

APPEL A PROJETS 2018

Projets sélectionnés

PEPER : Prédiction de la Prosomation d'énergie renouvelable par apprentissage

*Hossam AFIFI, Télécom SudParis; Jordi BADOSA, Ecole polytechnique;
Florence OSSART, CENTRALESUPELEC GEEPS*



Le monde de l'énergie électrique fait face à des changements structurels important : les usages de l'électricité sont en constante augmentation et les enjeux climatiques imposent une augmentation de la part des énergies renouvelables dans la production (solaire et éolien). Le réseau électrique centralisé évolue vers une structure décentralisée, formée de sous-ensembles qui associent production, stockage et consommation à l'échelle locale et coopèrent pour couvrir l'ensemble des besoins à plus grande échelle. La prosomation (adaptation du comportement de consommation en fonction de l'énergie produite et disponible) est alors un point-clé pour assurer l'équilibre du réseau. La gestion énergétique efficace et coopérative d'un tel système s'appuie sur la prévision du comportement des différents acteurs du réseau (producteurs et consommateurs), l'échange de données entre eux (coopération), à différentes échelles de temps et d'espace. L'objectif du projet PEPER est de rassembler des données pertinentes sur les différents acteurs du réseau, et d'exploiter les techniques d'apprentissage pour développer des algorithmes de prévision de la production et de la consommation de chaque acteur, puis de coopération entre eux.

Valorisation des Données pour la Recherche d'Emploi (VADORE)

*Bruno CREPON, ENSAE ; Michele SEBAG, CNRS ; Marco CUTURI, ENSAE ;
Christophe GAILLAC, ENSAE ; Philippe CAILLOU, LRI*



Le contexte du projet est celui du chômage en France. Le chômage est un phénomène multi-causes, dépendant notamment des facteurs limitant l'offre et la demande de travail. Ce projet se focalise sur le chômage frictionnel, lié aux imperfections informationnelles, dus aux coûts de collecte, de traitement et de diffusion de l'information, ainsi qu'à l'asymétrie d'information entre offreurs et demandeurs d'emploi (DE), et aux limitations cognitives des individus. Ces imperfections sont une des raisons pour laquelle certains emplois restent inoccupés alors même qu'une demande d'emploi importante est observée dans les mêmes secteurs d'emploi. L'idée centrale du projet est de mobiliser l'ensemble des informations disponibles pour améliorer l'appariement des DE et des emplois vacants. Le projet s'appuie sur la mobilisation de l'ensemble considérable des informations sur les DE et les entreprises, parmi lesquelles certaines (les données textuelles notamment) sont encore inexploitées. Ces informations seront exploitées pour élaborer deux fonctionnalités, de nature technique et d'inspiration économique différentes, les évaluer et les comparer rigoureusement.

Bad Nudge - Bad Robot ? : Nudge et Éthique dans l'interaction verbale homme-machine

Laurence DEVILLERS, LIMSI-CNRS ; Serge PAJAK, RITM, EA 7360, Université Paris-Sud



Les objets connectés et spécifiquement les agents conversationnels comme Google Home, apportent une nouvelle dimension dans l'interaction, à savoir la parole, et pourraient devenir un moyen d'influence des individus. Ils ne sont pour l'instant ni régulés, ni évalués et très opaques. Les 'nudges', issues des sciences comportementales, consistent à inciter les individus à changer de comportement sans les contraindre et en utilisant leurs biais cognitifs. Souvent utilisés à bon escient pour la santé, les nudges sont des outils efficaces de politique publique. Mais l'utilisation d'interface vocale personnalisée pourrait amplifier les phénomènes de manipulation. Bad Nudge - Bad Robot ? est un projet interdisciplinaire pour répondre à cinq verrous de recherche sur le nudge dans l'interaction homme-machine : la définition des nudges et des bad nudges ; le développement d'outils d'évaluation des nudges ; la mise en œuvre de stratégies de nudging avec des robots montrant l'impact des nudges ; le développement d'un agent évaluateur et une étude longitudinale sur un objet connecté vocal existant (ex: Google Home) ; l'élaboration à l'aune de nos disciplines de préconisations éthiques et juridiques de politique publique pour la régulation de ces objets dans la société. Une plus grande transparence des systèmes permettrait l'amélioration de la confiance et de la compréhension des systèmes par les utilisateurs, notamment pour des personnes vulnérables comme les enfants et les personnes âgées.

