

MACHINE LEARNING FOR BIG DATA

<http://machinelearningforbigdata.telecom-paristech.fr>

EN PARTENARIAT AVEC

criteo.

 **SAFRAN**
AEROSPACE · DEFENCE · SECURITY

PSA PEUGEOT CITROËN



BNP PARIBAS
La banque d'un monde qui change

AVEC LE SOUTIEN DE LA



CONTACTS PRESSE


Dominique Célier

01 45 81 75 17 • 06 87 11 95 90
dominique.celier@telecom-paristech.fr

Stéphane Menegaldo

01 45 81 70 95 • 07 82 75 17 30
stephane.menegaldo@telecom-paristech.fr

www.telecom-paristech.fr

 Suivez les débats et participez sur #MLBigData

SOMMAIRE

•	Editorial par Stéphan Cléménçon	3
•	Présentation de la Chaire d'enseignement et de recherche « Machine Learning for Big Data »	4
■	Machine Learning ?	4
■	Enjeux du <i>Big Data</i> et du <i>Machine Learning</i> pour les entreprises	4
■	Un marché de l'emploi en plein essor	4
•	La vision des partenaires	5
■	Criteo	5
■	PSA Peugeot Citroën	5
■	Safran	6
■	BNP Paribas	6
•	Les axes de recherche	7
■	Axe 1 : apprentissage par renforcement et optimisation/simulation stochastique	7
■	Axe 2 : graph-mining et analyse des réseaux sociaux	7
■	Axe 3 : ranking et détection d'anomalies	7
■	Axe 4 : cloud learning et algorithmes d'apprentissage distribués	7
■	Axe 5 : grande dimension - apprentissage et séries/flux de données temporelles	7
•	Activités de recherche	8
■	Interventions	8
■	Sélection de publications	9
•	Le Big Data, une compétence au cœur de Télécom ParisTech	10
■	Les Chaires d'enseignement et de recherche	10
■	L'innovation « Big Data » à Télécom ParisTech	10
•	Les formations au Big Data de Télécom ParisTech	11
■	Mastère Spécialisé « Big Data : Gestion et Analyse des Données Massives »	11
■	CES (Certificat d'Etudes Spécialisées) « Data Scientist »	11
■	Master « Data Sciences »	11
■	Formations courtes	11
■	Formations de sensibilisation au Big Data pour cadres dirigeants	11
■	MOOC « Fondamentaux pour le Big Data »	12
■	Séminaires Big Data du jeudi	12
•	Les partenaires et chercheurs de la Chaire	13
■	Criteo	13
■	Safran	13
■	PSA Peugeot Citroën	13
■	L'équipe académique	14
•	Biographies	15
•	Fiche technique de la Chaire « Machine Learning for Big Data »	16
•	A propos de	17
■	Télécom ParisTech	17
■	Criteo	17
■	PSA Peugeot Citroën	17
■	Safran	17
■	BNP Paribas	17

Editorial par Stéphan Cléménçon



L'accélération du développement et des usages des technologies de l'information et de la communication nous a indéniablement projeté dans une nouvelle ère numérique, celle du « Big Data », où la collecte, le stockage et l'accès à des données toujours plus massives offrent la perspective de progrès considérables dans de nombreux secteurs d'activité, poursuivant dans la voie tracée par les grands acteurs de l'internet.

L'omniprésence des capteurs, des systèmes embarqués, des réseaux sociaux, combinée au déploiement de briques technologiques permettant le traitement en ligne de masses de données peu structurées, laisse envisager la possibilité de suivre en temps réel des phénomènes complexes tels que l'évolution des comportements d'utilisateurs d'un portail sur le web, de superviser et optimiser le fonctionnement d'infrastructures sophistiquées, comme les réseaux de transports d'énergie, les flottes d'aéronefs ou de véhicules individuels et permet d'entrevoir la conception de nouveaux services, toujours plus personnalisés, en accord avec la granularité de l'information désormais disponible.

Le Big Data selon moi, c'est à la fois l'élaboration et la mise en œuvre de solutions technologiques nous permettant de plonger dans le bain de l'information numérique et la conviction partagée par de plus en plus d'acteurs, dans des domaines aussi variés que la Science, la Médecine, l'Industrie, les Services ou encore la Défense, que les sciences et techniques de l'information peuvent conduire à une amélioration sans précédent de nos connaissances et savoir-faire dans de nombreux champs de l'activité humaine.

Toutefois, certaines promesses ne pourront être tenues que si dans le même temps, des progrès substantiels sont réalisés dans le domaine du traitement mathématique des masses de données. C'est précisément le but du « Machine Learning » que d'élaborer des algorithmes permettant d'extraire et exploiter l'information enfouie dans des données, parfois bruitées et d'une dimension explosive, souvent à des fins prédictives ou d'optimisation. De ce point de vue, le Big Data, c'est aussi un certain nombre de challenges scientifiques pour les mathématiciens et informaticiens contribuant au développement de cette discipline en plein essor. Si l'on devait ne s'en tenir qu'aux « 3 V » du Big Data, on pourrait en effet facilement identifier les problèmes mathématiques relatifs à la contrainte du temps réel, requérant de revisiter les algorithmes afin de traiter l'information « à la volée » (Vélocité), aux difficultés du « passage à l'échelle » (Volume), et à la nécessité de traiter une information complexe portée par exemple par des séries temporelles, du texte ou encore des images ou des vidéos (Variété).

Perçu à tort comme une discipline visant à remplacer l'expertise d'un opérateur humain par des machines réalisant un traitement automatisé défini par des données, le Machine Learning a au contraire pour objectif de nous aider à exploiter des masses de données renfermant une information qu'il nous est absolument impossible d'embrasser sans un traitement mathématique adéquat, mis en œuvre au moyen de ressources informatiques spécifiques. Le Machine Learning est aujourd'hui à l'œuvre dans de nombreux domaines (internet, commerce, sécurité, gestion des risques, bioinformatique) et s'incarne avec succès dans des applications telles que la vidéosurveillance, la maintenance prédictive, la publicité en ligne ou les moteurs de recommandation par exemple.

On peut ainsi anticiper, sans grand risque de se tromper, que ce corpus de connaissances et techniques à l'interface des mathématiques et de l'informatique, en progrès constant depuis quelques décennies, sera encore à l'origine de nombreuses innovations à fort impact sociétal, économique ou scientifique pour peu que son potentiel soit compris par un public de plus en plus large, qu'il soit maîtrisé par un nombre croissant d'ingénieurs et de cadres techniques et qu'il se confronte aux enjeux de la société moderne. C'est le tout sens de la Chaire « Machine Learning for Big Data » de Télécom ParisTech.

