



# Société Francophone de Classification

Bulletin numéro 21

Novembre 2007

**Président** : Y. Lechevallier

**Vice-présidents** : R. Verde et G. Venturini

**Secrétaire** : A. Hardy

**Trésorier** : B. Leclerc

**Éditeur** : F.-X. Jollois

Société Francophone de Classification, 61 rue de Bruxelles, 5000 Namur, Belgique

<http://sfc.enst-bretagne.fr>

*Correspondance* : A. Hardy, Facultés Universitaires Notre-Dame de la Paix, 8 rempart de la vierge, 5000 Namur, Belgique

Ce bulletin est le vôtre, faites-le vivre en envoyant toutes les informations que vous jugerez utiles à la communauté de la SFC à l'éditeur en utilisant l'adresse suivante : [francois-xavier.jollois@univ-paris5.fr](mailto:francois-xavier.jollois@univ-paris5.fr)

## Sommaire

Mot du président	1
Assemblée générale	2
Annonces	4
Parutions	5
Conférences	7
Divers	7

## Mot du président

Cher(e)s collègues et ami(e)s de la Classification,

Voici notre bulletin annuel qui est réalisé par notre éditeur du bulletin, François-Xavier Jollois. J'en profite pour féliciter nos deux élus aux postes de vice-présidents : Rosanna Verde et Gilles Venturini qui ont été élus cette année à Paris lors de l'Assemblée Générale. Bienvenue à Gilles au bureau et merci à Djamel pour son travail à la SFC et bonne chance comme président d'EGC. Merci à Rosanna d'avoir accepté de renouveler son mandat de vice-présidence.

Je félicite Bernard Fichet de son élection comme *additional member* au Conseil de l'IFCS. Maintenant nous avons deux membres francophones sur les quatre *additional members* du Conseil de l'IFCS ce qui confirme notre activité au sein de l'IFCS et le rôle de la SFC dans le domaine de la classification. J'en profite pour remercier Bernard d'avoir accepté la co-présidence du comité scientifique de nos journées 2008 qui seront conjointes avec celles de CLADAG.

Nos journées 2007 se sont déroulées à Paris à l'ENST et, au nom de la SFC, je remercie Irène Charon, Georges Hébrail et Olivier Hudry pour l'organisation impeccable

et l'accueil qu'ils nous ont réservés et aussi souligner le travail du comité scientifique. Il est prévu qu'une sélection de papiers correspondant aux communications soit publiée dans *Mathématiques et sciences humaines* sous la présidence d'Olivier Hudry. Sur les six conférenciers invités il y avait deux conférenciers européens. La veille de nos journées s'est déroulée au CNAM une journée en l'honneur d'Edwin Diday qui est un des membres fondateurs de la SFC.

Durant cette assemblée générale nous avons retenu Grenoble comme lieu des journées de la SFC en 2009. Il faut se préparer à repartir à l'extérieur de l'hexagone dans la région de Naples au mois de juin 2008. Grâce à notre vice-présidente Rosanna Verde, les 11-13 juin 2008 nous aurons nos journées avec le groupe italien de classification CLADAG à Caserta, jolie petite ville près de Naples. Ce sont les premières journées communes entre la SFC et CLADAG. J'espère que nous serons nombreux à Caserta pour présenter devant nos collègues italiens nos travaux de recherche. Personnellement je suis heureux de pouvoir participer à ces journées franco-italiennes car mon premier congrès, hors hexagone, s'est passé à Naples. Il avait été organisé, de manière magistrale, par Carlo Lauro et m'avait permis de m'intégrer facilement dans la communauté francophone naissante de l'analyse des données. Aussi j'espère que les jeunes chercheurs francophones en classification pourront, au cours de ces journées, avec leurs collègues italiens, créer de nouveaux axes de recherche autour de la classification.

Les 3èmes Rencontres Inter-Associations (RIA's2007) ont eu lieu à Toulouse les 12 & 13 Mars 2007 sur le thème de la recherche et de l'extraction d'information. L'intervenante de la SFC a été Annie Morin et nous avons eu de

nombreuses interventions de doctorants sur le thème de la classification et, je l'espère, futurs membres de la SFC (voir le site : <http://www.irit.fr/RIA07>).

Grâce à notre trésorier, Bruno Leclerc, nous pouvons maintenant devenir membre affilié de la CSNA (Classification Society of North America) et ainsi recevoir le *Journal of Classification* et en 2008 les membres de la SFC pourront bénéficier d'une réduction importante sur l'abonnement au nouveau journal ADAC (Advances in Data Analysis and Classification). Des informations plus précises sur ces avantages vous seront communiquées en début de l'année 2008.

Je voudrais terminer mon propos en vous rappelant que cette année il y a les élections du président et du trésorier de la SFC, l'appel à candidature sera bientôt lancé.

A tous et à toutes, à bientôt.

Yves Lechevallier  
Président de la SFC

---

## Assemblée Générale

Compte-rendu de l'Assemblée Générale de la S.F.C. Paris - 6 septembre 2007.

L'Assemblée Générale est ouverte sous la présidence de Yves Lechevallier.

### 1. Mot du Président de la SFC (Yves Lechevallier)

Yves Lechevallier remercie, au nom de la SFC, toute l'équipe de Paris pour l'organisation des Rencontres SFC'2007.

### 2. Elections (Yves Lechevallier)

Après dépouillement, on constate les résultats suivants :

- **Rosanna Verde** est réélue *Vice-Présidente de la SFC*.
- **Gilles Venturini** est élu *Vice-Président de la SFC*.

