rome, Italy) - "Cryptographie et Statistique". Edwin's address showed that his mind is, as always has been, a wealthy source of ideas and research projects, and made clear the reason for this international tribute. At the end of the day, everyone could make a toast with him at the cocktail organized by CNAM.

The Workshop was also the opportunity to present the Festschrift edited in honour of Edwin Diday. "Selected Contributions in Classification and Data Analysis", P. Brito, P. Bertrand, G. Cucumel, F. De Carvalho (Eds.), has been published by Springer (Heidelberg), in the prestigious collection "Studies in Classification, Data Analysis, and Knowledge Organization". The volume counts 59 contributions (647 pages) from more than 100 authors and contains a refereed collection of papers from internationally recognised researchers in the fields of data analysis and classification. It presents recent advances in data analysis, classification, clustering, multivariate statistics and data mining, from a methodological, theoretical, or algorithmic point of view, combined with a wide range of real-world applications which show how data analysis methods can be applied in specific domains. More information about this volume may be found at the internet address <http://www.springer.com/dal/home/new+%26+fo>

[rthcoming+titles+%28default%29?SGWID=1-40356-22-173750283-0.](http://www.modulad.fr/dac/Home.html)

We all join together in thanking Edwin for his generous contribution to the development of Data Analysis, wishing that he will be active and healthy for many years to come.

For more information about the Workshop, please refer to <http://www.modulad.fr/dac/Home.html> .

For the organizing committee, Paula Brito

Contributions for the coming issue of the IFCS newsletter can be sent to:

andre.hardy@fundp.ac.be

Text files are by preference in ascii or word, with pc format. For graphical materials, by preference GIF is used.

Editor:

André HARDY

Department of Mathematics

FUNDP - University of Namur

8, Rempart de la Vierge

B – 5000 Namur

Belgium

Tel : +32 81 72 49 05

Fax : +32 81 72 49 14