Ses partenaires industriels l'ont parfaitement compris, puisqu'associés à une équipe académique dynamique, ils permettent en particulier d'enrichir la recherche par des cas d'usages susceptibles de constituer, demain, des grands champs applicatifs du Machine Learning, d'animer la formation au moyen d'une connaissance « métier », offrant une formidable visibilité à nos élèves quant aux débouchés des filières « Big Data – Data Science ».

Je me réjouis de vous retrouver à l'occasion de cette première journée de la Chaire « Machine Learning for Big Data », où j'aurai le plaisir d'évoquer les objectifs poursuivis dans les domaines de la Recherche et de la Formation ainsi que les premiers résultats obtenus, d'inciter tous les acteurs de la Chaire à interagir selon diverses modalités afin de mettre en place un cercle vertueux garantissant le succès de cette aventure. »

Stéphan Cléménçon est Professeur à Télécom-ParisTech, Institut Mines-Télécom, au sein du Département TSI et anime le groupe de recherche STA. Il effectue ses travaux de recherche en mathématiques appliquées au LTCI UMR Télécom ParisTech/CNRS No. 5141. Ses thématiques de recherche se situent principalement dans les domaines du Machine Learning, des probabilités et des statistiques. Il est responsable du Mastère Spécialisé « Big Data » à Télécom ParisTech ainsi que du Certificat d'Études Spécialisées « Data Scientist » et titulaire de la chaire industrielle « Machine Learning for Big Data ».

Présentation de la Chaire d'enseignement et de recherche « Machine Learning for Big Data »

Créée en septembre 2013 avec le soutien de la Fondation Télécom et financée à hauteur de près de 2 M€ par quatre entreprises privées **partenaires : Criteo, PSA Peugeot Citroën, Safran et BNP Paribas**, la Chaire « Machine Learning for Big Data » est portée par le mathématicien Stéphan Cléménçon, Enseignant-Chercheur, Professeur au sein du Département du Traitement du Signal et des Images à Télécom ParisTech.

Proposant cinq axes de recherche méthodologiques, enrichis par des applications industrielles concrètes, cette Chaire a pour objectif d'animer, de façon collaborative et en interaction avec ses partenaires, une **activité de recherche de pointe en Machine Learning**, ainsi que de proposer **des programmes de formation afférents**. Ces formations participent à la visibilité auprès des étudiants et des professionnels d'une expertise susceptible d'offrir des opportunités stratégiques dans la décennie à venir.

Stéphan Cléménçon est responsable du Mastère Spécialisé « Big Data : gestion et analyse des données massives » et du Certificat d'Etudes Spécialisées (CES) « Data Scientist » à Télécom ParisTech. Dans la même équipe, le professeur Florence d'Alché-Buc est responsable de la formation pour le Master « Data Sciences » co-habilité avec l'Ecole polytechnique. Toutes ces formations sont animées par des experts de la Chaire « Machine Learning for Big Data ».

Toutes les activités de la Chaire sont présentées sur son site Internet : <http://machinelearningforbigdata.telecom-paristech.fr>

■ Machine Learning ?

Le traitement statistique des masses de données convoque à la fois mathématiques appliquées et informatique, à travers une discipline en plein essor : le Machine Learning ou apprentissage statistique.

La variété des données aujourd'hui disponibles (nombres, images, textes, signaux), leur grande dimension et leur volumétrie rendent souvent inopérantes les méthodes statistiques traditionnelles reposant sur le prétraitement humain et un long travail de modélisation. Le Machine Learning vise donc à élaborer et étudier des algorithmes, à vocation prédictive le plus souvent, permettant à des machines d'apprendre automatiquement à partir des données et à effectuer des tâches de façon performante.

Ces algorithmes ne se substituent pas à une expertise humaine : ils ont vocation à réaliser des tâches qui ne pouvaient pas être effectuées jusqu'à présent du fait de la complexité et de la masse des informations disponibles. Il s'agit de révéler l'information cachée dans les données. De plus, grâce à la phénoménale puissance de calcul dont nous disposons maintenant, le Machine Learning peut répondre à des exigences de temps réel, en traitant presque immédiatement une somme d'informations colossale.

Les applications sont très variées. Dans le secteur aérien, la maintenance prédictive fait appel au Machine Learning qui analyse les milliers de rapports de vol de la totalité d'une flotte pour en déduire les priorités de maintenance. En e-commerce, une recommandation personnalisée pour chacun des dizaines de milliers d'utilisateurs d'un site se base sur l'analyse de centaines de milliers de produits. Dans la finance, des milliers d'informations sont scannées chaque seconde, ce qui permet de détecter une anomalie parmi des millions d'instructions dans les flux monétaires.

On passe ainsi véritablement d'une approche « top-down », où le savoir-faire, l'intuition guidaient les décisions, à une approche « bottom-up » où les décisions s'appuient sur les données réelles, rendues lisibles par l'approche Big Data.

■ Enjeux du Big Data et du Machine Learning pour les entreprises

Au-delà du buzz médiatique dont il fait l'objet, le Big Data est un sujet stratégique majeur, au cœur d'enjeux économiques et sociétaux considérables. Son impact est désormais perçu dans presque tous les secteurs de l'activité humaine : de la recherche scientifique à la médecine en passant, entre autres, par la finance, le bâtiment, l'e-commerce, la défense ou les transports.

Les avancées technologiques, l'omniprésence des capteurs (systèmes embarqués, objets connectés, internet ...) et l'explosion des réseaux sociaux s'accompagnent d'un véritable déluge de données, propulsant les sciences de l'information au centre du processus de valorisation des masses de données. Au-delà de la collecte et du stockage, l'enjeu est de pouvoir les analyser afin d'optimiser les décisions et mettre au point de nouvelles applications, toujours plus performantes : maintenance prédictive dans les réseaux et transports, ciblage commercial, biométrie...

■ Un marché de l'emploi en plein essor

Le Big Data appelle de nouvelles compétences, crée de nouveaux métiers (*Data scientist, Data analyst...*) et requiert des cursus pluridisciplinaires pouvant combiner mathématiques, informatique, droit et business. Ces cursus sont élaborés en collaboration étroite avec les secteurs de l'Industrie et des Services. Reconnue pour son expertise en ingénierie des technologies de l'information et de la communication, Télécom ParisTech a tout naturellement été le premier établissement supérieur français à dispenser des formations spécialisées en Big Data. Ses étudiants aux profils de haut niveau sont dès à présent très convoités sur le marché de l'emploi français et à l'international.

