

Apprentissage de transformées rapides

Stage de recherche de M2

Équipe [QARMA](#), Laboratoire d'Informatique et Systèmes ([LIS](#))

Contact: Valentin Emiya, Luc Giffon - prenom.nom@lis-lab.fr

Mots clés : informatique, apprentissage automatique, optimisation, transformées rapides.

Durée : six mois environ.

Rémunération : environ 500€ mensuels selon montant réglementaire.

Possibilités de débouché en doctorat : oui, en concertation avec le candidat.

Contexte scientifique. Ce stage de recherche s'inscrit dans une problématique d'élaboration de modèles d'apprentissage automatique économes en quantité de calculs et en espace de stockage. L'étude d'opérateurs linéaires rapides constitue une piste de réflexion prometteuse dans cette optique [1].

Les opérateurs linéaires rapides tels que la transformée de Hadamard peuvent être exactement exprimés comme des produits de matrices creuses, dites parcimonieuses. Cette considération a récemment ouvert la voie à la conception d'algorithmes permettant d'apprendre automatiquement les motifs de parcimonie ainsi que les poids de ces factorisations [2, 3]. Certains de ces travaux sont allés plus loin et ont émis l'hypothèse que n'importe quel opérateur linéaire pourrait être approximé par une factorisation de matrices parcimonieuses, en offrant ainsi une approximation rapide.

Objectifs et déroulement. L'objectif de ce stage est d'étudier différents axes d'amélioration des modèles et algorithmes d'optimisation existants permettant d'approximer un opérateur linéaire par un produit de facteurs parcimonieux. L'un de ces axes d'amélioration est l'apprentissage des motifs de parcimonie : en effet, du fait de la combinatoire sous-jacente, les méthodes employées jusqu'alors sont pour l'instant assez peu satisfaisantes car biaisées dans la structure des matrices apprises ou instables numériquement.

Ce stage se déroulerait en trois parties : (i) une étude bibliographique des modèles d'intérêt et des algorithmes d'optimisation associés ; cette étude pourra s'accompagner d'une mise à niveau sur certains aspects afin de compléter les connaissances du stagiaire ; (ii) l'élaboration, sur la base d'heuristiques ou résultats théoriques et en fonction des compétences et motivations du candidat, d'une ou plusieurs améliorations à apporter aux modèles et aux algorithmes ; (iii) enfin, une étude empirique des bénéfices de la méthode sera réalisée dans le cadre d'expérimentations rigoureuses et reproductibles.

Équipe d'accueil, encadrement. Le stagiaire intégrera l'équipe d'apprentissage automatique de Marseille ([QARMA](#)) au sein du Laboratoire d'Informatique et Systèmes ([LIS](#)). L'encadre-

ment sera assuré par Valentin Emiya et Luc Giffon, avec des interactions possibles avec d'autres membres de l'équipe [QARMA](#), en fonction des orientations choisies.

Profils recherchés : compétences générales en informatique et mathématiques, bases en apprentissage automatique, attirance pour le sujet, programmation en Python/Matlab, ouverture d'esprit, motivation pour la recherche, bonne capacité de travail. Des compétences en optimisation seraient un plus.

Procédure de candidature. Les étudiants intéressés sont invités à prendre contact le plus tôt possible avec Valentin Emiya et Luc Giffon. Après un premier échange informel permettant de donner plus d'informations sur le stage, le candidat sera invité à envoyer un dossier de candidature (CV, lettre de motivation, relevés de notes des années passées, éventuels rapports de projets/stages réalisés et lettres de recommandation), en vue d'un entretien.

Références.

- [1] L. Giffon, V. Emiya, L. Ralaivola, and H. Kadri. QuicK-means : Acceleration of K-means by learning a fast transform. technical report available online, September 2019.
- [2] L. Le Magoarou and R. Gribonval. Flexible Multilayer Sparse Approximations of Matrices and Applications. *IEEE Journal of Selected Topics in Signal Processing*, 10(4), June 2016.
- [3] T. Dao, A. Gu, M. Eichhorn, A. Rudra, and C. Ré. Learning fast algorithms for linear transforms using butterfly factorizations. In *Proceedings of ICML, 2019*.