Le Président remercie Djamel Zighed, Vice-Président sortant, pour l'excellent travail qu'il a effectué dans le cadre de son mandat.

### 3. Bilan de SFC'2007

Au nom du comité d'organisation, Olivier Hudry remercie la SFC de leur avoir confié la tâche intéressante d'organiser le congrès de la SFC.

- 80 personnes sont présentes au congrès. Parmi les conférenciers, on a déploré seulement trois défections non annoncées.
- Le bilan financier devrait être légèrement positif, malgré le fait qu'un subside espéré de 6000 € du CNRS ne pourra finalement pas être obtenu.
- Les recettes viennent principalement des frais d'inscriptions, de SPAD (les sacs et leur contenu), du CNRS (contribution financière pour les conférenciers invités) et de l'ENST.
- Les dépenses concernent essentiellement les aspects logistiques : repas, repas de gala, pauses café, conférenciers invités, ...
- Les pays d'où sont issus les participants : France, Belgique, Suisse, Italie, Portugal, Allemagne, Canada, Costa Rica, Brésil.

- Une sélection d'articles seront publiés dans la revue " Mathématiques et Sciences Humaines ". Les personnes intéressées doivent soumettre une version longue de leur papier avant le 31 octobre. Olivier Hudry remercie les responsables de la revue " Mathématiques et Sciences Humaines " d'avoir accepté de publier un numéro spécial avec les articles sélectionnés de SFC'2007.

- Olivier Hudry remercie les sponsors : l'ENST, SPAD, le CNRS ainsi que toutes les personnes du comité d'organisation.

Le Président remercie une fois encore très chaleureusement le comité d'organisation et le comité scientifique pour l'excellent travail fourni pour l'organisation de SFC'2007.

### 4. Rapport moral (André Hardy, Secrétaire de la SFC)

#### (a) Membres de la SFC

Voici l'évolution des membres de la SFC depuis 1994.

1994	1995	1996	1997	1998
63	82	63	123	116
1999	2000	2001	2002	2003
119	110	113	118	118
2004	2005	2006	2007	
125	100	96	74+...	

Tous les membres en ordre de cotisation à la SFC reçoivent les Bulletins de la SFC et de l'IFCS.

- (b) Bulletin de la SFC. L'éditeur est François-Xavier Jollois. Le numéro 20 est paru en novembre 2006. Le prochain numéro paraîtra en novembre 2007.

#### (c) Rencontres de la SFC

Voici le nombre de participants aux différentes Rencontres de la SFC.

Lieu	Année	Nombres
Brest	1992	50
Tours	1994	80
Namur	1995	55
Vannes	1996	103
Lyon	1997	107
Montpellier	1998	105
Nancy	1999	82
Pointe-à-Pitre	2001	113
Toulouse	2002	120
Neuchâtel	2003	80
Bordeaux	2004	120
Montréal	2005	86
Metz	2006	70
Paris	2007	80

#### (d) SFC'2007

Les 14<sup>èmes</sup> Rencontres de la SFC ont lieu à Paris du 5 au 7 septembre 2007.

#### (e) SFC'2008

Les 15<sup>èmes</sup> Rencontres de la SFC auront lieu au Belvedere di S. Leucio - Caserta - Italie les 11, 12 et 13 juin 2008. Il s'agit de rencontres

conjointes avec celles de la CLADAG, le groupe de Classification et d'Analyse des Données de la Société Italienne de Statistique (SIS).

(f) Contacts avec d'autres sociétés

- La SFC est membre fondateur de l'IFCS ; des informations sur la SFC sont reprises dans chaque bulletin de l'IFCS ; quatre membres de la SFC participent au Conseil de l'IFCS (Patrice Bertrand, Bernard Fichet, Jean-Paul Rasson et André Hardy). André Hardy est *Publication Officer* de l'IFCS. Il fait donc aussi partie de l'Exécutif de l'IFCS.
- La SFC a des accords privilégiés avec la *Gesellschaft für Klassifikation* et avec l'*Associação Portuguesa de Classificação et Análise de Dados (CLAD)*.
- Une convention a été signée entre la SFC et la SFdS. Un membre de la SFC fait partie du Conseil de la SFdS (Yves Lechevallier). Un membre de la SFdS fait partie du Bureau élargi de la SFC (Christian Derquenne). Une cotisation conjointe est prévue depuis 1999.
- La SFC a parrainé, pour la septième année consécutive, les *Journées francophones d'Extraction et de Gestion des Connaissances* (Namur, Belgique, 23-26 janvier 2007)
- Une collaboration scientifique a lieu entre des associations françaises et francophones (AFIA, ARIA, EGC, INFORSID, SFdS et SFC).
  - La représentante de la SFC est Pascale KUNTZ
  - Les troisièmes Rencontres Inter-Associations se sont déroulées à Toulouse, les 12 et 13 mars 2007, sur le thème *Recherche et extraction d'information*. Le conférencier représentant la SFC était Annie Morin.
- CSNA (Classification Society of North America) : une possibilité est donnée aux membres de la SFC de rejoindre la CSNA comme *membre affilié*, avec quelques avantages substantiels
  - Le CD *Classification Literature Automated Search Service* qui inclut un certain nombre d'ouvrages numérisés
  - L'abonnement au *Journal of Classification*, et la disponibilité en ligne de la totalité des anciens numéros de ce Journal

(g) Site Web et envois électroniques

La SFC est sur Internet. Le site est pris en charge par François-Xavier Jollois. L'adresse du site est <http://sfc.enst-bretagne.fr>. Des liens ont été créés avec d'autres sites pouvant intéresser la SFC (IFCS, SFdS, ...). Cette année, plusieurs annonces ont été faites aux membres de la SFC par courrier électronique (congrès, annonces diverses, ...).