La vision des partenaires

■ Criteo

« L'excellence de la formation en mathématique en France est un vrai atout pour les entrepreneurs. Notre aventure Criteo nous a montré que les Etats-Unis facilitent la vie de l'entrepreneur, mais aussi que l'excellence de la formation, la loyauté et la fidélité des équipes de recherche en France sont des atouts formidables pour mener à bien de grands projets. »

Les recherches en Machine learning se font in vivo, il est donc important aussi que les professionnels transmettent leur expertise terrain aux data scientists et gestionnaires de données massives pour continuer de faire avancer les développements algorithmiques dans ce domaine. Criteo est donc fier de participer au développement des formations spécialisées proposées dans le cadre de la chaire dédiée au Machine Learning à Telecom ParisTech. C'est bien sûr un moyen pour nous de former de futurs talents, mais aussi par le partage de contribuer à développer des méthodologies et bonnes pratiques applicables au monde de l'entreprise comme au monde de la recherche. »

Romain Niccoli, Chief Technical Officer, Directeur Général Délégué et co-fondateur de Criteo

■ PSA Peugeot Citroën

« Tous les processus d'un groupe automobile sont supportés par des systèmes d'information et donc exploitent ou créent des données numériques en quantités extrêmement importantes. Par ailleurs, un des objets connectés le plus répandu dans un futur proche sera sans conteste le véhicule. Lui aussi est générateur de nombreuses données de fonctionnement. »

Forte de ce constat, PSA se retrouve donc avec un gisement de données souvent inexploité : produit, qualité, relation client, fonctionnement du véhicule...

La maturité de certaines technologies du Big Data nous a permis de mieux comprendre l'approche. A travers des pilotes de type « proof of concept », nous avons testé quelques cas d'usage pour différentes composantes du business: analyse des coûts de garantie de nos véhicules, activité bancaire, exploitation de données d'essais, analyse des informations issues de réseaux sociaux, services associés à des véhicules connectés, ...

Cela nous conforte dans l'idée que l'approche prédictive, permise par le Big Data, serait propice à une meilleure efficacité et un meilleur ciblage de nos actions. Ce serait par ailleurs, tout au moins dans un premier temps, un élément de différenciation par rapport à la concurrence.

Il nous est apparu évident qu'au-delà de l'enjeu technologique et de la maîtrise des outils, il fallait mener à bien une évolution nécessaire de la culture interne et des compétences. Et que cette transformation traverserait l'entreprise dans tous ses métiers. D'où notre participation à la Chaire Machine Learning for Big Data de Télécom ParisTech. »

Bernard Cohen, IT Benchmarking, Innovation, PSA Peugeot Citroën

■ Safran

« L'analyse de données massives est déjà présente dans certaines activités de Safran, qu'il s'agisse d'aéronautique avec le monitoring des moteurs et équipements en services dans les compagnies aériennes, ou de sécurité avec les applications de l'identification biométrique. Elle prend une place de plus en plus importante, elle est en effet une composante de la nouvelle révolution industrielle aux termes de laquelle les données, constituant un continuum du cycle de vie des produits depuis la conception jusqu'aux usages, en passant par la production, deviendront un atout pour la compétitivité des industriels du secteur de l'aéronautique, la défense et la sécurité.

Pour les clients de Safran la valeur d'usage comptera autant que la valeur intrinsèque des produits : il en résultera une transformation des modèles d'affaire basée sur des services de haute valeur ajoutée orientés vers la maximisation de la disponibilité opérationnelle des systèmes aériens ou de sécurité. Ainsi la capacité à mettre au point par apprentissage des algorithmes de surveillance permettant de discriminer les spécificités d'usage de chaque client, la fouille de données d'exploitation des flottes, ou la modélisation des comportements opérationnels et économiques propres à chaque compagnie aérienne seront déterminant pour la compétitivité des services proposés par Safran dans le domaine du transport aérien.

Vis-à-vis des enjeux des activités de Safran, deux verrous scientifiques principaux se présentent en analyse de données : d'une part la très haute performance requise pour les techniques d'apprentissage et la maîtrise théorique de celle-ci, dans des contextes aéronautique, défense et sécurité particulièrement exigeants en terme de justification de performance ; d'autre part la capacité à exploiter l'ensemble de l'expertise déjà acquise par Safran sur ses équipements et systèmes dans les approches d'apprentissage automatiques. Les activités de la Chaire Machine Learning contribueront à forger les compétences nécessaires pour aborder ces enjeux. »

Alain Coutrot, Directeur Adjoint R&T, Safran

■ BNP Paribas

« La révolution numérique, la technologie et la gestion intelligente de l'information sont au cœur du plan de développement du Groupe BNP Paribas et soutiennent une conviction forte : la transformation digitale sera structurante pour l'avenir de la banque. Le principal défi consiste à offrir aux clients l'expérience bancaire digitale avec tout le choix et la souplesse que cela implique, tout en assurant la sécurité de leurs opérations et de leurs données.

Avec plus de 500 millions de contacts digitaux par an avec ses clients, BNP Paribas est d'ores-et-déjà un grand acteur du digital. Le développement d'outils et de nouveaux algorithmes dans l'analyse de données massives et l'élaboration de modèles prédictifs offrent de nouvelles perspectives pour renforcer la relation avec nos clients et nos prospects, stimuler notre capacité d'innovation et proposer des services toujours plus adaptés comme par exemple :

- identifier, en temps réel, de comportements atypiques ou anormaux pour prévenir les fraudes à la carte bancaire ou aux virements.

- construire des modèles comportementaux basés sur des modèles statistiques pour aider les clients dans la gestion de leur budget, prodiguer des conseils pertinents et dignes de confiance.

C'est pour relever ces défis et accompagner, dans les domaines de la science des données et de l'apprentissage statistique, la montée en compétence de nos équipes que s'inscrit notre participation à la Chaire Machine Learning for Big Data de Télécom ParisTech. »

Pedro Cutillas, Head of SSC CRMRBIS, Coordination SSC SP MIB, BNP Paribas

Les axes de recherche

Le programme de recherche et d'enseignement de la Chaire porte sur les thèmes suivants :

■ **Axe 1 : apprentissage par renforcement et optimisation/simulation stochastique**

Dans de nombreuses situations, la « Machine » produisant automatiquement des décisions/prédictions interagit avec l'environnement d'où émanent les données (ex : phénomène de « market impact » en finance), l'apprentissage doit alors s'appuyer sur un nécessaire compromis entre « exploration » et « décision » et peut être formulé comme un problème d'optimisation stochastique dont la résolution requiert l'utilisation de méthodes de simulation avancées. La recherche d'une balance optimale entre « exploration » et « exploitation » peut parfois s'appuyer sur l'expertise humaine, permettant de définir les actions/décisions optimales et d'identifier plus aisément les paramètres des modèles régissant l'évolution du système. On parle alors d'apprentissage actif.

