



Colloquim STATIM'09
« Statistiques pour le traitement de l'image »

22 et 23 janvier 2009
Université Paris 1 Panthéon-Sorbonne

Organisation : Charles Bouveyron & Vincent Vigneron



Contexte et objectifs

La statistique joue de nos jours un rôle croissant dans de nombreux domaines scientifiques, notamment en traitement de l'image. En effet, les progrès scientifiques réalisés ces dernières années ont par exemple permis d'augmenter sensiblement les capacités de mesure, en imagerie particulièrement, et il est à présent difficile pour un opérateur humain de traiter de façon exhaustive ces données dans un temps raisonnable. La statistique prend alors le relais sur l'humain pour analyser de façon automatique ces données dans le but d'aider les opérateurs à la prise de décision. De plus, le traitement de l'image pose de nouveaux problèmes théoriques en statistique. Par exemple, la classification à partir de données incertaines, le traitement de données corrélées ou la discrimination de populations déséquilibrées sont des problèmes particulièrement présents en analyse d'images.

Il est par conséquent nécessaire de réfléchir sur l'usage de la statistique en analyse d'image et de proposer des méthodes statistiques adaptées aux problèmes modernes. Outre l'importance de la performance des méthodes proposées, elles devront également apporter une meilleure compréhension des phénomènes observés.

Le colloquium « Statistiques pour le traitement de l'image » se propose de faciliter les contacts entre les statisticiens et la communauté des « traiteurs d'images » et de faire ainsi germer de nouvelles idées en organisant sur 2 jours 4 sessions thématiques.

Organisation et partenaires

Ce colloquium est principalement financé par le Conseil Scientifique de l'Université Paris 1 Panthéon-Sorbonne et est accueilli dans les locaux du laboratoire SAMOS-MATISSE du Centre d'Economie de la Sorbonne.

Dates et lieu

Le colloquium se tiendra les 22 et 23 janvier 2009 dans les locaux de l'Université Paris 1 Panthéon-Sorbonne au Centre Pierre Mendès France, 90 rue de Tolbiac, 75013 Paris (métro Olympiades). L'accès aux différentes salles du colloquium se fait par les ascenseurs rouges.

Les exposés seront donnés dans la salle des thèses au 22ème étage.

Les pauses-café auront lieu en salle 13 au 20ème étage.

Programme du jeudi 22 janvier 2009

Les exposés ont lieu dans la salle des thèses au 22ème étage

9h–12h30 : Session « Problèmes nouveaux en apprentissage statistique »

- 9h00 : J.-M. Poggi, *Forêts aléatoires : importance et sélection de variables*
- 10h00 : Pause-café en salle C-20-13 (20ème étage)
- 10h30 : J. Jacques, *Apprentissage adaptatif en classification et régression*
- 11h30 : C. Bouveyron, *Classification robuste et apprentissage faiblement supervisé*

14h–17h30 : Session « Imagerie bio-cellulaire »

- 14h00 : B. Chalmond, *Microscopie cellulaire 3D haute résolution*
- 15h00 : Pause-café en salle C-20-13 (20ème étage)
- 15h30 : S. Lelandais, *Analyse d'images bio-cellulaires : reconnaissance morphologique et comptage*
- 16h30 : O. Lezoray, *Diffusion sur graphes, construction de classificateurs parcimonieux : deux approches pour le traitement d'images microscopiques*

Programme du vendredi 23 janvier 2009

Les exposés ont lieu dans la salle des thèses au 22ème étage

9h–12h30 : Session « Reconnaissance de formes dans des images »

- 9h00 : J. Hérault, *Images et scènes : réduire la variabilité pour mieux catégoriser*
- 10h00 : Pause-café en salle C-20-13 (20ème étage)
- 10h30 : C. Schmid, *Apprentissage semi-supervisé pour la classification d'images*
- 11h30 : J. P. Tarel, *Approche semi-quadratique et applications à la perception de l'environnement routier*

