

Application de l'IA à la prédiction des performances de bovins sur la base de spectres infra-rouge.

Mots-clefs : Intelligence Artificielle, Statistiques, Génétique, Spectres

Contact : julie.aubert@inrae.fr, hugo.gangloff@inrae.fr, tristan.mary-huard@agroparistech.fr

Contexte

En génétique animale, on s'intéresse à la prédiction d'indices génétiques de performance des vaches laitières en termes de qualité du lait. La prédiction de ces indices est réalisée à partir d'échantillons de lait analysés par spectroscopie infra-rouge. A cet effet, on considère un large jeu de données constitué dans le cadre du programme de phénotypage et génotypage pour la compréhension et la maîtrise de la composition fine du lait Phénofinlait. 800 000 spectres infra-rouges ont été collectés sur le lait de 57 000 vaches laitières. On dispose par ailleurs d'informations génétiques pour une sous-partie des animaux du panel : plus de 8000 vaches ont été génotypées pour plusieurs centaines de milliers de marqueurs bi-alléliques. On dispose aussi des indices de performances de qualité et de production pour l'ensemble des vaches du dispositif.

Sujet

L'objectif du stage est de mettre au point une méthode de prédiction d'indices sur la base des données spectrométriques collectées sur chaque individu. La méthode développée sera basée sur la méthodologie des réseaux de neurones convolutifs (Convolutional Neural Networks). On cherche notamment à savoir si la prédiction réalisée à partir de spectres permet d'obtenir des performances comparables à celles obtenues à partir de données génétiques. Dans un deuxième temps, on cherchera à combiner les deux types d'information (génétique et spectrale) afin d'optimiser les performances de prédiction. Enfin, on évaluera la pertinence d'une prédiction conjointe de l'ensemble des indices conjointement par rapport à une prédiction de chaque indice séparément.

L'implémentation de la méthode sera réalisée en Python, à l'aide de la bibliothèque pytorch. Le stagiaire aura accès à un serveur de calcul haute performance. Le stagiaire aura aussi l'occasion de collaborer avec les collègues généticiens de l'unité mixte de recherche Génétique Animale et Biologie Intégrative et de la fédération des entreprises de conseil et service en élevage Eliance, qui ont participé à la constitution du jeu de données.

Profil du/de la candidat(e)

- Master ou école d'ingénieur en apprentissage statistique/mathématique appliquée,
- Développement Python et expérience avec des bibliothèques de machine learning (scikitlearn ou autre),
- Intérêt pour les applications en sciences du vivant.

Conditions d'exercise

- Financement : projet DeepPhenomics du Metaprogramme INRAE DIGIT-BIO
- Lieu : MIA Paris-Saclay, Campus Agro Paris-Saclay, Université Paris Saclay, AgroParisTech INRAE, 22 place de l'Agronomie à Palaiseau
- Collaborateurs : Julie Aubert, Hugo Gangloff, Tristan Mary-Huard
- Durée : 5 à 6 mois à partir du printemps 2023

Références

[1] Pedregosa, Fabian, et al. "Scikit-learn : Machine learning in Python." *Journal of Machine Learning Research* 12 (2011) : 2825-2830.