■ **Axe 2 : graph-mining et analyse des réseaux sociaux**

Dans les applications liées à l'Internet ou à l'exploration de réseaux sociaux par exemple, les données se représentent naturellement sous la forme d'un graphe, dont la dimension est souvent trop grande pour permettre une visualisation directe. L'extraction automatique des propriétés de réseaux de très grande dimension est l'un des axes de recherche considéré dans cette Chaire, ses applications vont de l'étude de la diffusion d'information au sein d'un réseau social à l'analyse du Web caché.

■ **Axe 3 : ranking et détection d'anomalies**

Dans les applications telles que le traitement des bases de données numériques massives pour le design de moteurs de recherche ou de recommandation, le but recherché n'est pas d'apprendre à prédire un label associé de façon probabiliste à une observation (comme c'est le cas en classification supervisée) mais d'apprendre à ranger les valeurs possibles pour le vecteur aléatoire d'observation dans un même ordre que celui induit par la probabilité a posteriori. Toutefois, le « ranking » ne se résume pas au problème (supervisé) de l'apprentissage d'un ordre mais peut également se référer à l'agrégation d'ordres ou de préférences, avec des applications dans le domaine des méta-moteurs de recherche ou en « database middleware ».

■ **Axe 4 : cloud learning et algorithmes d'apprentissage distribués**

Les réseaux (Internet, réseaux sociaux, etc.) ont conduit à une véritable explosion des bases de données. A titre d'exemple, fin 2013, le volume d'Internet représentait 1 yottaoctet (10^{24} octets) et en 2014 environ 50 milliards de pages sont indexées par Google, pour 3 milliards d'utilisateurs. De telles masses de données ne peuvent être stockées que de façon distribuée. Au-delà du problème de stockage, c'est l'analyse de ces « nuages de données » qui constitue aujourd'hui un véritable défi. C'est la raison pour laquelle nous mènerons un effort de recherche très important sur le thème de la décentralisation asynchrone d'algorithmes d'apprentissage supervisé et non supervisé « on-line » (ex : algorithmes de consensus, gossip), ainsi que sur les architectures logicielles permettant l'abstraction de cette décentralisation, élaboration d'algorithme d'apprentissage distribué et « on-line » sous des contraintes de capacité explicites (temps de calcul, mémoire, etc.).

■ **Axe 5 : grande dimension - apprentissage et séries/flux de données temporelles**

Les techniques de traitement du signal (ex : filtrage, analyse harmonique computationnelle, séparation de sources) sont encore largement méconnues dans le domaine du Machine Learning, une connaissance approfondie de ces dernières pourrait conduire au développement d'approches multivariées massives, multi-échelles/fréquentielles à des fins de prédiction/exploration. La rapidité avec laquelle certaines bases de données sont actualisées (finance, e-commerce, Internet...) parfois en temps réel, motive la recherche de méthodes d'apprentissage performantes dans un cadre séquentiel et adaptatif. La capacité à représenter efficacement des données complexes est souvent un aspect clef de l'apprentissage. Au-delà de la recherche d'une formulation mathématique des problèmes d'apprentissage statistique mentionnés, de solutions algorithmiques, d'un cadre de validité théorique pour ces dernières et de preuves expérimentales de concept, les travaux menés s'attacheront à la question du contrôle et de l'évaluation statistique de la performance des approches proposées.

Activités de recherche

Les chercheurs membres de la Chaire publient régulièrement des travaux rattachés aux cinq axes de recherche. Ils interviennent également dans de grands événements nationaux et internationaux. Ici n'est présenté qu'un aperçu de ces activités : l'ensemble des publications et des interventions peut être consulté sur le site Internet, rubrique Recherche : <http://machinelearningforbigdata.telecom-paristech.fr>

■ Interventions

IEEE 2014 - International Conference on Big Data

Washington, Etats-Unis, 27-30 octobre 2014 - <http://cci.drexel.edu/bigdata/bigdata2014>

La Conférence internationale IEEE 2014 sur le Big Data fournit une tribune de premier plan pour la diffusion des résultats les plus récents dans la recherche, le développement et les applications du Big Data. Des publications de haute qualité sont présentées sur tous les aspects du Big Data, en mettant l'accent sur les 5 V (Volume, Vitesse, Variété, Valeur et Vérité) : science et fondations du Big Data, infrastructure, gestion de données, search et mining, confidentialité des données et sécurité, applications. Stéphane Cléménçon présentera deux articles : "Scaling up M-estimation via sampling designs: the Horvitz-Thompson stochastic gradient descent" dans la session principale, le 30 octobre et "Multiresolution analysis of incomplete rankings with applications to prediction" dans le workshop Scalable Machine Learning: Theory and Applications le 27 octobre.

CIKM 2014 - International Conference on Information and Knowledge Management

Shanghai, China, 3-7 novembre 2014 - <http://cikm2014.fudan.edu.cn>

CIKM est une conférence de haut niveau dans les domaines de la recherche d'information, de la gestion des connaissances et des bases de données, réunissant des chercheurs et des praticiens des trois communautés. Son objectif est d'identifier les problèmes difficiles auxquels sera confronté le développement des futurs systèmes de connaissances et d'information, et à façonner les futures orientations de la recherche par le biais de la publication de résultats de recherche appliquée et théorique de haute qualité.

Luis Garriga présentera un article concernant l'axe 2 de la Chaire : graph mining et analyse des réseaux sociaux.

Conférence NIPS 2014 - Neural Information Processing Systems Foundation

Palais des Congrès de Montréal, Canada, 8-12 décembre 2014 - <https://nips.cc/Conferences/2014>

Joseph Salmon y a présenté un article co-écrit avec Jean Lafond, Éric moulines et Olga Klopp : "Probabilistic low-rank matrix completion on finite alphabets".

Eric Sibony a présenté un article dans le cadre du workshop Analysis of Rank Data : "Borda Count approximation of Kemeny's Rule and pairwise voting inconsistencies" qui se rattache à l'axe 3 de la Chaire : ranking et détection d'anomalies.

Colloque Technion France 2014 : Making Sense of Big Data

Maison de la Chimie, Paris, lundi 15 décembre 2014 de 8h30 à 18h30

Le Technion est l'Institution technologique Israélienne de référence au niveau international dans les domaines scientifique et technologique. Elle a permis la réalisation de nombreux succès dans tous les domaines d'excellence liés au Big Data. Le prochain colloque scientifique annuel de l'Association Technion France aura lieu le lundi 15 décembre 2014.

