

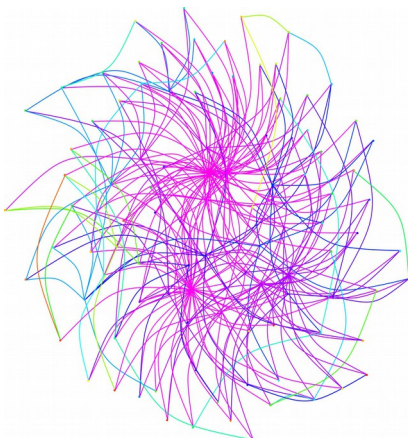
SCHEME, DATA MINING & MACHINE LEARNING OUTILS OPÉRATIONNELS.

Quelles informations utiles puis-je retirer de mes mesures de contrôle ? Comment extraire de l'information d'une base de données incomplète pour créer un modèle rapidement ? Quel est le niveau de fiabilité et de généralité de mon modèle prédictif ?

Voilà les questions auxquelles SCHEME peut répondre.

Trouver des liens et les transcrire en modèle

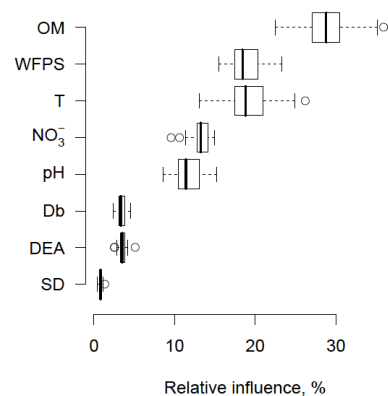
On retrouve les approches de Machine Learning dans de nombreux domaines opérationnels, comme la reconnaissance de motifs (reconnaissance de caractères, d'objets) et le développement d'intelli-



Machine Learning = Révéler les liens

gences artificielles. Avec la baisse des coûts de la puissance de calcul et de la mesure, les applications dans le domaine industriel et de l'environnement sont en plein développement.

L'objectif principal des approches de Machine Learning est de produire un modèle qui représente globalement les caractéristiques et les interactions contenues dans une base de données. Ce dernier permet une analyse fine de l'information contenue dans la base de données tout en étant très performant et simple d'usage. Ces approches sont particulièrement intéressantes quand les liens entre variables explicatives et processus sont a priori inconnus ou complexes. Les Machines Learning permettent souvent d'aller plus loin que les approches statistiques classiques qui nécessitent une connaissance préalable sur les liens et des hypothèses de travail contraignantes.



Exemple de hiérarchisation des variables explicatives du processus de dénitrification par BRT

Le type d'approche que nous utilisons sont **peu sensibles aux valeurs aberrantes, s'accommode des trous dans les données.** Elle permet d'évaluer l'**importance relative** des variables sur la réponse, **d'identifier les variables explicatives**, de les **hiérarchiser** et d'explicitier les liens.

Les algorithmes utilisés chez SCHEME

En fonction de la base de données, du problème, et des contraintes techniques d'usage, SCHEME utilise communément

les techniques suivantes :

- Réseaux de Neurones Artificiels
- Boosted Regression Trees (BRT)
- Support Vector Machines
- Random Forest

Pour qui et comment ?

La méthode s'applique à tous types d'activité et sur tous types de base de données. Ces dernières peuvent être de tailles relativement modestes : 9 mesures par variable à tester pour les algorithmes les plus robustes.

La procédure chez SCHEME est normalisée :

- En amont : analyse rapide de la base de données pour évaluer la faisabilité
- Analyse complète de la base de données
- Création du modèle
- Formation à l'utilisation du modèle et livraison du code

A titre informatif, **la réalisation peut être effectuée en une semaine.**

Pourquoi ?

Cela peut s'appliquer à des problèmes continus (un taux de production de méthane), de classification (discriminer des composants organiques selon leurs signatures spectrales), de présence - absence (distribution spatiale d'espèces invasives) , d'analyse de signal.

Les thématiques sont diverses : écologie, hydrologie, biogéochimie, physique nucléaire, procédés industriels, production d'énergie, biotechnologie, etc...

Références

SCHEME a été missionnée par CYLERGIE - SUEZ ENVIRONNEMENT afin de contrôler leurs procédés de production d'énergie. L'équipe a développé un modèle hybride physique/machine learning pour le National Institute of Water and Atmospheric Research (NIWA) en Nouvelle Zélande pour prédire les concentrations en phosphore et en azote sur l'ensemble des cours d'eau non-instrumentés du territoire national.