Thèse Cifre : « Fouilles de données massives pour l'identification d'empreintes comportementales de l'exposition aux contaminants en écotoxicologie aquatique »

Partenaire industriel employeur : Société Viewpoint www.viewpoint.fr
Laboratoire d'accueil : INRAE, UR RiverLy, Laboratoire d'écotoxicologie, Lyon-Villeurbanne.

Encadrement scientifique :

- ***A Chaumot*** *- INRAE Laboratoire d'écotoxicologie* www6.ara.inrae.fr/lyon-riverly-ecotox
- ***J-L Bertrand-Krajewski -*** *INSA Lyon Laboratoire Deep (EA7429)* deep.insa-lyon.fr

Ecole doctorale : E2M2, Université Lyon 1

Mots-Clefs : Data mining, Classification, Apprentissage, IA, Vidéo-tracking, Analyse d'image, Analyse comportementale, Diagnostic, Biosurveillance, Contamination chimique, Biodiversité

Problématique

La préservation de la ressource en eau et de la biodiversité aquatique sont des préoccupations sociétales majeures dont une des prérogatives est de limiter le rejet de contaminants toxiques dans l'environnement aquatique. C'est dans ce cadre que l'auto-surveillance et l'évaluation de la toxicité des rejets industriels et urbains font progressivement leur apparition dans la législation. Dans ce contexte, un partenariat entre la société *Viewpoint* - spécialisée dans le développement de dispositifs d'analyse vidéo pour le suivi du comportement d'organismes dans le contexte d'études médicales - et le laboratoire d'écotoxicologie de INRAE Lyon a abouti au co-développement du dispositif *ToxMate* capable d'évaluer sur site, en ligne et en temps réel la qualité des eaux via l'analyse par vidéo-tracking du comportement locomoteur de 3 espèces d'invertébrés aquatiques (crustacé, escargot, sangsue). S'appuyant sur ce développement, la thèse proposée s'inscrit dans un projet voulant lever un verrou majeur en écotoxicologie et aux retombées fortes dans le domaine de la gestion environnementale : l'identification des substances responsables de la toxicité observée dans l'environnement. S'inscrivant ainsi dans une question scientifique majeure - le lien entre diversité des substances présentes au sein des cocktails de contaminants de l'environnement et variabilité de la réponse biologique entre espèces - notre hypothèse est qu'il doit être possible d'identifier dans la complexité des réponses comportementales mesurées sur une diversité d'espèces, des patrons de réponses spécifiques qui peuvent être reliés à la nature des contaminants (notion d'empreintes comportementales). Cette hypothèse s'appuie d'une part sur des résultats préliminaires que nous avons obtenus lors d'exposition à une quinzaine de micropolluants de diverses classes (métaux lourds, médicaments, insecticides, herbicides, biocides…) sur les 3 espèces d'invertébrés aquatiques aujourd'hui suivies dans le dispositif *Toxmate*, et d'autre part sur une étude menée il y a une dizaine d'années dans le domaine des neurosciences chez le poisson (Rihel et al. 2010 Science 327(5963): 348-351).

Ainsi, au regard de nos retours d'expérience qui montrent des premières différences de réponses comportementales entre espèces et face au constat que la quantité d'information délivrée par le *ToxMate* (1 image toutes les 40 ms sur 30 jours sur 48 individus suivis indépendamment) est aujourd'hui sous-exploitée en termes de traitement du signal, l'objectif de la thèse est de réaliser la preuve de concept, en laboratoire et sur site, que l'on peut établir par l'utilisation de méthodes statistiques avancées et d'apprentissage automatique (machine learning, IA) une relation entre substances et descripteurs comportementaux de 3 invertébrés aquatiques choisis parmi 3 groupes bien distincts de la diversité animale. Pour mener à bien la mise en œuvre de ces techniques de fouille de données, le travail de thèse associera le laboratoire Deep de l'INSA Lyon dont l'expertise est

reconnue pour l'interprétation des données massives générées dans le domaine de la métrologie de l'environnement.

Méthodologie

S'appuyant sur la disponibilité d'un jeu de données vidéo-tracking massif (dont l'acquisition sera finalisée en parallèle du démarrage de la thèse lors d'une étude expérimentale sur une centaine de substances), la thèse aura en charge la mise en œuvre d'analyses statistiques et de modélisations pour l'identification des empreintes comportementales. Une première réflexion devra être portée sur la définition de métriques calculées à partir des données brutes de position de 16 individus par espèce, métriques devant permettre de décrire la diversité comportementale observée entre espèces lors de l'exposition sur 48h aux différentes substances. Pour identifier des empreintes comportementales propres à chaque type de contaminants, la première approche sera de recourir à différentes techniques statistiques de type analyse discriminante, ou encore de classification supervisée par apprentissage automatique (random forests, réseaux de neurones artificiels) en s'appuyant sur une classification des contaminants fonction de leurs modes d'action ou de leurs propriétés structurales. Une sortie attendue sera d'établir un algorithme d'analyse permettant d'identifier une empreinte substance-spécifique pour différents composés d'intérêt (notamment industriel). Une seconde stratégie d'analyse consistera à définir une classification sans a priori des patrons comportementaux (approches multivariées ACP, clustering, cartes auto-adaptatives, …) se basant là aussi sur un très grand nombre de métriques calculées à partir des profils comportementaux. Une projection a posteriori des substances dans ces classifications sera réalisée pour tenter d'associer sans a priori les substances à ces empreintes ainsi définies de novo.

Profil recherché

Titulaire d'un **Master** dans le domaine des **Sciences des données** (data mining, machine learning, IA) attiré(e) par les applications dans le domaine biologique au sens large (santé, environnement). Capacité d'initiative et forte aptitude au travail en **équipe interdisciplinaire** sont ici des prérequis.

Contact

Arnaud Chaumot INRAE : arnaud.chaumot@inrae.fr
Jean-Luc Bertrand-Krajewski INSA : jean-luc.bertrand-krajewski@insa-lyon.fr
Alexandre Decamps VIEWPOINT : adecamps@viewpoint.fr

Démarrage prévu automne 2020.

Candidature

Envoyer CV et lettre de motivation au plus tard le 31 mai 2020 aux trois contacts indiqués ci-dessus.