14h–17h30 : Session « Apprentissage statistique en grande dimension »

- 14h00 : M. Verleysen, *Le compromis intrusion/extrusion pour la projection et la visualisation de données en grande dimension*
- 15h00 : Pause-café en salle C-20-13 (20ème étage)
- 15h30 : A. Mohammad-Djafari, *Lien entre la notion de Copula et la tomographie*
- 16h30 : S. Canu, *Apprentissage et méthodes à noyau en grande dimension*

Forêts aléatoires : importance et sélection de variables

Jean-Michel Poggi

Equipe de Probabilités, Statistique et Modélisation

Université Paris-Sud

Email : jean-michel.poggi@math.u-psud.fr

On s'intéresse à la méthode des forêts aléatoires d'un point de vue méthodologique. Introduite par Leo Breiman en 2001, elle est désormais largement utilisée tant en classification qu'en régression avec un succès spectaculaire. On vise tout d'abord à confirmer les résultats expérimentaux, connus mais épars, quant au choix des paramètres de la méthode, tant pour les problèmes dits « standards » que pour ceux dits de « grande dimension » (pour lesquels le nombre de variables est très grand vis à vis du nombre d'observations). Mais la contribution principale de cet article est d'étudier le comportement du score d'importance des variables basé sur les forêts aléatoires et d'examiner deux problèmes classiques de sélection de variables. Le premier est de dégager les variables importantes à des fins d'interprétation tandis que le second, plus restrictif, vise à se restreindre à un sous-ensemble suffisant pour la prédiction. La stratégie générale procède en deux étapes : le classement des variables basé sur les scores d'importance suivie d'une procédure d'introduction ascendante séquentielle des variables.

Il s'agit d'un travail en collaboration avec R. Genuer et C. Tuleau.

Références :

R. Genuer, J.-M. Poggi & C. Tuleau, *Random Forests : some methodological insights*, Rapport de Recherche INRIA, N° 12345, Nov. 2008, 32 p. <http://hal.inria.fr/inria-00340725/fr/>

Apprentissage adaptatif en classification et régression

Julien Jacques

Laboratoire Paul Painlevé, Université Lille 1

Email : *julien.jacques@polytech-lille.fr*

Nous proposons des techniques adaptatives en apprentissage statistique permettant de transférer l'information acquise sur une population de référence P vers une population cible P^* . Etant donné un échantillon d'individus issu d'une population P , décrits par des variables explicatives et regroupés en classes définies *a priori*, la classification supervisée consiste à construire une règle de classement de ces individus, qui sera utilisée pour classer de nouveaux individus dont la classe serait inconnue. Mais lorsque ces individus sont issus d'une population P^* différente de celle qui a servi à l'apprentissage de la règle de classement, cette dernière n'est plus utilisable. En établissant un lien entre P et P^* , il devient possible d'estimer la nouvelle règle de classement de façon parcimonieuse. Nous verrons comment ce lien peut être défini dans le cas de données continues puis de données binaires. Dans un cadre de régression, cette idée sera reprise afin de permettre d'obtenir une estimation peu coûteuse d'un modèle de régression sur une population cible, en utilisant l'information relatée par un modèle semblable mais établi sur la population de référence.

Références :

C. Bouveyron & J. Jacques, *Adaptive linear models in regression*, Preprint HAL n°00305987, Equipe SAMOS, Université Paris 1 Panthéon-Sorbonne, July, 2008.