RGPD et Cloud Personnel : de l'Empowerment à la REsponsabilité (GDP-ERE)

*Célia ZOLYNSKI, UVSQ ; Mélanie CLEMENT-FONTAINE, UVSQ ; Nicolas ANCIAUX, INRIA ;
Philippe PUCHERAL, USVQ/INRIA ; Guillaume SCERRI, UVSQ/Inria*

Alors que se profile un monde bouleversé par l'intelligence artificielle et l'exploitation des données personnelles, la place des individus et la maîtrise de leurs données se sont imposées comme des questions centrales dans le nouveau règlement européen sur la protection des données (RGPD) et la loi pour une République numérique. Le projet GDP-ERE poursuit un double objectif : analyser l'impact des architectures de cloud personnel sur les enjeux de responsabilité et confronter cette analyse aux règles édictées par le RGPD ; et proposer des évolutions législatives et technologiques permettant de mieux capturer le partage de responsabilité nécessaire entre les différentes parties et apportant à chacune les outils appropriés pour les endosser. La portabilité consacre un droit de l'individu à la récupération de ses données personnelles, et ouvre des perspectives d'empowerment et de développement de nouveaux usages tels que le personal big data et le big personal data, réalisés sous le contrôle de l'individu. Le cadre juridique se limite pour l'instant à prescrire ce droit à la portabilité tout en reconnaissant qu'il s'accompagne de nouvelles formes de responsabilités, sans toutefois en préciser l'articulation entre individus, fournisseurs de plateformes et fournisseurs de services et sans prendre en compte la variété des solutions techniques de cloud personnels.

L'objectif du projet GDP-ERE est d'analyser ce double mouvement, juridique et technique, afin d'établir plus précisément les responsabilités inhérentes à l'empowerment, en conformité avec les notions juridiques existantes de responsable de traitement, de sous-traitants et de tiers, et les exemptions du RGPD, ainsi qu'à envisager la préconisation de plateformes offrant un niveau de responsabilité graduel aux individus, de manière adaptée à la technologie

MissingBigData : missing data in the big data era

Julie JOSSE, CMAP/CNRS ; Gaël VAROQUAUX, INRIA



Le « big data », souvent observationnel et composé, plutôt qu'expérimental et homogène, pose des défis de données manquantes. Nous proposons d'utiliser des modèles plus puissants qui peuvent bénéficier des grands échantillons de données, en particulier des auto-encodeurs, pour imputer des valeurs manquantes. Pour éviter de biaiser les conclusions, nous étudierons l'imputation multiple et les conditions sur la dépendance dans les données. Notre projet vise à diminuer les facteurs de risque en matière de santé avec la prédiction de meilleurs résultats et l'identification des facteurs de risque de résultats indésirables. Nous recherchons une solution opérationnelle, de la méthodologie à la mise en œuvre, qui intègre la diversité et le volume des données. Nous nous éloignons également des études classiques en considérant plusieurs types de données manquantes. Ce qui sera une première, mais qui semble réalisable compte tenu des résultats de Mohan et Pearl (2018).

