

PROPOSITION DE STAGE

M2 INFORMATIQUE ou BIONFORMATIQUE



Titre : Réseaux de neurones pour la prédiction et la classification des ARN non-codants

Mots-clés : «Machine learning, Réseaux de neurones, Cartes SOM, Apprentissage supervisé, Apprentissage non-supervisé, développement logiciel, bioinformatique, prédiction d'ARN non-codants

Maîtres de stage :

Fariza Tahi, Professeur des Universités, Equipe AROBAS, laboratoire IBISC, Université Paris-Saclay, Université d'Evry, Genopole

Farida Zehraoui, Maître de Conférences, Equipe AROBAS, laboratoire IBISC, Université Paris-Saclay, Université d'Evry

Lieu du stage : Laboratoire IBISC, IBGBI, Université d'Evry, 23 Boulevard de France, 91000 Evry.

Durée de stage : 6 mois

Contact : fariza.tahi@univ-evry.fr

Résumé :

Les ARN non-codants (ARNnc) sont depuis quelques années au cœur de recherches particulièrement attrayantes et intensives, notamment dans le domaine biomédical. En effet, les ARNnc sont impliqués dans un grand nombre de maladies, telles que le cancer, le diabète, les maladies neuro-dégénératives, et suscitent un intérêt croissant dans leur utilisation comme biomarqueurs ou comme cibles thérapeutiques. Grâce aux nouvelles générations de séquençage (NGS), le profilage détaillé des molécules d'ARNnc a été considérablement améliorée. Un grand nombre de classes et de types d'ARNnc ont été découverts, mais beaucoup d'autres classes sont encore à découvrir.

L'identification de ces ARNncs est donc un enjeu important, et avec les NGS qui génèrent des volumes considérables de données RNAseq notamment, leur prédiction par des méthodes *in silico* est nécessaire. Ces méthodes permettront en effet d'identifier des ARNncs potentiels pouvant ensuite être validés par des méthodes expérimentales.

Le but de ce stage est de développer une méthode d'apprentissage automatique multi-sources combinant l'apprentissage supervisé et le non-supervisé, afin d'identifier les ARNncs connus et de découvrir de nouvelles classes, en utilisant des sources de données hétérogènes représentant différentes caractéristiques potentielles associées à ces ARN : leur structure secondaire, la présence de certains motifs dans leur séquence, des caractéristiques épigénétiques, etc. L'algorithme se basera sur des réseaux de neurones supervisés de type perceptron multi-couches (ou MLP) et des réseaux de neurones non-supervisés de type cartes auto-organisatrices (ou SOM pour Self Organizing Maps) pour permettre une meilleure visualisation et interprétation des résultats.

Ce stage fait suite aux travaux menés dans le cadre de la thèse de Ludovic Platon (voir les publications ci-dessous), et s'inscrit dans la continuité des travaux de prédiction des ARNncs et de leurs structures menés dans l'équipe AROBAS depuis plusieurs années. Ces travaux ont donné lieu à plusieurs algorithmes et logiciels publiés dans des revues internationales et mis à disposition de la communauté scientifique via le serveur web : <http://EvryRNA.ibisc.univ-evry.fr>

Bibliographie :

- L. Platon, F. Zehraoui, F.Tahi. Localized Multiple Sources Self-Organizing Map. International Conference on Neural Information Processing, ICONIP (3) 2018 : 648-659.
- L. Platon, F. Zehraoui, A. Bendahmane, F. Tahi. IRSOM, a reliable identifier of ncRNAs based on supervised Self-Organizing Maps with rejection. Bioinformatics, 34(17), i620-i628, 2018.
- L. Platon, F. Zehraoui, F.Tahi. Self-Organizing Maps with supervised layer. In Proceedings of the 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM+). 2017: 161-168.
- Boucheham A, Sommard V, Zehraoui F, Boualem A, Batouche M, Bendahmane A, Israeli D, Tahi F. PLoS One 2017, 16;12(6):e0179787.

