

Recrutement d'un Post-Doctorant à l'École Nationale de l'Aviation Civile (ENAC)

1 Contexte

Nous souhaitons développer, au sein du groupe probabilités et statistiques du laboratoire MAIAA de l'ENAC, des compétences en "Big Data" et en "grands réseaux".

1.1 Big Data

Le terme "Big Data" (ou grande quantité de données) désigne une masse considérable de données hétérogènes et réparties, mais il désigne également les outils informatiques mis en œuvre pour les analyser, dans le but d'en retirer une valeur ajoutée supplémentaire. Techniquement, le "Big Data" invite à relever de nombreux défis, notamment concernant les capacités de stockage et d'analyse. Le Big Data intervient dans des domaines très variés comme : le marketing (données web, commerce en ligne ...), l'aéronautique et l'automobile (informations fournies par des milliers de capteurs, modèles numériques avec plusieurs milliards de mailles), la documentation interne ou externe, les réseaux sociaux, la géo-localisation, la santé et la biostatistique.

Par conséquent, le Big Data demande des compétences spécifiques en informatique (stockage, capture, recherche), en statistique (analyses des données, prospectives) et visualisation (visualisations de grandes quantités de données). En particulier, le coût de la lecture de grandes quantités de données réparties à travers le web conduit à la création de nouveaux algorithmes de traitement minimisant le nombre d'accès à ces données. En outre, il semblerait qu'à la différence de l'approche classique, consistant à récolter des données spécifiques à une problématique que l'on désire modéliser, le Big Data nécessite une nouvelle approche consistant à inférer des modèles à partir de grandes quantités de données.

L'utilisation des techniques liées au Big Data pourrait conduire dans l'aviation civile à des améliorations dans divers domaines tels que la sécurité (safety), l'efficacité des vols (en termes de déviation par rapport à une route pré-définie) et la réduction des retards.

Nous nous sommes par ailleurs rapprochés des écoles d'ingénieurs et de l'Université de Toulouse dans le cadre de l'appel à projet du programme IDEX « Développer des programmes de formation de haut niveau en ingénierie » sur le thème « Statistique et Informatique pour le Big Data ». Si ce projet venait à être accepté, le post-doctorant y contribuerait parfaitement.

1.2 Grands réseaux

Le terme réseau s'emploie ici pour désigner un système d'interactions ou de relations entre diverses entités (personnes, neurones, villes, entreprises,...) pouvant être représenté de façon abstraite sous la forme d'un graphe : les sommets du graphe étant les entités et les arêtes matérialisant les relations ou interactions. L'adjectif « grand » est utilisé avec le sens d'un très grand nombre de sommets.

La plupart des réseaux étudiés en pratique (réseaux sociaux, réseaux de transport, réseaux de neurones,...) présentent une architecture complexe, c.-à-d. dont les propriétés diffèrent des réseaux étudiés avant les années 2000 (réseaux réguliers et réseaux aléatoires de type Erdős-Rényi). Les recherches entreprises depuis 2000 ont contribué au développement d'une « science des réseaux » qui cherche à spécifier les propriétés partagées par les réseaux du monde réel en vue d'obtenir un modèle générique. En outre, l'accroissement de données disponibles a aussi permis d'appliquer ces recherches dans des disciplines spécifiques telles que par exemple : la sociologie, la politique, la biologie et les transports. Comme exemple propre à l'aviation civile, mentionnons le réseau dont les sommets sont les aéroports répartis à travers le monde et dont les arêtes représentent l'existence d'une liaison aérienne directe entre les deux aéroports associés aux sommets. Ce réseau peut aussi être orienté s'il est tenu compte du sens de la liaison aérienne, et pondéré si le nombre annuel de vols est associé à chaque arête, mais

aussi en associant à chaque sommet la capacité de l'aéroport par exemple. La création et l'étude de ce réseau pourraient servir dans des études d'économie du transport aérien, de complexité du trafic, etc. A ce titre, nous avons participé avec Sonia Cafieri du groupe MAIAA/optimisation à l'appel à projet du programme Université de Toulouse -IDEX « Transversalité » sur le thème « Toulouse Complex Network ». Si ce projet venait à être accepté, le post-doctorant pourrait y apporter sa contribution.

L'étude de ces réseaux fait appel à divers outils mathématiques dont principalement la théorie des graphes, les probabilités et les statistiques. Par ailleurs, calculer les caractéristiques de ces réseaux et/ou les générer aléatoirement à parti d'un modèle nécessitent la mise au point de nouveaux algorithmes. Pour clore cette partie, notons le lien avec le Big Data en raison de l'utilisation d'une grande quantité de données hétérogènes et l'inférence de propriétés du réseau obtenu. D'où l'idée d'associer ces deux problématiques dans cette fiche de poste.

2 Missions

Le candidat contribuera au démarrage d'une activité Big Data au sein du groupe probabilités et statistiques en collaboration avec nos partenaires du projet IDEX. A ce titre, outre des travaux de recherches, le candidat mènera des actions de diffusion et de sensibilisations auprès des enseignants, des chercheurs et des élèves sur les apports dans l'aviation civile de cette activité et sur les méthodes et développements en cours. Il contribuera également au développement de l'activité sur les grands réseaux, avec pour objectif une application dans le domaine de l'aviation civile. Nous serons aidés par le LEEA pour la collecte de données. Cette activité de recherche sera aussi accompagnée d'une activité de diffusion.

Le candidat pourra être éventuellement mis à contribution dans des tâches d'enseignement auprès des élèves ingénieurs IENAC et à l'encadrement de stagiaires (ingénieurs, Master, ...). Il s'attachera aussi à collaborer avec les partenaires institutionnels et industriels du laboratoire.

3 Profil

Le candidat :

- justifiera d'une très bonne connaissance en statistique et en probabilité, principalement dans l'analyse des données ;
- justifiera d'une bonne capacité à l'utilisation et la programmation des outils informatiques utilisés dans le "Big data" (bases de données, Hadoop, logiciel R, logiciels pour les graphes, ...);
- justifiera d'un bon niveau en anglais.

4 Formalités

Le poste est à pouvoir sur 2 ans au sein du groupe probabilités-statistiques du laboratoire MAIAA de l'ENAC (Toulouse). La date limite de candidature est fin juin 2014, et une réponse sera communiquée autour du 15 juillet. Le contrat commencera début octobre 2014.

5 Contact

Pascal Lezaud
ENAC
7 avenue Edouard Belin
CS 54005
31055 Toulouse Cedex FRANCE

<mailto:lezaud@recherche.enac.fr>
<http://www.enac.fr>
<http://maiaa.recherche.enac.fr>