

Appel stages de master

Institut de Convergence DATAIA

Hiver 2019

1 Introduction

L'Institut DATAIA est l'institut de convergence français spécialisé en sciences des données, intelligence artificielle et société, ou comment les technologies big data et l'intelligence artificielle se croisent avec les sciences humaines et sociales pour une intelligence artificielle au service de l'humain. Sélectionné par l'Agence Nationale de la Recherche (ANR) dans le cadre des programmes d'Investissements d'Avenir", l'Institut DATAIA a pour vocation de regrouper et de structurer des expertises pluridisciplinaires de grande ampleur et de forte visibilité pour mieux répondre à des enjeux majeurs des sciences des données, de l'intelligence artificielle et de leurs applications grâce au décloisonnement entre les mathématiques, les sciences informatiques et les sciences juridiques, économiques et sociales. Il mobilise à son lancement plus de 130 chercheurs et enseignants-chercheurs de quatorze établissements (Université Paris-Saclay, Inria, CEA, CentraleSupélec, CNRS, Ecole polytechnique, ENSAE, HEC, IFP-EN, IMT, INRA, Université d'Evry-Val d'Essonne, Université Paris-Sud, Université Versailles - Saint-Quentin-en-Yvelines), ainsi qu'un programme d'affiliation industrielle permettant la collaboration étroite avec les entreprises. L'Institut DATAIA soutient la recherche en Science des données au sens large —incluant l'informatique, les mathématiques et les sciences humaines et sociales— à travers un appel annuel dédié aux stages.

Pré-requis

- Toute proposition devra s'appuyer sur la collaboration d'au moins deux personnes de deux établissements partenaires de l'institut DATAIA (voir la liste en annexe 2) n'appartenant pas au même laboratoire ni au même établissement hébergeur.
- La proposition doit rentrer dans le cadre des problématiques de DATAIA décrites en annexe.
- Toute communication (publication, présentation) devra mentionner le financement DATAIA :

This research was supported by DATAIA convergence institute as part of the "Programme d'Investissement d'Avenir", (ANR-17-CONV-0003) operated by [Partner] XXX

2 Soumission

Soumission	1 ^{er} Mars 2019
Décision	31 Mars 2019
Contact	viviane.hoang@inria.fr
Adresse de soumission	submission-dataia@inria.fr .

Proposition (2 pages): Les partenaires décrivent l'objectif du stage, son lien avec DATAIA, et listent les collaborations antérieures des partenaires (le cas échéant). Il n'est pas indispensable d'avoir un candidat pour la soumission du projet.

Soutien: Si le projet est accepté, DATAIA financera l'étudiant pour 4 à 6 mois au niveau de gratification habituel (environ 550 euros par mois).

Engagements: Le rapport de stage devra être adressé à DataIA à l'issue du stage. Une présentation brève pourra être demandée lors des journées scientifiques annuelles de DATAIA

Information requisite: La proposition devra nécessairement indiquer l'institut en charge de l'inscription de l'étudiant (nom, adresse, informations bancaires, contact administratif).

Annexe 1: Axes de recherche de DATAIA

Le programme scientifique de chaque proposition devra aborder explicitement au moins un des quatre axes de recherche esquissés ci-dessous, qui sont les quatre défis interdisciplinaires de DATAIA.

DES DONNÉES AUX CONNAISSANCES, DES DONNÉES À LA DÉCISION

La disponibilité croissante de données massives repousse les frontières techniques dans de nombreux champs. D'une part, la nature hétérogène, semi-structurée, incomplète ou incertaine des données remet en cause les modèles statistiques usuels ainsi que les algorithmes dédiés à la décision. D'autre part, la gestion des données soulève de nouvelles contraintes d'opérabilité telles que la sécurité, l'intégrité et la traçabilité.

En outre, produire de la connaissance nécessite de construire des modèles qui fournissent des décisions explicables, statistiquement valides et calculables. L'acceptation des résultats requiert aussi que la confidentialité et la loyauté soient renforcées. En parallèle, de nouveaux développements en optimisation doivent permettre d'améliorer les procédures d'estimation. Défis :

- Données hétérogènes, complexes, incomplètes, semi-structurées et/ou incertaines
- Données massives : algorithmique et structuration des données
- Apprentissage en ligne, méthodologie pour les données massives, méthodes efficaces
- Amélioration du stockage, calcul et estimation pour la science des données
- Modélisation des interactions entre agents (humains ou artificiels) par théorie des jeux.
- Représentation et algorithmes multi-échelle et multimodaux
- Analyse théorique de méthodes heuristiques (théorie de la complexité, géométrie de l'information, théorie des chaînes de Markov)
- Coévolution Humain-Machine dans les systèmes autonomes : agents conversationnels, voitures, robots sociaux

APPRENTISSAGE ET INTELLIGENCE ARTIFICIELLE

Récemment, la recherche en apprentissage profond a réalisé des avancées spectaculaires en vision par ordinateur et en traitement du langage naturel. Au-delà de l'arrivée des données massives, de l'accroissement de la puissance de calcul et des efforts de conception, les causes de ces progrès, encore mal connues, soulèvent au moins trois questions. Quelle théorie de l'apprentissage permettra

d'analyser les architectures profondes ? Comment gérer la compositionnalité de ces architectures et leur capacité à appréhender des objets plus complexes ? Comment ouvrir la boîte noire pour mettre à jour les représentations apprises ?

