PhD Position: « Massive data mining for the identification of behavioral fingerprints of contaminant exposure in aquatic ecotoxicology »

Industrial employer partner: Viewpoint www.viewpoint.fr
Host academic Laboratory: INRAE, UR RiverLy, Ecotoxicology Lab, Lyon-Villeurbanne.

Scientific supervision:

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

Keywords: Data mining, Classification, Deep Learning, AI, Video-tracking, Image analysis, Behavioral Analysis, Diagnostic, Biomonitoring, Chemical contamination, Biodiversity

Scientific issue and objectives:

The preservation of water resources and aquatic biodiversity are major societal concerns, one of those prerogatives is to limit the release of toxic contaminants into the aquatic environment. It is within this framework that self-monitoring and evaluation of the toxicity of industrial and urban effluents are progressively appearing in legislation. In this context, a partnership between Viewpoint – a company specialized in the development of video analysis devices for monitoring the behavior of organisms in the context of medical studies - and the ecotoxicology laboratory of INRAE in Lyon led to the co-development of the ToxMate device. This device proposes on-site, online and real-time assessment of water quality via video-tracking analysis of the locomotor behavior of 3 aquatic invertebrate species (crustaceans, snails, leeches). Based on this development, the proposed PhD is part of a project that aims to remove a major obstacle in ecotoxicology and with strong implications in the field of environmental management: the identification of the substances responsible for the toxicity observed in the environment. This will address the major scientific question of the link between the diversity of substances present in environmental contaminant cocktails and the inter-species variability of the biological response. Our hypothesis is that it should be possible to identify specific response patterns within the complexity of behavioral responses measured on a diversity of species that can be linked to the nature of the contaminants (notion of behavioral fingerprints). On the one hand, this hypothesis is based on preliminary results which we obtained on the three species of aquatic invertebrates currently monitored in the ToxMate device during exposure to about fifteen micropollutants of various classes (heavy metals, drugs, insecticides, herbicides, biocides, etc.). On the other hand, we have based this hypothesis on a study conducted on fish about ten years ago in the field of neuroscience (Rihel et al. 2010 Science 327(5963): 348-351).

Thus, taking into account that the amount of information delivered by the ToxMate (1 image every 40 ms over 30 days on 48 individuals followed independently) is today under-exploited in terms of signal processing, the objective of this PhD thesis is to carry out the proof of concept that the use of advanced statistics and machine learning methods can establish a relationship between the nature of substances and the behavior of 3 aquatic invertebrates chosen from 3 distinct groups of animal diversity. To carry out the implementation of these data mining techniques, the thesis work will involve the laboratory Deep of INSA Lyon, whose expertise is recognized for the interpretation of massive data generated in the field of environmental metrology.

Methodology:

Relying on the availability of a massive videotracking data set (whose acquisition will be finalized in parallel with the start of the PhD work during an experimental study on about a hundred of substances), the PhD will be in charge of implementing statistical analyses and modeling approaches for the identification of behavioral fingerprints. The first step will involve the definition of metrics calculated from the raw position data of 16 individuals per species, metrics that should lead to a description of the behavioral diversity observed between species during exposure over 48 hours to different substances. Then, to identify behavioral fingerprints specific to each type of contaminant, the first approach will use different statistical techniques such as discriminant analysis, or supervised classification by automatic learning (random forests, artificial neural networks) based on a classification of contaminants according to their modes of action or structural properties. An expected output will be an algorithm allowing the identification of a substance-specific fingerprint for different compounds of interest (notably industrial interest). A second analysis strategy will consist in defining a classification of behavioral patterns with no a priori (multivariate ACP approaches, clustering, self-organizing maps,...) based again on a large number of metrics calculated from behavioral profiles. Afterwards, a projection of the substances in these classifications will be carried out in order to associate the substances to these patterns defined de novo.

Profile:
Holder of a **MSc degree** in the field of **Data Sciences** (data mining, machine learning, AI) interested in biological applications (medicine, environment). Both an ability to take initiative and a strong ability to work in an **interdisciplinary** team are mandatory.

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

Scheduled start date: Autumn 2020

Application:
Send CV and cover letter by 31 May 2020 at the latest to the three contacts indicated above.