Classification robuste et apprentissage faiblement supervisé

Charles Bouveyron

SAMOS-MATISSE, CES, UMR CNRS 8174

Université Paris 1 Panthéon-Sorbonne

Email : *charles.bouveyron@univ-paris1.fr*

En classification supervisée, la supervision humaine est requise pour associer des *labels* aux observations d'un jeu d'apprentissage qui sont ensuite utilisés pour construire un classifieur. Cependant, dans de nombreuses applications, la supervision humaine peut s'avérer être imprécise, difficile ou coûteuse. Dans de telles situations, les classifieurs supervisés peuvent être fortement destabilisés et produire des décisions instables. Pour pallier ce problème, nous proposons une méthode de classification robuste au bruit sur les *labels* qui se base sur un modèle de mélange. L'idée de la méthode proposée est de confronter une modélisation non supervisée des données avec l'information de supervision portée par les *labels* afin de détecter des inconsistences. La méthode est ensuite capable de construire un classifieur robuste prenant en compte les inconsistences détectées. Nous montrons également que cette approche peut être étendue à la classification faiblement supervisée dans laquelle la supervision est fournie par des informations faibles ou vagues.

Références :

C. Bouveyron & S. Girard, *Robust supervised classification with mixture models : Learning from data with uncertain labels*, Preprint HAL n°00325263, Equipe SAMOS, Université Paris 1 Panthéon-Sorbonne, September, 2008.

Microscopie cellulaire 3D haute résolution

Bernard Chalmond

CMLA, UMR CNRS 8536

Ecole Normale Supérieure de Cachan

Email : bernard.chalmond@cmla.ens-cachan.fr

We address the problem of volume reconstruction from a sequence of cross-sections in the case where the positions of the cross-sections are unknown. We propose an integrated statistical formulation of the problem leading to an energy minimisation algorithm, both for registration and reconstruction. 3D deconvolution using a PSF with varying orientation completes the algorithmic setup. This method has been developed in the context of micro-rotation confocal microscopy and experiments will illustrate the ability of this method to reconstruct efficiently non-adherent living cells.

This work is in collaboration with Y. Yu and A. Trowé.

Analyse d'images bio-cellulaires : reconnaissance morphologique et comptage

Sylvie Lelandais

Laboratoire IBISC - CNRS FRE 3190

Université d'Evry-Val-D'Essonne

Email : sleland@cemif.univ-evry.fr

Dans cet exposé nous présentons un travail qui est mené conjointement par deux équipes du laboratoire IBISC : l'équipe DYNAMIC, constituée de biologistes expérimentaux, et l'équipe TADIB, spécialisée en traitement de données et d'images. L'objectif final de ce travail est de proposer aux biologistes un logiciel d'analyse automatique d'images qui, à partir de vues prises sur un microscope apotome ZEISS, permette d'évaluer le pourcentage de cellules appartenant à des classes prédéterminées. En effet, l'équipe DYNAMIC étudie les occurrences de cellules morphologiquement différentes et potentiellement plus invasives. Pour cela ils travaillent sur la potentialité de migration de cellules cancéreuses mises en culture. A ce stade, les cellules saines ayant évolué vers l'état cancéreux peuvent opter pour deux types de migration cellulaire : la migration mésenchymateuse, migration lente caractérisée par des cellules de forme allongée, ou la migration amiboïde, migration rapide caractérisée par des cellules de forme ronde et « blebbante » qui est conjointe à un échappement cellulaire de la tumeur primaire avec comme pronostic une forte probabilité de création de métastase. La question qui se pose aux biologistes est de savoir pourquoi des cellules passent du mode de migration mésenchymateuse au mode de migration amiboïde. L'hypothèse faite par l'équipe DYNAMIC est que cette évolution du comportement est liée au micro-environnement cellulaire et en particulier à la présence de la molécule PAI-1 (Plasminogen Activator Inhibitor-1) qui « encouragerait » ce comportement « métastatique ». La validation d'une telle hypothèse permettrait d'envisager de modifier le micro-environnement d'une tumeur primaire pour éviter la multiplication des sites cancéreux à travers le développement de métastases. Le problème est que, pour valider cette hypothèse, de nombreuses observations sont nécessaires. Ces observations se font sur des cultures de cellules vivantes photographiées au microscope. Sur ces images, on va compter le nombre de cellules de chaque classe : non migrante (ronde et lisse), mésenchymateuse (allongée) et amiboïde (ronde et blebbante), et évaluer les pourcentages de chaque classe en fonction de la composition du micro-environnement et de la durée écoulée. Le travail de culture est déjà long et délicat. Rajouter à cela des opérations de comptage et de classification « manuelle » rendent le travail particulièrement fastidieux.