(h) Divers

- Habituellement, l'Assemblée Générale de la

SFC se tient lors des rencontres de la SFC.

- Prochaines élections au Conseil d'Administration de la SFC : les mandats des Président et Trésorier arrivent à échéance.

Le rapport moral est approuvé à l'unanimité.

Une suggestion est faite : inclure dans le fascicule des résumés des communications, lors d'un congrès de la SFC, quelques informations sur la SFC (noms des membres du Conseil d'Administration de la SFC, adresse du site web, ...).

5. **Rapport financier** (Bruno Leclerc)

Rapport Financier pour l'année 2007	
<b>Recettes :</b>	
Cotisations 2005 perçues durant l'année 2006 :	32 €
Cotisations 2006 perçues durant l'année 2006 :	1 215 €
Cotisation 2007 perçue durant l'année 2006 :	16 €
TOTAL	1 263 €
<b>Dépenses :</b>	
IFCS dues pour 2005 et 2006 (frais de virement inclus) :	194,91 €
Frais de tenue de compte :	13 €
TOTAL	207,91 €
Avoir de la SFC au 1er janvier 2006	9 496,45 €
Avoir de la SFC au 1er janvier 2007	10 551,54 €

**Le rapport financier est approuvé à l'unanimité.**

- Bruno Leclerc suggère que les rapports financiers futurs soient dorénavant rédigés par année civile.
- On suggère de réfléchir à l'existence d'une inscription réduite *senior* ou *retraité* pour les Rencontres de la SFC.

6. Bulletin de la SFC et site internet.

La rédaction du Bulletin de la SFC et la tenue du site web incombent à François-Xavier Jollois. François-Xavier envisage de revenir à la publication de deux numéros du Bulletin de la SFC par an. Il encourage les membres de la SFC à lui faire parvenir toutes les informations sur les activités pouvant intéresser les membres de la SFC : thèses, numéros spéciaux de revues, livres, workshop, congrès, ...

7. Revue ADAC

Hans Bock rappelle que la revue *ADAC* est officiellement supportée par l'IFCS (Fédération Internationale des Sociétés de Classification) au même titre que le *Journal of Classification*. Cette revue est publiée par Springer. Elle a été initiée par les Sociétés Allemande, Italienne et Japonaise de Classification. Deux numéros sont parus. Springer accorde une réduction importante aux membres de la SFC. Pour obtenir ce tarif réduit, il faut remplir le formulaire prévu à cet effet, et le renvoyer au trésorier de la SFC (Bruno Leclerc), qui le renvoie à Springer. Le paiement doit se faire par carte de crédit.

8. Cotation SFC et cotation conjointe SFC/SFdS

On décide de ne pas modifier les montants des cotisations à la SFC pour 2008, sauf si la SFdS modifie le montant de ses cotisations. Dans ce cas les cotisations de la SFC seraient adaptées de manière à garder le différentiel habituel de 1 €.

9. SFC 2009

Ahlame Douzal présente la candidature de Grenoble. Les 16e rencontres de la Société Francophone de Classification (SFC) se tiendraient à l'université Joseph Fourier (<http://www.ujf-grenoble.fr>) à Grenoble du 9 au 11 septembre 2009. Elles seraient organisées par le laboratoire TIMC-IMAG (<http://www-timc.imag.fr>) de l'université Joseph Fourier. Le programme scientifique comporterait des sessions invitées et des contributions couvrant de larges thématiques connexes à la classification. Un intérêt particulier serait accordé à l'interdisciplinarité et à l'interaction entre la théorie et la pratique. Les membres du comité d'organisation seraient Ahlame Douzal, Gilles Bisson, Anne Guérin, Eric Gaussier, Cécile Amblard et Jérôme Gensel. La première annonce de la conférence serait publiée courant septembre 2008 (voir <http://sfc.enst-bretagne.fr>).

**Le projet est voté à l'unanimité.**

Deux remarques sont faites :

- Si les organisateurs peuvent inclure dans les thèmes du congrès leurs sujets de recherche, il est indispensable que les thèmes traditionnels des congrès de la SFC soient toujours présents.
- On demande que les membres organisateurs des rencontres de la SFC soient membres de la SFC.

10. SFC 2010

Henri Ralambondrainy réitère la volonté de l'Ile de la Réunion d'organiser les rencontres de la SFC. Les deux problèmes principaux sont l'éloignement par rapport à l'hexagone, et le coût d'un séjour en période touristique. Les dates idéales sont mars ou décembre. Le prix d'un billet d'avion pourrait alors descendre à environ 500 euros. Des propositions plus précises seront faites l'année prochaine.

11. SFC 2008

**Les 15ièmes Rencontres de la SFC auront lieu au Belvedere di S. Leucio - Caserta - Italie les 11, 12 et 13 juin 2008.** Il s'agit de rencontres conjointes avec le groupe de Classification et d'Analyse des Données de la Société Italienne de Statistique (CLADAG). Quelques informations :

- Des compagnies low cost effectuent des vols entre différentes villes européennes et Naples
- Le membre de la SFC proposé pour co-présider le comité scientifique est Bernard Fichet. La CLADAG devrait désigner un second co-président.
- Rosanna Verde et Antonio Irpino feront partie du comité d'organisation.
- Un résumé de 4 pages, en français, en italien ou en anglais, sera demandé aux personnes qui proposeront une communication. La présentation se fera dans les trois langues, mais on conseille d'utiliser des slides en anglais. Après un processus de

sélection, des articles longs en anglais seront publiés dans des actes publiés chez SPRINGER. Les articles longs en français seront publiés dans un numéro spécial de la revue RNTI.