Défis

- Apprentissage automatique innovant et IA : sens commun, adaptabilité, généralisation
- Apprentissage profond et apprentissage adversarial
- Apprentissage automatique et hyper-optimisation
- Optimisation pour l'apprentissage, e.g. améliorations des méthodes de gradient stochastiques, optimisation Bayésienne), optimisation combinatoire
- Lien apprentissage-modélisation, intégration d' priori dans l'apprentissage
- Reproductibilité et apprentissage robuste
- Inférence statistique et validation
- Compositionnalité des architectures profondes.

TRANSPARENCE, IA RESPONSABLE ET ÉTHIQUE

La confiance numérique s'instaure à partir de l'implantation de méthodologies éthiquement responsables à travers la transparence et la redevabilité des systèmes algorithmiques ; de la régulation de la collecte, de l'usage et du traitement de données personnelles ; du renforcement de la régulation au travers de procédures numériques appropriées.

La confidentialité par conception est une forme de régulation qui inclut la protection des données personnelles dans toutes les étapes de la collecte et du traitement. Le traçage des outils appliqués aux données doit également être développé de manière à faciliter l'explication du modèle pour les experts comme pour les utilisateurs rendant les systèmes algorithmiques auditables. Les principes de confidentialité, bien que faciles à formuler, nécessitent de modifier les infrastructures de stockage et de traitement, avec d'importants impacts législatifs, sociologiques et économiques. Les techniques de transparence des systèmes algorithmiques seront développés en focalisant sur : l'équité, la loyauté et la non-discrimination et la redevabilité-par-construction.

Défis

- Responsabilité et explicabilité par conception
- Transparence et équité par conception

- Audit des systèmes algorithmiques : non-discrimination , loyauté, biais techniques, neutralité, équité
- Mesure de la confiance et de l'appropriation du numérique
- Progressive user- analyse de donnée progressive (monitoring interactif des systèmes de décisions)
- Responsabilité du traitement de l'information et de la prise de décision : contrôle d'usage des données et vérification des faits
- Découverte causale, traçabilité des inférences à partir des données sources, interprétabilité des architectures profondes

PROTECTION, RÉGULATION ET ÉCONOMIE DE LA DONNÉE

Les entreprises impliquées dans l'économie de la donnée ont continuellement besoin de repenser leur structuration : elles doivent adopter une organisation orientée-projet avec des changements rapides dans l'allocation de ressources. L'économie de la donnée soulève aussi des problèmes de concentration et de monopole.

Un petit nombre d'entreprises (GAFAM) détiennent la plupart des données. Cette concentration de marché peut conduire à une compétition déloyale dont l'innovation dans les petites et moyennes entreprises risque de pâtir. Les citoyens attendent des gouvernements qu'ils interviennent dans l'économie digitale pour éviter trop de concentration et de monopole. Les gouvernements doivent empêcher la fuite des informations pour préserver la souveraineté des états et le respect des réglementations.

Défis

- respect de la vie privée par conception GDPR
- Apprentissage respectueux de la vie privée(*differential privacy*)
- Développement de méthodologies éthiquement responsables, et de technologies pour réguler la collection, l'usage et le traitement des données personnelles, et l'exploitation des connaissances tirées de ces données
- Sécurité informatique des chaînes de traitement de données
- Sécurité/crypto : block-chain et tiers de confiance

Les projets pourront aborder des sujets plus spécifiques à des domaines d'application. Une liste de ces domaines est donnée ci-dessous à titre indicatif. Toutefois, il est demandé que les projets abordent des questions de science des données d'intérêt général, c'est-à-dire non restreintes spécifiquement à un domaine d'application.

- Energie : optimisation de la gestion (production et distribution), et régulation.
- Sante, nutrition personnalisée et bien-être.
- Urbanisation Mobilité (véhicule connecté et autonome, villes intelligentes...)
- Analyse pour la finance les assurances
- Internet des personnes et des objets
- e-Sciences.

Annexe 2: Liste des partenaires de DATAIA

- Agro ParisTech
- CentraleSupélec
- CEA
- CNRS
- Ecole polytechnique
- ENS Paris-Saclay
- ENSAE
- ENSTA ParisTech
- HEC Paris
- IFP-EN
- Inria
- Institut Mines-Telecom (Télécom ParisTech, Télécom Sud Paris, Institut Mines-Télécom Business School)
- Inra
- Université d'Evry Val d'Essonne.
- Université Paris-Saclay
- Université Paris-Sud
- Université Versailles St-Quentin-en-Yvelines