ANR-11-INBS-0010

Machine learning Cheminformatics

About MetaboHUB

MetaboHUB (MTH) is the national French metabolomics and fluxomics infrastructure. Launched in 2013, MTH is a leading international infrastructure serving more than 700 scientists worldwide. MetaboHUB gathers 5 regional facilities including more than 80 permanent staffs, 15 NMRs, 43 MS, robotic and computational platforms. MTH aims at pushing forward the field to develop metabolomics and fluxomics from single cell to population. Your contribution will serve a broad range of researchers in the fields of biotechnologies, Human health and nutrition and plant science. Joining MTH, you will be involved in cutting edge research within a highly skilled and motivated consortium.

About MTH-Metabolome-IDF

The Metabolome-IDF platform has been specialized for more than 15 years in metabolomics, lipidomics, glycomics by mass spectrometry (MS) and data science for biomarker discovery in health. You will join the data science team (Odiscé) at CEA Saclay which develops innovative methods in applied mathematics and statistics, for the high-throughput processing (signal processing), integrative statistical analysis (machine learning) and annotation (cheminformatics) of MS data within large cohorts. The data and algorithms will be made publicly available to the community as software libraries (R/Bioconductor or Python) and Galaxy/Workflow4Metabolomics modules.

The mission

Determining the 2D structure of a compound given its MS/MS spectrum, i.e. the list of (mass/intensity) tuples of its fragments, is a major challenge in metabolomics [1]. Current reference method rely on the prediction of a molecular property vector (fingerprint) using a set of Support Vector Machines, that can be subsequently matched to those from known compounds in databases. Performances, however, are limited to 30% of correct structures [2]. Recently, alternatives have been suggested using artificial neural networks to further take into account the interactions between features [3].

Key responsibilities

You will first benchmark the open source prediction tools against the consortium's data, as well as against the CASMI challenge data. The model will then be enriched with new input features and output molecular properties, and the architecture will be optimized to improve the performances. Finally, the algorithms will be implemented into FAIR computational workflows for high-throughput and reproducible structure recommendation.

[1] Nguyen et al. (2019) Recent advances and prospects of computational methods for metabolite identification: a review with emphasis on machine learning approaches. *Briefings in Bioinformatics*, 20, 2028–2043.

[2] Schymanski et al. (2017) Critical Assessment of Small Molecule Identification 2016: automated methods. *Journal of Cheminformatics*, 9, 22.

[3] Fan et al. (2020) MetFID: artificial neural network-based compound fingerprint prediction for metabolite annotation. *Metabolomics*, 16, 104.

Profile

Master or PhD in
Machine learning/Deep learning
Cheminformatics

Skills

Scikit-learn, PyTorch
RDkit

Informations

CDD
12 months contract, full-time position
Starting from: January 2024

How to apply?

The application should contain the following attachments:

1. motivation letter
2. CV including contact details of two references
3. relevant diplomas or university certificates

Contact

Etienne Thévenot
Data sciences team (Odiscé)
UMR « Médicaments et Technologies pour la Santé » (SPI/LI-MS & LI2A)
CEA, Centre de Saclay
F-91191 Gif sur Yvette, France
etienne.thevenot@cea.fr

More informations

www.metabohub.fr

https://odisce.github.io

Post Reference: P4-E