12. Postes vacants au Conseil d'Administration de la SFC

**Les postes de Président et de Trésorier seront déclarés vacants au Conseil d'Administration de la SFC.** Un appel à candidature sera lancé en temps utile.

13. Divers

Le Président de la SFC propose d'adjoindre au Conseil d'Administration trois membres supplémentaires, vu les contacts de plus en plus nombreux que nous avons avec d'autres sociétés.

André HARDY, Secrétaire de la SFC  
Bruno Leclerc, Trésorier de la SFC

---

## Annonces

### Nomination

Nous avons la joie d'annoncer la récente nomination de Bernard Fichet (SFC) au poste de membre additionnel au conseil de l'IFCS.

De plus, Claus Weihs (GfKL) a lui aussi été nommé comme membre additionnel au conseil de l'IFCS.

— — —

### MIXMOD

Mixmod 2.0 est disponible.

Cette nouvelle version majeure inclut le traitement des données qualitatives (grâce aux modèles de mélanges multivariés multinomiaux). Bien évidemment, MIXMOD traite toujours les données quantitatives (grâce aux modèles de mélanges multivariés gaussiens).

MIXMOD (MIXture MODelling) est un logiciel permettant de traiter des problématiques d'estimation de densités, de classification ou d'analyse discriminante. Il propose un large choix d'algorithmes (EM) et ses versions stochastique (SEM) ou de classification (CEM) afin de maximiser la vraisemblance ou la vraisemblance complétée.

MIXMOD peut être utilisé sur des données quantitatives (grâce aux modèles de mélanges gaussiens multidimensionnels) ou qualitatives (grâce aux modèles de mélanges multinomiaux multidimensionnels). De plus, des critères (BIC, ICL, NEC, CV) permettent de sélectionner le meilleur modèle parmi plusieurs modèles parcimonieux proposés (14 modèles sont proposés dans le cadre quantitatif et 5 dans le cadre qualitatif).

Ecrit en C++, MIXMOD est interfacé avec Scilab et Matlab et distribué sous la licence GNU (GPL).

<http://www-math.univ-fcomte.fr/mixmod/fr/index.php>

— — —

### MClust

La nouvelle version de MClust 3 est disponible depuis le 16 novembre, sous le format d'un package du logiciel R. Il est dédié aux modèles de mélanges gaussien, à la classification, la discrimination et l'estimation de densités.

<http://www.stat.washington.edu/mclust>

— — —

### Thèses et HDR

Voici une liste non-exhaustive de thèses et d'habilitation à diriger la recherche soutenue au cours de l'année 2007 :

- Alissar NASSER, *Contribution à la classification non-supervisée par machines à noyaux*, Thèse soutenue le 23/10/2007 à ULCO - Centre Universitaire de la Mi-Voix, Calais.
- Sergiu Theodor CHELCEA, *Méthodes de classification pour l'analyse des usages des NTIC*, Thèse soutenue le 23/03/2007, INRIA
- Juliette BLANCHET, *Modèles markoviens et extensions pour la classification de données complexes*, Thèse soutenue le 10/10/2007, INRIA
- Matthieu VIGNES, *Modèles markoviens graphiques pour la fusion de données individuelles et d'interactions : application à la classification de gènes*, Thèse soutenue le 30/10/2007, INRIA
- Pierre GANCARSKI, *Classification automatique collaborative multi-stratégies. Application à la télédétection*, HDR soutenue le 25/10/2007, LSIIT, ULP Strasbourg

---

## Parutions

### MODULAD

Le numéro 36 de la revue en ligne Modulad est sorti en mai 2007. C'est un numéro spécial sur l'analyse de flux de données, issu de l'atelier dédié à cette thématique qui s'est déroulé en Italie, à Caserta, en mars 2007.

<http://www.modulad.fr>

— — —

### ADAC

Le deuxième numéro de la nouvelle revue *Advances in Data Analysis and Classification*, éditée par Springer, est paru en août de cette année. Celle-ci est éditée par H.-H. Bock, W. Gaul, A. Okada et M. Vichi. Les articles du premier numéro sont accessibles en ligne sur le site internet de l'éditeur. Voici le sommaire :

1. *Editorial*, Hans-Hermann Bock
2. *Fuzzy clustering based on nonconvex optimisation approaches using difference of convex (DC) functions algorithms*, Hoai An Le Thi, Hoai Minh Le et Tao Pham Dinh
3. *Variable selection in discriminant analysis based on the location model for mixed variables*, Nor Idayu Mahat, Wojtek Janusz Krzanowski et Adolfo Hernandez

4. *Fast calibrations of the forward search for testing multiple outliers in regression*, Marco Riani et Anthony C. Atkinson

5. *Convex ordering criteria for Lévy processes*, Jan Berghentum et Ludger Rüschendorf

<http://www.springerlink.com>

— — —

### Journal of Classification

Le dernier numéro de la revue *Journal of Classification*, éditée elle-aussi par Springer, est parue en juin 2007. Voici le sommaire :