Après avoir exposé le mode opératoire permettant d'acquérir les images, nous nous attarderons sur les différentes difficultés liées à ces images : gradient de luminosité d'orientation variable, faible rapport signal à bruit, éclairage rasant provoquant une rupture des contours. Si chacun de ces problèmes peut être résolu, la présence de l'ensemble de ces difficultés nécessite la mise en place d'une chaîne complète de prétraitements permettant d'obtenir une image correcte des écarts-types, image servant de support au calcul des composantes connexes présentes dans l'image. Par ailleurs, un filtrage par une dif-

férence de gaussiennes appliqué sur l'image des écarts-type permet d'obtenir une image dite de « halo » mettant en évidence la position du centre des cellules et rendant possible l'opération de comptage. Cette approche par filtrage est comparée à une approche par transformée de Hough. Les résultats présentés illustrent les limites respectives de chaque méthode. Enfin, une segmentation par Ligne de Partage des Eaux (LPE) est opérée sur une carte des distances réalisée sur chaque composante connexe, les germes de la LPE étant les centres précédemment obtenus. Il est ainsi possible d'isoler un maximum de cellules sur lesquelles seront calculés cinq paramètres morphologiques utilisés dans l'étape de classification.

D'où l'intérêt du développement d'un logiciel automatique de traitement et analyse des images qui dégage les biologistes de cette tâche et leur permet de prétendre à une validation de leur hypothèse sur une analyse d'une population cellulaire de grande taille.

Ce travail est réalisé en collaboration avec T.Q. Syed, A. Cartier-Michaud, V. Vigneron, C. Charrière-Bertrand, C. Montagne, G. Barlovatz-Meimon, M. Malo.

Référence :

M. Malo, C. Charrière-Bertrand, E. Chettaoui, C. Fabre-Guillevin, F. Maquerlot, A. Lackmy, A. Vallée, F. Delaplace, and G. Barlovatz-Meimon. *The PAI-1 swing : Microenvironment and cancer cell migration*. Accepté aux « Comptes Rendus Biologie », 2006.

V. Vigneron, S. Lelandais, C. Charriere-Bertrand, M. Malo, A. Ugon, and G. Barlovatz-Meimon. *Pro or cons local vs. global imagery information for identifying cell migratory potential*. In 15th European Signal Processing Conference (EUSIPCO 2007), pages 443-448, Poznań, Poland, September 2007.

Diffusion sur graphes, construction de classificateurs parcimonieux : deux approches pour le traitement d'images microscopiques

Olivier Lezoray

GREYC, UMR CNRS 6072, Équipe Image
Université de Caen
Email : *olivier.lezoray@info.unicaen.fr*

Les pathologistes utilisent leur outil de travail, le microscope, depuis le 17^{ème} siècle et la plus grande partie de leur analyse est visuelle. C'est pourquoi l'imagerie pathologique est devenue un environnement d'imagerie médicale d'importance grandissante qui présente de nombreux défis. Un de ces défis réside dans la conception de systèmes d'aide au diagnostic par une analyse automatique ou semi-automatique. Nous présentons deux approches pour le traitement d'images microscopiques. La première se place dans le cadre d'une analyse entièrement automatisée et constitue un système de screening de lames microscopiques. La seconde se place dans le cadre d'une analyse semi-automatique où le pathologiste s'intéresse uniquement à des images choisies qu'il souhaite analyser interactivement.

