

Compressive sensing et reconstruction de matrices/graphes

Projet

La collecte de grands flux de données de grande dimension (telles que les graphes par exemple) imposent de gérer de façon optimisée le recueil et le stockage de l'information. À titre d'exemple, le stockage à chaque instant d'un graphe entier est souvent impossible sur une longue durée, voire parfois inutile au vue de la question posée concernant cette suite de graphes.

Par ailleurs lors de la collecte de données, il est presque systématique d'être confronté à des données manquantes. Pour les matrices/graphes, ceci peut être vu comme l'observation partielle d'un(e) graphe (matrice) qu'il s'agit pourtant de reconstruire le plus complètement possible. Ce problème est par exemple en lien avec les systèmes de recommandation et le célèbre challenge Netflix.

Des stratégies dites de "compressive-sensing" ont été développées au début des années 2000. Elles visent notamment à intégrer la compression à l'étape préliminaire d'acquisition des données afin de limiter les coûts de calculs et de stockage. Toutefois, ces approches ont été proposées à l'origine pour estimer des vecteurs sparses dans un contexte de grande dimension, et non des objets plus complexes tels que des matrices ou des graphes.

Ces approches peuvent également servir à la complétion de matrices observées partiellement dans un contexte de données manquantes.

Les objectifs du stage sont :

1. étudier l'idée originelle du compressive-sensing,
2. étudier les développements récents en compressive-sensing appliqué aux matrices de faible rang,
3. étudier comment ces approches peuvent être utilisées pour reconstruire le graphe d'origine en distinguant 2 cadres:
 - l'opérateur doit choisir comment recueillir l'information sur le graphe en vue de sa reconstruction ultérieure,
 - l'observation partielle du graphe est imposée,
4. mettre en œuvre les approches étudiées précédemment sur des exemples afin d'en évaluer empiriquement les performances.

Prérequis

Outre de bonnes compétences en probabilités et statistique, le candidat utilisera R ou matlab pour réaliser des simulations afin de vérifier empiriquement ses résultats.

Encadrant : Alain Celisse, maître de conférences.

Length: 4-6 mois.

Opportunity: Le stage pourrait donner lieu à une thèse.

Laboratory: MODAL équipe-projet Inria, Lille.

Contact: Alain Celisse (celisse@math.univ-lille1.fr).

Suite: Le sujet proposé pourrait faire l'objet d'une thèse.

Bibliographie

- Compressive sampling, Emmanuel Candes, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006
- A Simpler Approach to Matrix Completion, Benjamin Recht, Journal of Machine Learning Research 12 (2011) 3413-3430