1. *Editorial*, Willem J. Heiser
2. *The Haar Wavelet Transform of a Dendrogram*, Fionn Murtagh
3. *On a General Transformation Making a Dissimilarity Matrix Euclidean*, Jacques Benasseni, Mohammed Bennani Dosse et Serge Joly
4. *Default Priors for Neural Network Classification*, Herbert K.H. Lee
5. *Simultaneous Component and Clustering Models for Three-way Data : Within and Between Approaches*, Maurizio Vichi, Roberto Rocci et Henk A.L. Kiers
6. *Initializing K-means Batch Clustering : A Critical Evaluation of Several Techniques*, Douglas Steinley et Michael J. Brusco
7. *Asymmetric Agglomerative Hierarchical Clustering Algorithms and Their Evaluations*, Akinobu Takeuchi, Takayuki Saito et Hiroshi Yadohisa
8. *Book Reviews*

<http://www.springerlink.com>

— — —

### Éléments de classification : aspects combinatoires et algorithmiques

de François Brucker et Jean-Pierre Barthélemy, dans la collection *Méthodes stochastiques appliquées* dirigée par Nikolaos Limnios et Jacques Janssen, Hermes-Lavoisier, Paris, septembre 2007, 438 pages.

<http://www.lavoisier.fr>

L'objectif de la classification est de regrouper des objets d'étude selon des critères de ressemblance et de séparer ceux qui sont dissemblables selon des critères de séparation.

Éléments de classification met l'accent sur des comparaisons à l'aide de dissimilarités. Lorsque les données sont qualitatives, des pratiques de recodage permettent de se ramener dans un cadre proprement métrique (appelé espace de représentation).

Après une étude des modèles classiques (partitions, hiérarchies, etc.), cet ouvrage s'intéresse à des modèles admettant des classes « empiétantes » et établit des théorèmes de bijection entre ces systèmes de classes et des modèles de dissimilarités.

En dépit de la difficulté algorithmique des problèmes de classification, il met en évidence diverses instances polynomiales que des algorithmes exacts peuvent résoudre. Il présente aussi, lorsque la situation s'y prête, des algorithmes heuristiques.

— — —

## Data clustering : theory, algorithms & applications

de Gan Guojun, Ma Chaoqun, Wu Jianhong, *ASA-SIAM Series on statistics & applied probability*, Lavoisier, août 2007, 480 pages.

<http://www.lavoisier.fr>

Cluster analysis is an unsupervised process that divides a set of objects into homogeneous groups. This book starts with basic information on cluster analysis, including the classification of data and the corresponding similarity measures, followed by the presentation of over 50 clustering algorithms in groups according to some specific baseline methodologies such as hierarchical, centre-based, and search-based methods. As a result, readers and users can easily identify an appropriate algorithm for their applications and compare novel ideas with existing results.

The book also provides examples of clustering applications to illustrate the advantages and shortcomings of different clustering architectures and algorithms. Application areas include pattern recognition, artificial intelligence, information technology, image processing, biology, psychology, and marketing.

Suitable as a textbook for an introductory course in cluster analysis or as source material for a graduate-level introduction to data mining. Readers learn how to perform cluster analysis with the C/C++ and MATLAB programming languages. Provides examples of clustering applications to illustrate the advantages and shortcomings of different clustering architectures and algorithms. Suitable for use as a textbook for an introductory course in cluster analysis or as handbook/reference for graduates and researchers.

— — —

## Cluster Analysis for Data Mining and System Identification

de János Abonyi et Balázs Feil, Birkhäuser Basel, Springer, août 2007, 303 pages.

<http://www.springer.com>

This book presents new approaches to data mining and system identification. Algorithms that can be used for the clustering of data have been overviewed. New techniques and tools are presented for the clustering, classification, regression and visualization of complex datasets. Special attention is given to the analysis of historical process data, tailored algorithms are presented for the data driven modeling of dynamical systems, determining the model order of nonlinear input-output black box models, and the segmentation of multivariate time-series. The main methods and techniques are illustrated through several simulated and real-world applications from data mining and process engineering practice.

The book is aimed primarily at practitioners, researchers, and professionals in statistics, data mining, business intelligence, and systems engineering, but it is also accessible to graduate and undergraduate students in applied mathematics, computer science, electrical and process engineering. Familiarity with the basics of system identification and fuzzy systems is helpful but not required.

— — —

## Data Streams : Models and Algorithms

de Charu C. Aggarwal (Ed.), *Series : Advances in Database Systems*, Vol. 31, Springer, 2007, 354 pages.

<http://www.springer.com>

This book primarily discusses issues related to the mining aspects of data streams and it is unique in its primary focus on the subject.

This volume covers mining aspects of data streams comprehensively : each contributed chapter contains a survey on the topic, the key ideas in the field for that particular topic, and future research directions. The book is intended for a professional audience composed of researchers and practitioners in industry. This book is also appropriate for advanced-level students in computer science.

— — —

## Advances in Fuzzy Clustering and its Applications

de J. Valente De Oliveira et W. Pedrycz (Eds.), John Wiley & Sons, avril 2007, 454 pages.