Dans le cadre d'une analyse entièrement automatique, nous avons retenu le schéma classique de segmentation, caractérisation et classification. La segmentation consiste à extraire les cellules d'une image de microscopie, à les caractériser, puis à les classer selon leur malignité. Plusieurs de nos travaux précédents ont montré que des schémas de segmentation hybrides combinant classification de pixels et opérations morphologiques permettent d'obtenir des segmentations d'images microscopiques de très bonne qualité. Cependant, avec une analyse automatique, la masse de données générée est énorme et nous devons faire face à des données de grandes dimensions et en grand volume. Ceci s'accompagne inévitablement de redondance et de bruit et a deux conséquences : des temps d'apprentissage très long et des fonctions de décisions complexes. Afin de pallier ces problèmes, nous avons défini une méthode de sélection de modèle qui permet d'utiliser des SVM dans le cadre de la classification de pixels en microscopie. Ceci nous permet d'obtenir des classificateurs parcimonieux dont l'apprentissage est rapide, et dont les fonctions de décisions restent efficaces tout en étant peu complexes et donc rapides. La sélection de modèle que nous avons mis en œuvre permet la sélection des exemples pertinents, des attributs pertinents et des hyper-paramètres du classificateur. L'espace de recherche étant trop grand, nous utilisons une méta-heuristique (la recherche tabou) qui effectue la sélection de modèle selon un compromis entre attache aux données et complexité de la fonction de décision. Avec ces éléments, nous obtenons des classificateurs SVM parcimonieux (peu de vecteurs supports) que nous utilisons d'une part dans un schéma de segmentation morphologique, mais également pour la classification de cellules.

Dans le cadre d'une analyse semi-automatique, nous nous intéressons à une représentation par graphe de l'image à analyser et proposons un ensemble de méthodes basées graphe pour la segmentation d'images microscopiques. Nous montrons que traitement local et

non local peuvent être unifiés par la définition de dérivées discrètes sur graphes pondérés. Ces dérivées discrètes permettent de transcrire équations aux différences partielles continues en équations aux différences partielles sur graphes de topologies arbitraires. Nous nous plaçons dans un cadre variationnel et nous proposons un analogue discret de la régularisation continue qui opère sur des graphes de topologies arbitraires. La formulation proposée amène à une famille de méthodes simples et rapides basées sur le p -Laplacien pondéré et dont les paramètres sont le degré de régularité p , la structure du graphe et la fonction de similarité pour les poids du graphe. Cela nous permet de disposer d'un ensemble d'outils qui peuvent être combinés pour résoudre de nombreux problèmes en traitement d'images microscopiques (filtrage, segmentation, etc.). Dans ce cadre, nous montrons l'intérêt d'utiliser des représentation simplifiées du graphe grille d'une image par notamment les graphes d'adjacence de régions. Nous illustrerons les potentialités de l'approche sur des problèmes de segmentation interactive en histologie et cytologie.

Le système visuel : Réduire la variabilité pour mieux catégoriser Images et Scènes

Jeanny Hérault

Laboratoire GIPSA-lab
INPG Grenoble
Email : *jeanny.herault@inpg.fr*

Les premiers niveaux du système visuel des primates sont maintenant bien connus. Dans cet exposé, nous présentons leur architecture et leurs fonctions comme un modèle pour le traitement et l'analyse des images. Tout y apparaît comme bien adapté à la statistique des images pour en réduire la redondance et les variabilités.

La rétine, par des fonctions spécifiques, extrait les informations utiles contenues dans le signal spatio-temporel des images de notre monde visuel :

- Un filtrage spatio-temporel à variables non séparables qui compense le spectre en $\frac{1}{f}$ des images,
- Une compression localement adaptative qui réduit la variabilité des éclaircissements et des contrastes,
- Un codage des couleurs qui est particulièrement efficace,
- Un échantillonnage spatialement variant qui est bien adapté à l'interprétation de notre monde 3D.