Stéphane Cléménçon est intervenu dans la session N°3 : Data Science : Recherche, éducation & formation aux Big Data.

Conférence COLT 2015

Université Pierre et Marie Curie, Paris, du 3 au 6 juillet

La conférence entend la théorie de l'apprentissage au sens large, par exemple : design et analyse d'algorithmes d'apprentissage, apprentissage non supervisé ou semi-supervisé, deep learning, théories bayésiennes, prise de décision et incertitude, traitement du langage naturel...

Daniel A. Spielman, Professeur à l'Université de Yale et spécialiste des matrices de graphes Laplaciennes, était invité à COLT par la Chaire Machine Learning for Big Data. Il a présenté un article avec Rasmus Kyng, Anup Rao et Sushant Sachdeva : Algorithms for Lipschitz Learning on Graphs.

Les chercheurs de la Chaire Machine Learning for Big Data ont par ailleurs présenté deux articles :

- Learning the dependence structure of rare events: a non-asymptotic study par Nicolas Goix, Anne Sabourin et Stéphane Cléménçon.
- Low Rank Matrix Completion with Exponential Family Noise par Jean Lafond.

■ Sélection de publications

La liste complète des publications est disponible sur :

<http://machinelearningforbigdata.telecom-paristech.fr/fr/pages/publications-scientifiques>

Axe 1 : apprentissage par renforcement et optimisation/simulation stochastique

Online Influence Maximization - S. Lei, S. Maniu, L. Mo, R. Cheng, and P. Senellart

Scalable, Generic, and Adaptive Systems for Focused Crawling - G. Gouriten, S. Maniu et P. Senellart

On the complexity of A/B testing - E. Kaufmann, O. Cappé and A. Garivier

Axe 2 : graph-mining et analyse des réseaux sociaux

Fast Rule Mining in Ontological Knowledge Bases with AMIE+ - Luis Galárraga, Christina Teflioudi, Fabian Suchanek, Katja Hose

Knowledge Bases for Web Content Analytics - Johannes Hoffart, Nicoleta Preda, Fabian M. Suchanek, Gerhard Weikum

Discovering Meta-Paths in Large Heterogeneous Information Networks - C. Meng, R. Cheng, S. Maniu, P. Senellart et W. Zhang

Axe 3 : ranking et détection d'anomalies

MRA-based Statistical Learning from Incomplete Rankings - Eric Sibony, Stéphan Cléménçon, Jérémie Jakubowicz

On Anomaly Ranking and Excess-Mass Curves - Nicolas Goix, Anne Sabourin, Stéphan Cléménçon

Anomaly Ranking as Supervised Bipartite Ranking - S. Cléménçon, S. Robbiano (Télécom ParisTech).

Axe 4 : cloud learning et algorithmes d'apprentissage distribués

A Distributed Frank-Wolfe Algorithm for Communication-Efficient Sparse Learning - Auteurs: A. Bellet, Y. Liang, A. Bagheri Garakani, M.-F. Balcan and F. Sha

On-Line Learning Gossip Algorithm in Multi-Agent Systems with Local Decision Rules - S. Cléménçon, P. Bianchi, J. Jakubowicz & G. Morral Adel

Axe 5 : grande dimension - apprentissage et séries/flux de données temporelles

Operator-valued kernel-based vector autoregressive models for network inference - Néhémy Lim, Florence d'Alché-Buc, Cédric Auliac, George Michailidis

Mind the duality gap: safer rules for the Lasso - Olivier Fercoq, Alexandre Gramfort, Joseph Salmon

Probabilistic low-rank matrix completion on finite alphabets - J. Lafond , O. Klopp, É. Moulines , J. Salmon

Le Big Data, une compétence au cœur de Télécom ParisTech

Première grande école française d'ingénieurs dans le domaine des Technologies de l'Information, Télécom ParisTech forme ses diplômés à innover et entreprendre dans un monde désormais numérique. Choisie par près de 1 500 étudiants chaque année dont 55% d'internationaux, l'école permet aujourd'hui à ses diplômés d'intégrer absolument tous les secteurs d'activité.

Avec des enseignements et une recherche d'excellence couvrant l'ensemble des technologies de l'information et des usages, Télécom ParisTech a mis au point un écosystème d'innovation unique, basé sur une forte interaction entre sa formation, son centre de recherche et ses deux incubateurs d'entreprises. Ses liens étroits avec l'Industrie en font un témoin privilégié de l'émergence du phénomène Big Data et de son impact technologique dans le domaine de la formation et de la recherche.

L'un des points forts de Télécom ParisTech réside depuis toujours dans ses compétences et son expertise quant au traitement de l'information structurée (signaux, réseaux, images, vidéos, données textuelles). L'ubiquité du Machine Learning dans ses activités de recherche et de développement combinée à une capacité d'adapter les méthodes d'analyse à la nature des données à traiter et aux systèmes d'information déployés, lui permet de contribuer significativement à l'état de l'art.

Le Big Data a été choisi comme l'un des six grands axes stratégiques de Télécom ParisTech.

En savoir plus sur www.telecom-paristech.fr/bigdata

■ Les Chaires d'enseignement et de recherche

Télécom ParisTech, en partenariat avec des entreprises, s'investi dans trois Chaires dans le domaine du Big Data au sein de son groupe, l'Institut Mines-Télécom.

Machine Learning for Big Data, créée en 2013 et pilotée par le professeur Stéphane Cléménçon au sein d'une équipe de réputation internationale dans ce domaine à l'interface des mathématiques et de l'informatique. Quatre entreprises prestigieuses en sont les partenaires : Criteo, PSA Peugeot Citroën, Safran, ainsi que BNP Paribas. Toutes quatre ont été motivées par l'ambition des travaux de l'équipe de Télécom ParisTech dans le domaine en pleine ébullition du Big Data.

Big Data and Market Insights, créée en 2014 en partenariat avec Télécom École de Management, portée par Talel Abdesslem, spécialiste des bases de données et de la fouille de données du Web et responsable de l'équipe Interaction, Cognition et Complexité du département Informatique et Réseaux. La Chaire, créée avec le soutien de la Fondation Télécom, est co-financée par Deloitte, le Groupe Rocher, SNCF et le Groupe BPCE.

En savoir plus : www.telecom-paristech.fr/recherche/chaieres/chaire-big-data-market-insights.html

Valeurs et politiques des informations personnelles, créée en 2013, en partenariat avec le Groupe Imprimerie Nationale, BNP Paribas, Orange, LVMH, Dassault Systèmes et Deveryware, avec la collaboration de la CNIL (Commission nationale de l'informatique et des libertés) est coordonnée par Claire Levallois-Barth, docteure en droit. Cette chaire traite des aspects juridiques, techniques, économiques et philosophiques qui concernent la collecte, l'utilisation et le partage des informations personnelles.