Smart Lawyer: Rating Legal Services in the Courtroom

David RESTREPO AMARILES, HEC Paris ; Michalis VAZIRGIANNIS, École polytechnique



Le 11 mai 2017, la Cour de cassation a reconnu l'importance des services en ligne et des plateformes d'évaluation comparative des avocats et des cabinets d'avocats, notamment par des classements et des notations, pour la protection des consommateurs de services juridiques. Toutefois, la Cour a également affirmé que de tels services doivent assurer un certain niveau de qualité : "[il] leur appartient [...], dans leurs activités propres, de délivrer au consommateur une information loyale, claire et transparente". Malheureusement, à ce jour, les fournisseurs de ces services s'appuient principalement sur des preuves anecdotiques. Le manque d'informations fiables sur la qualité des services juridiques fournis par les avocats dans les salles d'audience est un phénomène inquiétant et généralisé dans toutes les juridictions de l'Union européenne, mais aussi aux États-Unis et au Canada. Ce projet vise à combler cette lacune en combinant l'expertise juridique et la recherche en science des données. Il cherche à développer des outils de mesure significatifs et fiables de la performance juridique qui peuvent améliorer l'accès à la justice et la qualité des services juridiques, tout en aidant les cabinets d'avocats à évaluer la performance des avocats et la qualité des juridictions.

HistorIA : Grandes bases de données historiques. Fouille de données, exploration et explicabilité

Christophe PRIEUR, Telecom ParisTech ; Jean-Daniel FEKETE, Inria



Depuis le développement des méthodes big data et leur arrivée dans les sciences sociales, plusieurs initiatives très ambitieuses ont vu le jour se fixant pour objectif de changer la manière de faire de la recherche en histoire. Pourtant, le déploiement de ces nouvelles approches se heurte à de nombreuses réticences des historiens, qui, confrontés à la difficulté du dialogue interdisciplinaire, sont souvent sceptiques sur les fins mêmes d'une collaboration dans laquelle ils craignent, parfois à raison, d'être dépossédés d'un matériau qu'ils ont le sentiment de ne plus maîtriser lorsqu'il est transformé pour être intégré dans des bases de données. Tout autant ces procédures de transformation que celles d'analyse suscitent des doutes méthodologiques voire épistémologiques profonds, d'autant plus que les outils mis en œuvre sont souvent novateurs et n'ont donc pas pu bénéficier de nombreux retours d'expérience. Dans ce projet réunissant des chercheurs en histoire, en sciences sociales computationnelles et en visualisation d'information, nous souhaitons développer des grandes bases de données historiques en y appliquant des méthodes de fouille de données, notamment autour de l'analyse des réseaux de relations, tout en mettant en œuvre une approche itérative du processus d'exploration, fondée sur l'appropriation par les utilisateurs des procédures et outils mobilisés ainsi que des résultats des analyses. Pour cela, l'accent sera mis sur l'explicabilité des algorithmes et sur l'analyse progressive des données progressive et l'interaction humain-machine.

StreamOps : Plateforme Open Source pour la Recherche et l'Intégration d'Algorithmes pour l'Analyse de Flux de Séries Temporelles Massives

Cédric GOUY-PAILLER, CEA; Karine ZEITOUNI, Université de Versailles-St-Quentin-en-Yvelines ;



Ce projet a pour objectif de concevoir et mettre à disposition de la communauté une plateforme de recherche et d'intégration pour la récupération, la manipulation et l'analyse de données en flux, avec un focus initial sur les séries de mesures issues de capteurs, ayant les caractéristiques suivantes :

- S'appuyant sur des outils logiciels performants (débit d'entrée, robustesse, tolérance aux pannes, distribution multi-nœud). On s'appuiera notamment sur Kafka pour le passage de données entre les différents éléments modulaires. Un serveur de données optimisé pour l'ingestion et le traitement de séries temporelles sera couplé à la plateforme afin de répondre à certains cas d'usage nécessitant le stockage d'une partie du flux pour réutilisation ultérieure ou réapprentissage complet.
- Offrant un environnement modulaire pour le concepteur d'algorithmes aussi bien que pour l'utilisateur, avec un cœur de gestion/manipulation des flux, et des briques de traitement intégrables dans différents langages (R, python, java).
- Prenant en compte le prétraitement et la représentation de séries de mesures en flux au sein du serveur de données
- S'appuyant sur des briques scientifiques performantes de l'état de l'art
- Code cœur open-source
- Passage de la recherche à l'utilisation opérationnelle simplifiée grâce une infrastructure en micro-services