Le cortex visuel primaire qui procède à une analyse locale de l'image rétinienne :

- Une analyse fréquentielle en spectre d'énergie qui s'affranchit des translations,
- Des filtres spatiaux à profil radial Log-Normal pour mieux échantillonner le spectre des images,
- Une distribution Log-polaire des fréquences centrales pour mieux s'affranchir des effets de taille et de rotation, et pour estimer la perspective monoculaire.

Ces différentes propriétés seront illustrées par des exemples d'application à la catégorisation des scènes et à l'estimation de la perspective locale dans les images naturelles.

Apprentissage semi-supervisé pour la classification d'images

Cordelia Schmid & Jakob Verbeek

Equipe LEAR
INRIA Rhône-Alpes
Email : jakob.verbeek@inrialpes.fr

In the first part we are interested in finding images of people on the web, and more specifically within large databases of captioned news images. It has recently been shown that visual analysis of the faces in images returned on a text-based query over captions can significantly improve search results. The underlying idea to improve the text-based results is that although this initial result is imperfect, it will render the queried person to be relatively frequent as compared to other people, so we can search for a large group of highly similar faces. The performance of such methods depends strongly on this assumption : for people whose face appears in less than about 40% of the initial text-based result, the performance may be very poor. I will present a method to improve search results by exploiting faces of other people that co-occur frequently with the queried person. We refer to this process as ‘query expansion’. In the face analysis we use the query expansion to provide a query-specific relevant set of ‘negative’ examples which should be separated from the potentially positive examples in the text-based result set. We apply this idea to a recently-proposed method which filters the initial result set using a Gaussian mixture model, and apply the same idea using a logistic discriminant model. We evaluate the methods on a database of captioned news stories from Yahoo!News. The results show that (i) query expansion improves both methods, (ii) that our discriminative models outperform the generative ones, and (iii) our best results surpass the state-of-the-art results by 10% precision on average. In the second part we are interested in Conditional Random Fields (CRFs), which are an effective tool for a variety of different data segmentation and labeling tasks including visual scene interpretation, which seeks to partition images into their constituent semantic-level regions and assign appropriate class labels to each region. For accurate labelling it is important to capture the global context of the image as well as local information. We introduce a CRF based scene labelling model that incorporates both local features and features aggregated over the whole image or large sections of it. Secondly, traditional CRF learning requires fully labelled datasets. Complete labellings are typically costly and troublesome to produce. We introduce an algorithm that allows CRF models to be learned from datasets where a substantial fraction of the nodes are unlabeled. It works by marginalizing out the unknown labels so that the log-likelihood of the known ones can be maximized by gradient ascent. Loopy Belief Propagation is used to approximate the marginals needed for the gradient and log-likelihood calculations and the Bethe free-energy approximation to the log-likelihood is monitored to control the step size. Our experimental results show that incorporating top-down aggregate features significantly improves the segmentations and that effective models can be learned from fragmentary labellings. The resulting methods give scene segmentation results comparable to the state-of-the-art on three different image databases.

Références :

T. Mensink & J. Verbeek, *Improving People Search Using Query Expansions : How Friends Help To Find People*, European Conference on Computer Vision, 2008.

J. Verbeek & B. Triggs, *Scene Segmentation with CRFs Learned from Partially Labeled Images*, Advances in Neural Information Processing Systems, 2007.

L'approche semi-quadratique pour l'analyse des images de l'environnement routier

Jean-Philippe Tarel

Division ESE

Laboratoire Central des Ponts et Chaussées

Email : *jean-philippe.tarel@lcpc.fr*

L'approche semi-quadratique a été développée initialement en analyse d'image pour traiter de problèmes de restauration d'image. Le champ d'application de cette approche a été ensuite étendu et cela a permis d'établir la connection avec l'estimation robuste. L'approche semi-quadratique recoupe aussi en partie l'approche EM (Expectation-Maximization). Une lecture de l'approche semi-quadratique fondée sur les outils d'optimisation numérique sous contraintes a aussi été proposée. Cette dernière reformulation permet à notre avis une facilité d'application plus grande, à la diversité des problèmes que l'on rencontre en analyse d'images. Nous illustrerons donc le propos sur différents problèmes qui se posent en analyse d'images de l'environnement routier : appariement et reconstruction selon un modèle linéairement paramétré pour estimer la forme 3D de la chaussée par stéréovision, régression robuste et simultanée de plusieurs courbes pour la détection des marquages, et enfin régression de région pour la détection de la chaussée.