<http://www.wiley.com>

Fuzzy clustering is now a mature and vibrant area of research with highly innovative advanced applications. Encapsulating this through presenting a careful selection of research contributions, this book addresses timely and relevant concepts and methods, whilst identifying major challenges and recent developments in the area. Split into five clear sections, Fundamentals, Visualization, Algorithms and Computational Aspects, Real-Time and Dynamic Clustering, and Applications and Case Studies, the book covers a wealth of novel, original and fully updated material, and in particular offers :

1. a focus on the algorithmic and computational augmentations of fuzzy clustering and its effectiveness in handling high dimensional problems, distributed problem solving and uncertainty management.
2. presentations of the important and relevant phases of cluster design, including the role of information granules, fuzzy sets in the realization of human-centricity facet of data analysis, as well as system modelling
3. demonstrations of how the results facilitate further detailed development of models, and enhance interpretation aspects
4. a carefully organized illustrative series of applications and case studies in which fuzzy clustering plays a pivotal role.

This book will be of key interest to engineers associated with fuzzy control, bioinformatics, data mining, image processing, and pattern recognition, while computer engineers, students and researchers, in most engineering disciplines, will find this an invaluable resource and research tool.

— — —

## Selected Contributions in Data Analysis and Classification

de P. Brito, P. Bertrand, G ; Cucumel, F. De Carvalho (Eds.), *Series : Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, novembre 2007, 634 pages.

<http://www.springer.com>

This volume presents recent methodological developments in data analysis and classification. It covers a wide range of topics,

including methods for classification and clustering, dissimilarity analysis, graph analysis, consensus methods, conceptual analysis of data, analysis of symbolic data, statistical multivariate methods, data mining and knowledge discovery in databases. Besides structural and theoretical results, the book presents a wide variety of applications, in fields such as biology, micro-array analysis, cyber traffic, bank fraud detection, and text analysis. Combining new methodological advances with a wide variety of real applications, this volume is certainly of special value for researchers and practitioners, providing new analytical tools that are useful in theoretical research and daily practice in classification and data analysis.

---

## Conférences

### SFC 2008

#### Appel à communication

Les 15<sup>èmes</sup> Rencontres de la Société Francophone de Classification auront lieu du 11 au 13 juin 2008, au Palais Royal de Caserta, Italie. Ce seront les premières rencontres de la SFC organisées conjointement avec le Groupe de "Classification et Analyse des Données" de la Société Italienne de Statistique (CLADAG).

Quelques informations :

- Plusieurs hôtels sont disponibles, à ces dates, au centre ville de Caserta
- Caserta est joignable par avion depuis la France, la Belgique et la Suisse (aéroport : Naples Capodichino) (plusieurs vols sont assurés chaque jour par des compagnies nationales et « low cost ») ; à la sortie de l'aéroport, un bus de ligne a pour destination la « gare centrale » de Naples ; là il faut prendre soit un train, soit le bus, pour la gare de Caserta.
- Les droits d'inscription seront, dans la mesure du possible, presque les mêmes qu'à Paris
- Date limite pour la soumission de résumés de communications (maximum 4 pages) : 20 février 2008
- Une publication des meilleurs articles sera organisée après le colloque dans un livre Springer, série « Classification and Data Analysis » pour les papiers en anglais et dans la revue RNTI pour les papiers en français. Des informations complémentaires sont disponibles sur le site web : <http://www.sfc-cladag.unina2.it>

---

### SFdS

Les 40<sup>èmes</sup> Journées de Statistique de la Société Française de Statistique (SFdS) seront organisées conjointement avec le 38<sup>ème</sup> Congrès de la Société Statistique du Canada (SSC) au Centre des Congrès d'Ottawa (Canada) du 25 au 29 mai 2008.

[http://www.ssc.ca/2008/index\\_f.html](http://www.ssc.ca/2008/index_f.html)

---

### NCP'07

La conférence NCP'07 se déroulera du 17 au 21 décembre 2007, à l'INSA de Rouen, France.

<http://ncp07.insa-rouen.fr>

The Conference NCP07 marks the good period of 25 years of extensive and successful developments in these fields, and, at the same time, pays a tribute to Prof. Hoang Tuy on the occasion of his 80th birthday, in recognition of his pioneering achievements which considerably affected the field of Global Optimization. It will be held at the National Institute for Applied Sciences (INSA), Rouen, France during December 17-21, 2007.

The International Scientific Committee of NCP07 will consist of 100 members of about 30 countries all the world over. The Conference will be composed of featured symposia, plenary lectures, invited sessions, and contributed sessions. It aims to address to a large audience that includes both researchers from academia and practitioners from industry.

We are pleased to invite you to submit extended abstracts and proposals of special sessions. A session consists of 4 talks, 25 min each. Companies interested in presenting their products/software/methodologies are invited to contact the conference secretariat (see the menu item "Contacts"). Further information on the Conference can be found in the website of the Conference.

Special Issues dedicated to NCP07 will publish selected papers presented at this Conference. More details about the papers submissions can be found in the menu item "CFP - Special Issues".

---

### EGC 2008

La 8<sup>ème</sup> édition de la conférence francophone **Extraction et Gestion des Connaissances** (EGC'2008) aura lieu à l'INRIA Sophia Antipolis près de Nice, du 29 Janvier au 1er Février 2008.

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

---

### SFC 2009

La 16<sup>e</sup> rencontre de la Société Francophone de Classification (SFC) se tiendra à l'université Joseph Fourier (<http://www.ujf-grenoble.fr>) à Grenoble du 9 au 11 septembre 2009. Elle sera organisée par le laboratoire TIMC-IMAG (<http://www-timc.imag.fr>) de l'université Joseph Fourier. Le programme scientifique comportera des sessions invitées et des contributions couvrant de larges thématiques connexes à la classification. Un intérêt particulier sera accordé à l'interdisciplinarité et à l'interaction entre la théorie et la pratique. La première annonce de la conférence sera publiée courant septembre 2008 (voir le site de la SFC).