En savoir plus : www.informations-personnelles.org

■ L'innovation « Big Data » à Télécom ParisTech

ParisTech Entrepreneurs est l'incubateur de Télécom ParisTech. En 15 ans, il a accueilli plus de 300 startups innovantes du numérique. Beaucoup d'entre elles utilisent les technologies du Big Data. Parmi celles-ci : Adomik, acteur innovant de la publicité sur le Web ; FocusMatic, la plateforme d'analyse de Big (Social) Data ; iDMog, la plateforme de gestion intégrée des données de production de pétrole et de gaz et Safety Line, la solution innovante de gestion de la sécurité aérienne.

En savoir plus sur ParisTech Entrepreneurs : <http://entrepreneurs.telecom-paristech.fr>

Les formations au Big Data de Télécom ParisTech

■ Mastère Spécialisé « Big Data : Gestion et Analyse des Données Massives »

Le Mastère Spécialisé est un cursus professionnalisant qui s'adresse à un public de diplômés (ingénieurs et Masters), en poursuite d'études ou en reconversion, pour une promotion de 30 étudiants environ. Sur une durée de 16 mois, cette formation propose plus de 700 heures de cours, travaux pratiques et séminaires d'octobre à juillet et se conclut par un stage de 4 à 6 mois en entreprise et la soutenance d'une thèse professionnelle.

Son programme, en forte adéquation avec les besoins des entreprises en traitement de données massives, s'appuie sur de nombreuses études de cas, des retours d'expérience et un socle de connaissances à la fois techniques (informatique, mathématiques appliquées, droit) et business conduisant à l'exercice opérationnel du métier de « Data Scientist ». Les projets « Fil Rouge », proposés et tutorés par des professionnels, couvrent les différentes facettes du Big Data (acquisition, stockage et analyse des données, visualisation, mise en SaaS, aspects légaux, business model) et sont réalisés en groupe sur trois trimestres, offrant en particulier la possibilité de travailler sur des sujets Big Data en vraie grandeur.

En savoir plus : www.telecom-paristech.fr/formation-continue/masteres-specialises/big-data.html

■ CES (Certificat d'Etudes Spécialisées) « Data Scientist »

Le CES est destiné aux professionnels en situation d'emploi aspirant à accroître leurs compétences dans le domaine de la Science des Données (stockage, représentation, analyse statistique, visualisation). Très opérationnelle, programmée sur une période de 10 mois, la formation permet la maîtrise des techniques de gestion et d'analyse des Big Data et des principaux algorithmes du Machine Learning.

La formation se répartit en 12 sessions de deux jours sur le site de Télécom ParisTech, chaque session s'organisant en séances de cours, travaux dirigés et travaux pratiques et se ponctuant par le témoignage d'un professionnel. Une plate-forme d'e-learning permet aux futurs Data Scientists d'approfondir les concepts et méthodes présentés lors des sessions et de les valider.

En savoir plus : www.telecom-evolution.fr/fr/formations-certifiantes/ces-data-scientist

■ Master « Data Sciences »

Co-habilité avec l'Ecole polytechnique, en partenariat avec l'Université de Paris-Sud et l'ENSAE ParisTech.

Première formation de niveau Master en Big Data associant plusieurs acteurs académiques majeurs, ce Master Recherche vise un public d'étudiants désireux d'approfondir les mathématiques appliquées dans le champ de la Science des Données (statistique, optimisation, Machine Learning), en relation avec les briques technologiques permettant le « passage à l'échelle » (calcul et optimisation distribuée, Framework Hadoop...). Il comprend des cours avec un projet en groupe étalé sur 24 semaines et se conclut par un stage de 5 mois. Particulièrement équilibré entre mathématiques et informatique, ce Master est aussi très appliqué via des études de cas proposées par les entreprises.

Ce Master fait partie depuis la rentrée 2015 de la mention « Mathématiques et Applications » de l'offre des masters de l'Université Paris-Saclay et est destiné à devenir une référence dans le domaine des Data Sciences au niveau mondial.

En savoir plus : <http://universite-paris-saclay.fr/fr/formation/master/m2-data-sciences>

■ Formations courtes

Télécom Evolution propose 8 formations courtes (1 ou 2 jours) focalisées sur des compétences précises : infrastructures et architectures distribuées, Data Science et Machine Learning, sécurité, visualisation, Web sémantique, extraction de données, le langage R... Un module accessible à tout public initie sur deux jours les non spécialistes aux enjeux économiques, juridiques et techniques du Big Data.

En savoir plus : www.telecom-evolution.fr/fr/domaines/big-data

■ Formations de sensibilisation au Big Data pour cadres dirigeants

AgroParisTech, l'ENSAE ParisTech, Télécom ParisTech et HEC Paris se sont associées pour sensibiliser les cadres dirigeants aux enjeux et aux opportunités du big data. La mise en place de projets big data porteurs de valeur nécessite l'implication des cadres dirigeants qui doivent repenser business models et chaînes de valeur. Les 4 écoles ont donc choisi de proposer une série de formations spécifiquement destinées aux décideurs, quelle que soit la taille de leur entreprise. Trois programmes ciblés permettent de répondre à leurs interrogations spécifiques.

En savoir plus : www.telecom-paristech.fr/bigdata-dirigeants

■ MOOC « Fondamentaux pour le Big Data »

De nombreuses personnes souhaitent accéder aux métiers du big data en raison des importants besoins de recrutement dans ce domaine. Elles recherchent une solution flexible, accessible et compatible avec une activité professionnelle leur permettant d'acquérir le niveau prérequis en informatique et statistiques. Le MOOC « Fondamentaux pour le big data » répond à ces besoins et prépare efficacement au suivi des formations dans le domaine du big data.

Le Big Data offre de nouvelles opportunités d'emplois au sein des entreprises et des administrations et de nombreuses formations préparant à ces opportunités se sont mises en place. Suivre ces formations nécessite des connaissances de base en statistiques et en informatique que ce MOOC propose d'acquérir. Les compétences visées constituent un préalable indispensable dans les domaines de l'analyse, de l'algèbre, des probabilités, des statistiques, de la programmation Python et des bases de données.

Le MOOC est organisé en 6 semaines, et se compose de sept parties se terminant par un quiz validant les acquis des différentes sessions vidéos. Un quiz final faisant suite à un projet valide l'ensemble du MOOC.