Référence :

Le compromis intrusion/extrusion pour la projection et la visualisation de données en grande dimension

John A. Lee & Michel Verleysen

DICE – Machine Learning Group, Université catholique de Louvain

3 place du Levant, 1348 Louvain-la-Neuve, Belgique

Email : *michel.verleysen@uclouvain.be*

La réduction de dimension a pour ambition de produire des représentations en faible dimension d'ensembles de données en haute dimension. Un des objectifs principaux de la réduction de dimension est la visualisation de données (en dimension 2 ou 3). De nombreuses méthodes de réduction de dimension ont été proposées récemment. Une grande partie d'entre elles se basent sur l'optimisation d'un critère de respect de distances entre paires de points. Le critère peut être simple (souvent quadratique, éventuellement après transformation non-linéaire des données), permettant une optimisation de type algébrique. D'autres critères plus pertinents face aux objectifs de la réduction de dimension ont également été définis, permettant par exemple de sous-pondérer les paires de données éloignées, dont la distance est moins importante à prendre en considération pour un objectif de visualisation. Dans ce cas, l'optimisation du critère requiert en général des méthodes itératives basées sur le gradient. La variété des critères développés, ainsi que celle des méthodes d'optimisation utilisées, a donné lieu à un grand nombre de méthodes de réduction non-linéaire de dimension. La comparaison de ces méthodes n'est pas toujours aisée. Il est évident que chaque méthode est « meilleure » que d'autres, si le critère de comparaison utilisé est précisément celui optimisé par la méthode. Des critères de qualité plus objectifs ont alors été proposés dans la littérature. Souvent, ils se basent non plus sur une mesure de respect des distances entre paires de points lors de la projection, mais bien sur la notion de voisinage : les voisins d'un point sont-ils restés les mêmes lors de l'opération de projection ? Si la réponse est positive, la projection sera considérée comme bonne. Les critères proposés ne sont néanmoins pas non plus universels, ni exempts de subjectivité, et ce pour deux raisons. D'une part il faut définir quels voisinages sont utilisés pour la mesure de qualité, et comment éventuellement pondérer davantage les voisins proches que les voisins éloignés. D'autre part, sauf dans certains cas particuliers bien précis, il est impossible de préserver parfaitement tous les voisinages. Par exemple, si une sphère ou un cylindre doit être projeté sur un espace de dimension deux, il est impossible d'éviter en même temps les écrasements (des points loin sur la sphère seront projetés proches l'un de l'autre) et les déchirures (des voisins sur la sphère seront projetés loin l'un de l'autre). Cet exposé montrera d'une part que les critères de qualité des méthodes de projection non-linéaires doivent être envisagés sous la forme d'une paire de critères mesurant les intrusions et extrusions, respectivement liées aux écrasements et déchirements décrits ci-dessus. Il montrera d'autre part qu'en définissant proprement une matrice de « co-ranking » (intersection de voisinages dans les espaces avant et après projection), il est possible d'en déduire de façon simple et unifiée les différents critères de respect de voisinages actuellement utilisés dans la littérature.

Référence :

J. A. Lee & M. Verleysen, *Quality assessment of dimensionality reduction : Rank-based criteria*. Accepted for publication in Neurocomputing, in Press, available online 10 January 2009.

