



Offre de stage

Sujet : Graphes en grande dimension : méthodes mathématiques/statistiques de comparaison et débruitage

Possibilité de poursuivre sur une thèse

La Chaire Data Science and Artificial Intelligence for Digitized Industry and Services

Portée par Florence d'Alché-Buc, enseignante-chercheur dans le département Image, Données, Signal de Télécom Paris, la chaire DSAI réunit cinq partenaires industriels : Airbus Defence & Space, Engie, Idemia, Safran et Valeo Finance. Son objectif général est de développer, en liaison étroite entre les Parties, une formation et une recherche de niveau international.

Ses quatre principaux axes de recherche sont :

1. Analyse et prévision de séries temporelles (Predictive Analytics on Time Series) ;
2. Exploitation de données hétérogènes, massives et partiellement étiquetées (Exploiting Large Scale and Heterogeneous, Partially Labelled Data) ;
3. Apprentissage pour une prise de décision robuste et fiable (Learning for Trusted and Robust Decision) ;
4. Apprentissage dans un environnement dynamique (Learning through Interactions with a Changing Environment).

Description du stage

Encadrement

Stephan Clémenton, Télécom Paris

Jamal Najim, Université Gustave Eiffel

Lieu et dates du stage

Telecom Paris, 19 Place Marguerite Perey, 91120 Palaiseau

Université Gustave Eiffel, Bâtiment Copernic

5 Boulevard Descartes 77420 Champs-sur-Marne

Date de début du stage: Avril 2022 (peut -être pourvu plus tôt si besoin)

Équipe(s) d'accueil de la thèse

Equipe S2A à Télécom Paris

Equipe Modèles et Méthodes en Signal, Images et Données à l'UGE

Mots clés

Graphes en grande dimension, graphes aléatoires, méthodes spectrales pour les graphes, apprentissage statistique.

Sujet détaillé

Graphes en grande dimension : méthodes mathématiques et statistiques de comparaison et débruitage

Un graphe est un objet permettant d'encoder des informations de nature variée, souvent in-homogènes, entre agents : proximité, communautés, connectivité, etc. avec des structures d'interdépendance (statistique) parfois très complexes.

Les graphes en grande dimension apparaissent naturellement dans la production de données contemporaines et motivent de nombreux travaux en apprentissage statistique.

Dans le cadre de ce stage, on s'intéressera précisément à deux types de problèmes:

- (1) explorer les techniques de comparaison de graphes : comment comparer 2 graphes de dimensions possiblement différentes et dégager d'éventuelles similarités
- (2) explorer des techniques de débruitage/reconstruction sur graphe : en partant du principe qu'un graphe observé contient beaucoup de bruit mais qu'il existe une structure sous-jacente informative, comment extraire/estimer cette structure ?

Le stage commencera par une étude bibliographique visant à identifier l'état de l'art sur ces deux questions, en particulier les travaux mathématiques et statistiques permettant une quantification des erreurs de comparaison/estimation. Des simulations accompagneront l'étude bibliographique dans la perspective de tester les algorithmes et procédures proposés dans la littérature.

Après avoir identifié un nombre limité de problèmes se prêtant à une analyse quantitative, sachant qu'il y a de nombreux angles d'attaque en théorie des graphes : analyse spectrale des graphes, analyse combinatoire, statistique, etc., le stagiaire ébauchera le développement de solutions simulatoires et/ou analytiques pour les problèmes de comparaison et de débruitage. A travers ce stage, on s'attachera en particulier à explorer l'usage de la sparsification de graphes (voir e.g. [1],[2]) pour effectuer des tâches de modélisation, d'estimation et d'apprentissage, voir e.g. [3].

Profil du candidat

Formation avancée (M2) en probabilités et statistiques, bonne maîtrise de python. Une expérience en mathématiques discrètes sera appréciée mais n'est pas obligatoire.

Candidatures

Les candidatures doivent être envoyées, accompagnées d'un CV et d'un relevé de notes, à:

stephan.clemencon@telecom-paris.fr

jamal.najim@univ-mlv.fr

Références

- [1] Spectral Sparsification of Graphs. D. Spielman & S.H. Teng. *SIAM J. Comput.*, 40(4), 981–1025, 2011.
- [2] Spectral Sparsification of Graphs: Theory and Algorithms. J. Batson, D. Spielman, N. Srivastava & S.H. Teng. *Communications of the ACM*, vol. 56, No. 8, 2013.
- [3] On Graph Reconstruction via Empirical Risk Minimization: Fast Learning Rates and Scalability. G. Papa, S. Cléménçon & A. Bellet, *Neurips*, 2016.