---

## Divers

### Compte-rendu de l'atelier *Data Analysis and Classification in honor of Edwin DIDAY*

Une journée sur l'Analyse des Données et la Classification, et en hommage aux contributions scientifiques d'Edwin Diday, a eu lieu le 4 septembre 2007 à Paris au CNAM

(Conservatoire National des Arts et Métiers). Cette journée, qui a compté plus de 80 participants venant de 16 pays différents, fut l'occasion de réunir les chercheurs qui connaissent Edwin Diday, ainsi que ses anciens étudiants. La journée a commencé par une introduction teintée d'émotion où Pierre Cazes (Université Paris Dauphine, France) a résumé la carrière d'Edwin Diday et ses contributions scientifiques. Puis des chercheurs de renommée internationale ont accepté, pour cette journée, de présenter les exposés suivants :

- 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 Diday a ensuite prononcé quelques mots proposant notamment plusieurs nouveaux thèmes de recherche, et finalement un cocktail organisé par le CNAM a eu lieu à la fin de cette journée. La journée fut aussi l'occasion de présenter l'ouvrage édité en l'honneur d'Edwin Diday, qui a pour titre *Selected Contributions in Classification and Data Analysis* et qui a été publié par Springer (Heidelberg) dans la collection *Studies in Classification, Data Analysis, and Knowledge Organization*. Cet ouvrage, co-édité par P. Brito, P. Bertrand, G. Cucumel, F. De Carvalho, inclut 59 contributions (647 pages) de plus de 100 auteurs. Il s'agit d'un ensemble de textes proposés par des chercheurs reconnus au plan international dans les domaines de l'Analyse des Données et la Classification. Des avancées récentes sont présentées en analyse des données, en classification supervisée ou non supervisée, en statistique multidimensionnelle et en fouille des données, que ce soit d'un point de vue méthodologique, théorique ou algorithmique. Les approches proposées sont illustrées par une grande variété d'applications. On peut trouver plus d'informations concernant cet ouvrage sur le site Web suivant :

<http://www.springer.com>

Concernant le contenu et le déroulement de la journée du 4 septembre, on peut trouver des informations à l'adresse suivante :

<http://www.modulad.fr/dac/Home.html>

Patrice Bertrand

Pour le comité d'organisation de *Workshop on Data Analysis and Classification in honor of Edwin DIDAY*

— — —

### Compte-rendu de la 4ème Rencontre

### Internationale sur l'Analyse Statistique Implicative A.S.I. 4, de Castellon de la Plana (du 18 au 21 Octobre 2007)

Cette Quatrième Rencontre fait suite aux Journées intitulées *La fouille dans les données par la Méthode Statistique Implicative* organisées par Marc Bailleul à l'Institut de Formation des Maîtres de Caen les 23-24 Juin 2000. Cet ASI 1 a été suivi par le Colloque (ASI 2) à l'Université PUC de São Paulo du 9 au 11 Juillet 2003, organisé par Saddo Ag Almouloud ; puis par ASI 3 organisé du 6 au 8 Octobre 2005 à Palerme par Filippo Spagnolo. Ainsi, suite à l'intérêt porté à ces Journées, aux développements théoriques continus de l'Analyse Statistique Implicative (ASI), à la multiplication des recherches utilisant la méthode d'analyse pour explorer des données, c'est la 4ème fois que se rassemblaient des chercheurs représentant de nombreuses nationalités<sup>1</sup> sur le seul thème de l'ASI. La Rencontre a été organisée et présidée :

- par Pilar Orus (Comité d'organisation),
- et Régis Gras (Comité scientifique et de programme).

Elle a réuni plus de 60 participants. Le Professeur Eino Shin Suzuki, de l'Université de Kyushu (Japon) était conférencier invité.

Rappelons que la fouille dans des données (encore appelée *Knowledge Discovery in Databases* ou encore *Data Mining* dans la littérature anglo-saxonne) part, en général, du croisement de sujets (ou objets) et de variables binaires (propriétés ou attributs) ou non. Son objectif majeur consiste à conjecturer des modèles basés sur des relations quantitatives ou qualitatives et des structures induites à partir des données. Différentes méthodes, comme l'Analyse Factorielle des Correspondances (A.F.C.), la Classification Ascendante Hiérarchique (C.A.H.), sont communément utilisées pour de telles fouilles de données. Parmi elles, l'Analyse Statistique Implicative (A.S.I.), née de problématiques de didactique des mathématiques, fondée et développée par Régis Gras et son équipe, vise l'extraction de connaissances, d'invariants, par exemple des conceptions, etc., bref de règles inductives non symétriques consistantes. Elle attribue une mesure à des propositions du type *quand a est choisi, on a tendance à choisir b*. Pour cela, elle quantifie la qualité de ces règles sur la base statistique d'un nombre significatif de contre-exemples où la règle n'est pas vérifiée et où certains déséquilibres cardinaux sont observés parmi les exemples et les contre-exemples à l'implication et à sa contraposée. Le logiciel, dénommé CHIC (Classification Hiérarchique Implicative et Cohésitive), développé au début des années 90 par Saddo Ag Almouloud et depuis, par Raphaël Couturier, permet :

- de traiter différents types de variables (binaires, modales, fréquentielles, intervalles, floues),
- de quantifier la consistance, la qualité des règles extraites, la significativité de classes de règles,
- de quantifier la typicalité et la contribution de sujets ou de catégories de sujets à certaines règles et classes de règles,

<sup>1</sup>Belgique, Brésil, Chili, Chypre, Costa-Rica, Cuba, Espagne, France, Grèce, Italie, Japon, Slovaquie et Suisse.