En savoir plus : www.telecom-evolution.fr/fr/moocs/mooc-fondamentaux-pour-le-big-data

■ Séminaires Big Data du jeudi

En complément des formations du Mastère Spécialisé et du Master Recherche, des séminaires hebdomadaires sont proposés le jeudi à Télécom ParisTech avec des professionnels, acteurs du Big Data. L'objectif des séminaires est de présenter les cas, de les analyser et d'en discuter.

Ainsi, pendant la première année de formation (2013-2014), des spécialistes de Thales, Nextpreformance, Terradata, Syncsort, McKinsey et Coach & Boost sont venus échanger avec les étudiants. La saison 2014-2015 s'est ouverte avec une intervention de Gilles Babinet, « Digital champion » pour l'Europe.

Voir le programme des séminaires : <http://machinelearningforbigdata.telecom-paristech.fr/fr/tag/seminaire>

Les partenaires et chercheurs de la Chaire

■ Criteo



Nicolas Le Roux
Scientific Program Manager



Romain Niccoli
Chief Technical Officer, Directeur
Général Délégué et co-fondateur

■ Safran



Alain Coutrot
Directeur R&D Safran



Daniel Duclos
Responsable du Pôle Technologies du
Signal et de l'Information, Direction R&T



Vincent Bouatou
Directeur adjoint Recherche &
Technologie, Morpho (Safran)



Jérôme Lacaille
Expert Algorithmes Snecma (Safran)



Stéphane Gentric
Research Entity Manager, Senior
Expert, Research and Technology Unit

■ PSA Peugeot Citroën



Bernard Cohen
DP/DSIN/SUSI Benchmarking,
Innovation



Sylvain Allano
Directeur scientifique et technologies
futures



Jean-Pierre Dumoulin
Maître-Expert Big Data



Claudia Constant
Directeur de l'Université PSA Peugeot
Citroën



Jean-Marc Mousset
PSA Corporate University Partnerships,
Coordination & Projects

■ L'équipe académique

Biographies détaillées sur le site : <http://machinelearningforbigdata.telecom-paristech.fr/fr/pages/lequipe-academique>



Stéphane Cléménçon

Professeur à Télécom ParisTech et porteur de la Chaire Machine Learning for Big Data



Eric Moulines

Professeur à Télécom ParisTech



Olivier Cappé

Directeur de Recherche au CNRS



Pierre Senellart

Professeur à Télécom-ParisTech



Joseph Salmon

Maître de conférence à Télécom ParisTech



Fabian M. Suchanek

Maître de conférence à Télécom ParisTech



Slim Essid

Enseignant-chercheur à Télécom ParisTech



Alexandre Gramfort

Enseignant-chercheur à Télécom ParisTech



Ons Jelassi

Enseignante à la formation continue de Télécom ParisTech



Chloé Clavel

Enseignant-chercheur à Télécom ParisTech



Pascal Bianchi

Enseignant-chercheur à Télécom ParisTech



Laurence Likforman

Professeur Associé (HDR) à Télécom ParisTech



Florence d'Alché-Buc

Enseignant-chercheur à Télécom ParisTech



Gersende Fort

Chercheur CNRS à Télécom ParisTech



François Roueff

Enseignant-chercheur à Télécom ParisTech



Anne Sabourin

Enseignant-chercheur à Télécom ParisTech



Mauro Sozio

Enseignant-chercheur à Télécom ParisTech



Olivier Fercoq

Maître de conférences à Télécom ParisTech

Biographies



Yves Poilane

Directeur de Télécom ParisTech

Yves Poilane est président du jury du Concours Commun Mines-Ponts et de la Banque de Notes du même nom depuis 2007. Il assure plusieurs missions auprès des entreprises et dans l'enseignement supérieur : il est actuellement vice-président de l'association Pasc@line (visant à promouvoir les formations et métiers du numérique), président de la commission internationale de la Conférence des Grandes Ecoles (CGE) et président de la commission Enseignement de ParisTech (Institut des sciences et technologies de Paris).

Yves Poilane est également l'un des administrateurs de la Fondation de Coopération Scientifique Paris Saclay. Ingénieur général des Mines, diplômé de l'Ecole polytechnique en 1982 et de Télécom ParisTech en 1984, il a assuré, avant 2007, plusieurs postes opérationnels et de direction à France Télécom, Télécom Bretagne et chez Orange.



Stéphan Cléménçon

Professeur à Télécom-ParisTech, au sein du Département TSI.

Il anime le groupe de recherche STA. Il effectue ses travaux de recherche en mathématiques appliquées au LTCI UMR Télécom ParisTech/CNRS No. 5141. Ses thématiques de recherche se situent principalement dans les domaines du machine-learning, des probabilités et des statistiques. Il est responsable du Mastère Spécialisé « Big Data » à Télécom ParisTech ainsi que du Certificat d'Études Spécialisées « Data Scientist » et titulaire de la chaire industrielle « Machine-Learning for Big Data »



Nicolas Le Roux

Scientific Program Manager chez Criteo depuis octobre 2012.

Nicolas Le Roux a soutenu sa thèse (PhD) en Machine Learning de l'Université de Montréal en 2008. Il a ensuite travaillé sur les réseaux de neurones, les modèles statistiques à grande échelle et sur l'optimisation, à Microsoft Research Cambridge et à l'INRIA. Depuis 2012, il est Scientific Program Manager à Criteo, en charge du pilotage des projets scientifiques relatifs à la prédiction du clic et à la recommandation de produits.



Jean-Pierre Dumoulin

Chief Technical Officer Responsable Infrastructure, Stratégie et Télécoms PSA Peugeot Citroën depuis août 2010.

Jean-Pierre Dumoulin a auparavant occupé les fonctions de Responsable de Datacenter, Responsable du département « Outils CAO & Calcul » et Responsable de la stratégie technique et expertise opérationnelle chez PSA Peugeot Citroën.

Jean-Pierre Dumoulin a pour principales missions la définition et la mise en œuvre de la stratégie technique pour la DSI PSA Peugeot-Citroën, le pilotage des infrastructures et des services transversaux. Expert du Big Data, ce Centralien (Marseille, promotion 1987) et Supélec (1988) a initié dès 2011 le lancement de « Business 3.0 », projet d'entreprise visant à faire des infrastructures un vecteur de compétitivité, notamment grâce au Big Data.

Jean-Pierre Dumoulin est également impliqué dans deux associations professionnelles destinées à faire avancer l'IT en France : Vice-Président exécutif du CRIP (Club des Responsables d'Infrastructure et de Production) et Vice-Président exécutif / membre fondateur de CDO Alliance (association centrée sur les métiers du numérique).