Lien entre la notion de Copula et la tomographie

Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes

UMR CNRS 08506, Supélec

Email : djafari@lss.supelec.fr

Un problème important en statistiques est de trouver une loi de probabilité jointe entre deux variables à partir de ses lois marginales. Dans le cas de deux variables, les densités marginales $f_1(x)$ et $f_2(y)$ sont liées à leurs distribution jointe $f(x, y)$ via des intégrations suivant les deux axes horizontal et vertical. Ce problème ressemble à un problème de reconstruction d'image en tomographie où on chercherait à reconstruire une image à partir de seulement deux projections. Ce problème est un problème mal-posé au sens qu'il admet une infinité de solutions, et justement, la notion de « Copula » qui permet de caractériser l'ensemble des lois jointes avec des marginales données, peut être utilisé en tomographie pour caractériser l'ensemble des images compatibles avec deux projections. On pourra alors rechercher une solution particulière dans cet ensemble. A l'inverse, nous pensons que les techniques de reconstruction d'image peuvent apporter une nouvelle voie dans la détermination d'une loi jointe à partir de ses lois marginales, un sujet qui est très important en inférence statistiques.

An important problem in statistics is determining a joint probability distribution from its marginals. In 2D case, the marginal probability density functions $f_1(x)$ and $f_2(y)$ are related to their joint distribution $f(x, y)$ via the horizontal and vertical line integrals. So, the problem of determining $f(x, y)$ from $f_1(x)$ and $f_2(y)$ is an ill-posed inverse problem. In statistics the notion of copula is exactly introduced to obtain a solution to this problem. Interestingly, this is also a problem encountered in X ray tomography image reconstruction where $f(x, y)$ is an image representing the distribution of the material density and $f_1(x)$ and $f_2(y)$ are the horizontal and vertical line integrals. In this talk we try to link the notion of copula to X ray Computed Tomography (CT) and to see if we can use the methods used in each domain to the other one.

Ceci est un travail commun avec Doriane Pougaza et Jean Francois Bercher.

Apprentissage et méthodes à noyau en grande dimension

Stéphane Canu

Laboratoire LITIS
EA 4108, INSA de Rouen
Email : *scanu@insa-rouen.fr*

Les machines à noyaux constituent une classe d'algorithmes permettant d'extraire de l'information à partir de données dans un cadre non-paramétrique et fonctionnel. L'intérêt suscité par ces méthodes tient d'abord aux excellentes performances qu'elles ont permis d'obtenir, notamment sur des problèmes de grandes dimensions. Cette bonne tenue à la charge est due à la parcimonie de la solution et à la (comparativement) faible complexité de son calcul. L'intérêt des machines à noyaux réside aussi dans le fait qu'elles s'insèrent dans un cadre flexible et rigoureux qui recèle un grand potentiel.

Cet exposé se propose de faire le point sur différentes facettes des machines à noyaux en général, des SVM en particulier, et leurs applications en traitement du signal et des images.

Restauration

Pour les repas de midi, nous tenons à votre disposition des tickets repas d'une valeur de 9,50 Euros pour la cantine Avenance (en face de l'université, rue de Tolbiac).

Nous vous proposons également ci-dessous une liste de restaurants proches du lieu du colloque :

- Hawaï, 87, avenue d'Ivry : une cantine Vietnamiennne (Pho = 6 euros),
- Lao Lanxang, 105, avenue d'Ivry : restaurant Laotien (<15 euros),
- Espace Aurore, 23 rue Terre au Curé : cuisine française (<15 euros).

Remerciements

Nous tenons à remercier les personnes et organismes suivants qui nous ont aidé dans la réalisation de cet évènement :

- l'université Paris 1 Panthéon-Sorbonne,
- le laboratoire SAMOS-MATISSE du CES,
- le service TICE de l'université Paris 1,
- Marie Cottrell, directrice du laboratoire SAMOS,
- Omar Aboura, Camille Brunet, Aurélien Hazan,
- et tous les membres du SAMOS pour leur soutien