

Implantation d'une interface utilisateur pour l'exploration interactive d'un ensemble de motifs extraits

Stage de six mois à effectuer entre février et août 2023

Cette annonce propose un stage de 6 mois destiné à un·e étudiant·e en cinquième année d'informatique (M2 ou école d'ingénieurs). Le stage s'inscrit dans le projet ANR-20-CE23-0023 InvolvD¹. Le sujet principal est le développement d'une interface utilisateur, outil indispensable pour permettre aux expert·e·s en pharmacie de bénéficier de notre nouvel algorithme d'exploration de résultats expérimentaux. Le stage sera encadré par Ronan Bureau, Bertrand Cuissart et Etienne Lehembre. Le·La stagiaire sera accueilli·e au sein du GREYC, le laboratoire d'informatique de l'Université de Caen Normandie.

Contexte technique et scientifique

Dans le cadre d'InvolvD, nous avons récemment développé un algorithme visant à accompagner un·e expert·e dans la découverte d'un espace de données structurées. L'algorithme aillant fait ses preuves lors de tests avec des oracles synthétiques, nous souhaitons passer à une phase expérimentale concrète en impliquant des expert·e·s humain·e·s. Le stage consiste à réaliser l'interface utilisateur qui permettra à un·e pharmacien·e de choisir les parties des résultats qui l'intéressent au premier chef.

Les éléments d'étude sont des graphes étiquetés appelés pharmacophores [2] issus d'un calcul de fouille de données. L'ensemble des pharmacophores est structuré grâce à la relation d'inclusion entre les graphes. L'interface a pour but d'offrir à l'expert·e un moyen efficace de parcourir cette structure de façon à alimenter l'algorithme d'apprentissage par renforcement. L'objectif étant de limiter la frustration et le manque d'attention de l'expert·e [1], il est important que l'interaction ne se résume pas à une suite de questions - réponses.

Suite aux réponses de de l'expert·e, l'algorithme actualise l'intérêt de chaque pharmacophore pour l'analyse. Pour traduire cette évolution, la visualisation de la structure à parcourir doit évoluer en conséquence.

La réalisation de l'interface sera prolongée par un travail destiné à évaluer les performances de l'algorithme d'évaluation de l'intérêt des pharmacophores. Pour cette partie du travail, il sera indispensable d'avoir un échange de nature interdisciplinaire avec les chercheurs en pharmacie.

Enfin, le stage se conclura par un travail plus ouvert et axé sur le choix stratégique à associer au parcours des pharmacophores. On peut privilégier une stratégie d'exploitation

¹<https://involvd.greyc.fr/>

associée à un parcours plutôt de proches en proches, on peut opter pour une stratégie d'exploration qui privilégie les pharmacophores associés aux endroits peu explorés par l'analyse ou on peut imaginer des compromis entre ces deux stratégies.

Technologies envisagées La réalisation de l'interface s'appuiera sur une méthode classique MVC (Modèle - Vue - Contrôleur) où le modèle serait le code C++ fourni. Il sera nécessaire d'intégrer le code à un wrapper Python pour mettre en place les contrôleurs communiquant avec la vue qui utilisera Dash Cytoscape^{2,3} ; Cytoscape est un logiciel de visualisation de graphes déjà existant.

Apports du stage Le·la stagiaire sera intégré·e au sein de l'équipe CODAG du GREYC, laboratoire d'informatique normand. Le projet ANR InvolvD impliquant des chercheurs·euses de plusieurs laboratoires français, l'étudiant·e aura l'occasion d'échanger avec plusieurs spécialistes dans le contexte d'une recherche interdisciplinaire. Ces échanges seront accompagnés par une intégration dans le groupe chimie-informatique caennais, groupe qui compte une vingtaine de membres et qui se réunit mensuellement pour échanger. L'étudiant·e aura ainsi plusieurs occasions de présenter ses travaux dans un contexte collaboratif. De plus, le travail étant un travail de recherche académique, il se concrétisera par la rédaction d'une communication scientifique qui sera soumise à la communauté scientifique (poster, workshop, article de conférence, ou journal). Enfin, en réalisant ce stage, l'étudiant·e va acquérir des connaissances très intéressantes concernant la "chemoinformatique", domaine interdisciplinaire visant à réaliser des avancées informatiques pour mieux appréhender le monde de la chimie.

Contacts

- etienne.lehembre@unicaen.fr
- bertrand.cuissart@unicaen.fr

Pour candidater, envoyez-nous votre CV, une lettre de motivation, vos bulletins de notes du MASTER 1 et 2, si vous en disposez, ainsi que toute autre pièce jugée utile. Nous préférons recevoir les fichiers au format pdf.

References

- [1] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. Power to the people: The role of humans in interactive machine learning. *Ai Magazine*, 35(4):105–120, 2014.
- [2] Jean-Philippe Métivier, Bertrand Cuissart, Ronan Bureau, and Alban Lepailleur. The pharmacophore network: a computational method for exploring structure–activity relationships from a large chemical data set. *Journal of Medicinal Chemistry*, 61(8):3551–3564, 2018.

²<https://dash.plotly.com/cytoscape>

³<https://github.com/plotly/dash-cytoscape>