Alain Coutrot

Directeur adjoint Recherche et Technologie de Safran depuis mai 2005

Polytechnique (1970) – ENSTA (1975) - Ingénieur en Chef de l'Armement

Après un début de carrière à la DGA, Alain Coutrot rejoint le Groupe en 1983 où il dirige le développement des boosters d'Ariane 5 à la Société Européenne de Propulsion puis chez Europropulsion. Il revient en 1992 chez SEP-Villaroche Division petite propulsion et équipements spatiaux et en 1998, il intègre la Direction de la R&T de Snecma et devient Directeur de la R&T du Groupe Snecma en 2004.

Vice-Président de la Fondation de Recherche pour l'Aéronautique et l'Espace (mai 2005). Chevalier dans l'Ordre de la Légion d'Honneur.

Fiche technique de la Chaire « Machine Learning for Big Data »

Date de création :	Septembre 2013
Porteur de la chaire :	Télécom ParisTech (Institut Mines-Télécom)
Responsable académique :	Stéphane Cléménçon – Professeur à Télécom ParisTech
Partenaire académique :	Fondation Télécom
Partenaires entreprises :	Safran PSA Peugeot Citroën Criteo BNP Paribas
Axes de recherche :	Apprentissage par renforcement et optimisation/simulation stochastique Graph-mining et analyse des réseaux sociaux Ranking et détection d'anomalies Cloud learning et algorithmes d'apprentissage distribués Grande dimension - apprentissage et séries/flux de données temporelles
Equipe de recherche :	Stéphane Cléménçon – Professeur à Télécom ParisTech Eric Moulines – Professeur à Télécom ParisTech Olivier Cappé – Directeur de recherche au CNRS Pierre Senellart – Professeur à Télécom ParisTech Joseph Salmon – Maître de conférence à Télécom ParisTech Fabian Suchanek – Maître de Conférences à Télécom ParisTech Slim Essid – Enseignant-chercheur à Télécom ParisTech Alexandre Gramfort - Enseignant-chercheur à Télécom ParisTech Ousama J. Jelassi – Enseignante à Télécom ParisTech Chloé Clavel – Enseignant-chercheur à Télécom ParisTech Mauro Sozio – Maître de Conférences à Télécom ParisTech Pascal Bianchi – Enseignant-chercheur à Télécom ParisTech Florence d'Alché-Buc – Enseignant-chercheur à Télécom ParisTech Gersende Fort – Chercheur CNRS à Télécom ParisTech Laurence Likforman – Professeur associé à Télécom ParisTech François Roueff – Enseignant-chercheur à Télécom ParisTech Anne Sabourin – Enseignant-chercheur à Télécom ParisTech Olivier Fercoq – Maître de conférences à Télécom ParisTech
Site de la Chaire :	http://machinelearningforbigdata.telecom-paristech.fr

A propos de...

■ Télécom ParisTech

Télécom ParisTech forme à innover et entreprendre dans un monde numérique. Ses cursus diplôment ingénieurs, docteurs et professionnels tout au long de la vie et attirent 55 % d'internationaux. Toutes les disciplines des sciences et technologies de l'information et de la communication sont couvertes par ses enseignements et sa recherche. Celle-ci présente une expertise originale sur 6 axes stratégiques - Big Data, Très Grands Réseaux et Système, Confiance numérique, Interactions Réel-Virtuel, Modélisation, approche interdisciplinaire de l'Innovation - lui permettant de faire face aux défis majeurs du monde numérique.

Télécom ParisTech se positionne comme le futur Collège de l'innovation par le numérique de l'Université Paris-Saclay, dont l'ambition est de devenir l'un des premiers pôles d'innovation mondiaux.

www.telecom-paristech.fr

■ Criteo

Criteo diffuse au niveau mondial des publicités à la performance personnalisées. La société mesure le retour post-clic sur les ventes, rendant ainsi le ROI transparent et facile à calculer. Criteo emploie plus de 1 000 personnes aux États-Unis, en Europe et en Asie, et collabore avec plus de 6 000 annonceurs dans une cinquantaine de pays. Opérationnel sur toutes les grandes plates-formes d'enchères en temps réel, Criteo entretient également des relations directes avec plus de 7 000 éditeurs à travers le monde.

www.criteo.com

■ PSA Peugeot Citroën

Fort de trois marques de renommée mondiale, Peugeot, Citroën et DS, le Groupe a vendu 2,8 millions de véhicules dans le monde en 2013, dont 42% hors d'Europe. Deuxième constructeur automobile européen, il a réalisé un chiffre d'affaires de 54 milliards d'euros en 2013. Il s'affirme comme leader européen en termes d'émissions de CO₂, avec une moyenne de 115,9 grammes de CO₂/km en 2013. PSA Peugeot Citroën est présent dans 160 pays. Ses activités s'étendent aussi au financement (Banque PSA Finance) et à l'équipement automobile (Faurecia).

www.psa-peugeot-citroen.com

■ Safran

Safran est un groupe international de haute technologie, équipementier de premier rang dans les domaines de l'Aéronautique et de l'Espace (propulsion, équipements), de la Défense et de la Sécurité. Implanté sur tous les continents, le Groupe emploie 66 300 personnes pour un chiffre d'affaires de 14,7 milliards d'euros en 2013*. Composé de nombreuses sociétés, Safran occupe, seul ou en partenariat, des positions de premier plan mondial ou européen sur ses marchés. Pour répondre à l'évolution des marchés, le Groupe s'engage dans des programmes de recherche et développement qui ont représenté en 2013 des dépenses de 1,8 milliard d'euros. Safran est une société cotée sur Euronext Paris et fait partie de l'indice CAC 40.

* Le chiffre d'affaires 2013 retraité des impacts IFRS11 s'établit à 14,4 milliards d'euros

www.safran-group.com

■ BNP Paribas

BNP Paribas a une présence dans 75 pays avec plus de 185 000 collaborateurs, dont près de 145 000 en Europe. Le Groupe détient des positions clés dans ses deux grands domaines d'activité : Retail Banking & Services (comprenant Domestic Markets et International Financial Services) et Corporate & Institutional Banking. En Europe, le Groupe a quatre marchés domestiques (la Belgique, la France, l'Italie et le Luxembourg) et BNP Paribas Personal Finance est numéro un du crédit aux particuliers. BNP Paribas développe également son modèle intégré de banque de détail dans les pays du bassin méditerranéen, en Turquie, en Europe de l'Est et a un réseau important dans l'Ouest des États-Unis. Dans ses activités Corporate & Institutional Banking et International Financial Services, BNP Paribas bénéficie d'un leadership en Europe, d'une forte présence dans les Amériques, ainsi que d'un dispositif solide et en forte croissance en Asie-Pacifique.

www.bnpparibas.com