- de représenter par un graphe, pour un seuil de qualité choisi, des chemins ou des réseaux de règles et, par une hiérarchie, des méta-règles appelées aussi règles généralisées,
- de supprimer, d'ajouter, de conjointre des variables.
- de représenter par une hiérarchie ascendante les classes de similarité des variables.

Pendant ces quatre Journées, 28 communications, 2 conférences et 2 ateliers pratiques ont permis de rendre compte de travaux théoriques ou applicatifs et de résultats significatifs obtenus dans différents domaines par une fouille dans leurs propres données<sup>2</sup>. Ces travaux ont fait apparaître la place de l'ASI en bio-informatique, en psychologie sociale, en éducation, etc. de façon complémentaire et comparative à des méthodes factorielles ou taxonomiques. La conférence invitée présentée par Einoshin Suzuki a brossé un panorama général sur l'extraction de connaissances et, en particulier, sur les règles d'exception structurées. La conférence de Régis Gras a permis un rappel des concepts mathématiques fondamentaux de l'A.S.I., dont ses nouveaux développements (processus de Poisson, stabilité de l'indice d'implication, règle d'exception, par exemple), en insistant sur les problématiques qui y ont mené. Quelques travaux pratiques sur CHIC ont conduit à une meilleure prise en mains de ce logiciel et une compréhension approfondie de la théorie sous-jacente.

Un ouvrage de 410 pages, *Nouveaux Apports Théoriques à l'ASI et Applications*<sup>3</sup> regroupant les conférences et les communications et un fascicule utilitaire contenant leurs résumés et des informations pratiques ont pu être distribués à chacun des participants dès son arrivée. Au cours de l'année 2007, les membres du comité scientifique (30 chercheurs) avaient été mis à contribution pour lire, critiquer et homogénéiser les textes proposés, certains de ceux-ci n'ayant pas été retenus à la suite des échanges avec leurs auteurs. La mise en forme définitive et le tirage ont été effectués par Bruno Pinaud (Polytech Nantes) et par une équipe autour de Pilar Orus et Pablo Gregori à l'Université Jaume I de Castellon. Un site ouvert par F. Spagnolo ([http://math.unipa.it/~grim/asi/asi\\_index.htm](http://math.unipa.it/~grim/asi/asi_index.htm)) et celui spécifique de ASI 4 créé par P. Gregori (<http://www.asi4.uji.es/>) permettent d'accéder, non seulement à ces actes, mais également à ceux de ASI 1 (Caen 2000), ASI 2 (São Paulo 2003) et ASI 3 (Palerme 2005), ainsi qu'à des photos prises pendant les Journées.

Au cours de ces Rencontres, le choix d'une alternance équilibrée de communications théoriques ou appliquées et de travaux pratiques sur CHIC, est toujours souligné fa-

vorablement à la fois par les chercheurs chevronnés et par les jeunes chercheurs. Les débats, quelquefois très animés qui ont accompagné chacune des interventions ont été l'occasion de préciser et de spécifier certains points éclairant et guidant des recherches en cours ou à venir. Unaniment, les participants ont souhaité la poursuite de telles Rencontres. D'ores et déjà ASI 5 se profile. L'Université de Bratislava est la première à manifester l'intention de l'organiser. Mais d'autres options existent également.

L'organisation de ASI 4 par Pilar Orus, Pablo Gregori et leur équipe a été remarquable d'efficacité et de rigueur. Les activités scientifiques et la restauration quotidiennes se déroulaient à l'Université Jaume I. L'Hôtel Luz, au centre de la ville offrait un confort excellent. La convivialité, la bonne humeur ont régné pendant ces journées, toujours ensoleillées, ponctuées par un repas de gala très méditerranéen et une promenade au site enchanteur de Peniscola.

Rappelons, en les remerciant, les divers organismes ayant apporté leur soutien à ASI 4 :

- Escuela Superior de Tecnología y Ciencias Experimentales de la Universitat Jaume I
  - Departamento de Matemáticas de la Universitat Jaume I
  - Service de la Science et de la Technologie (Ambassade de France en Espagne)
  - Fundació Caixa Castelló-Bancaixa (Plan de Promoción de la Investigación 2007 de la Universitat Jaume I)
  - Diputació de Castelló
  - Generalitat Valenciana
  - Association pour la Recherche en Didactique des Mathématiques (ARDM)
  - Association Extraction et Gestion des Connaissances (EGC)
  - Laboratoire d'Informatique de Nantes (LINA)
  - Société PerformanSe SA, Carquefou- Nantes
- et le parrainage scientifique de :
- Sociedad de Estadística e Investigación Operativa (SEIO) de España,
  - Sociedad de Estadística e Investigación Operativa (SEIO) de España
  - Società Italiana di Statistica (SIS)
  - Société Française de Statistique (SFdS)
  - Société Francophone de Classification (SFC)
  - International Association for Statistical Education (IASE)

Régis Gras, Président du Comité Scientifique

<sup>2</sup>Les langues espagnole, portugaise, anglaise et française étaient tour à tour utilisées sans faire obstacle à la compréhension en raison de l'usage systématique de deux langues pour chaque intervention.

<sup>3</sup>Editeurs : Régis Gras, Pilar Orus, Bruno Pinaud, Pablo Gregori.