

Proposition de stage de M2
*Réseau bayésien étiqueté pour l'apprentissage
de la structure d'un réseau d'interactions
écologiques*

Contexte finalisé et problématique scientifique

Dans un écosystème, plusieurs types d'interactions peuvent exister entre les espèces : prédation, parasitisme, mutualisme,... Elles sont décrites sous la forme d'un réseau d'interactions écologiques, qui est un graphe dont les nœuds sont les espèces et où une arête indique une interaction directe entre deux espèces. Ce réseau est souvent partiellement connu car les interactions sont difficiles à observer. Néanmoins, on dispose parfois d'un certain nombre (souvent petit) d'observations indépendantes de l'état du système : présence-absence, classe d'abondance, comptage ou encore densité de chaque espèce. Ces données peuvent être exploitées pour apprendre la structure du réseau d'interaction.

Plusieurs approches probabilistes existent pour apprendre la structure d'un graphe à partir d'observations de variables exprimées en chaque nœud : apprentissage de la structure d'un réseau bayésien, d'un champ de Markov, ou encore d'un champ gaussien; apprentissage des poids dans un modèle de régression avec pénalité; calcul d'information mutuelle, ... Ces méthodes font chacune des hypothèses différentes. Ainsi le cadre des réseaux bayésiens permet d'apprendre un graphe orienté, alors que les autres approches apprennent un graphe non orienté. L'orientation des arêtes a du sens dans le cas d'interactions écologiques. Par contre les données sont souvent rares, et un modèle avec peu de paramètres comme un champ de Markov ou un champ gaussien sera plus facile à apprendre qu'un réseau bayésien décrit par ses tables de probabilités conditionnelles. Dans le cas d'observations de séries temporelles, le cadre des réseaux bayésiens dynamiques étiquetés (RBDE, Auclair et al. 2017) lève ce problème d'une part car le modèle est paramétré et d'autre part car il permet d'intégrer de la connaissance experte (structuration forte en niveaux trophiques, nombre limité de types d'interactions,...) pour aider l'apprentissage.

Dans ce sujet de stage, nous proposons de développer une méthode dans le cadre réseau bayésien, dédiée à l'apprentissage d'un réseau d'interactions à partir de données statiques, inspirée du cas dynamique des RBDE, puis de la confronter aux méthodes classiques d'apprentissage de réseaux.

Projet de Master

Il s'agira dans un premier temps de réaliser une étude bibliographique sur les principales méthodes probabilistes pour l'inférence de la structure d'un réseau à partir de données statiques.

Puis il s'agira de s'inspirer du cadre RDBE pour définir un modèle de RB étiqueté et une ou des lois a priori sur le graphe du modèle, adapté à l'application écologique, ainsi que la méthode d'apprentissage du graphe et des paramètres du RB.

Enfin, les performances de cette méthode et de celles de l'état de l'art seront comparées. Dans un premier temps, sur des exemples synthétiques, il s'agira d'essayer d'identifier les motifs qui sont les plus facilement retrouvés. Ensuite, les méthodes seront comparées sur un problème d'apprentissage du réseau d'interaction d'espèces marines présentes dans les forêts de kelp (Kusher et al. 2013).

Compétences attendues

- bon niveau en statistique ou machine learning
- pratique du logiciel Matlab
- goût pour les applications

Bibliographie

Auclair, E., Peyrard, N., Sabbadin, R. (2017). Labeled DBN learning with community structure knowledge. In *Proc. of the 27th European Conference on Machine Learning*.

Kushner, D. J., Rassweiler, A., McLaughlin, J. P., and Lafferty K. D. (2013). A multi-decade time series of kelp forest community structure at the California Channel Islands. *Ecology* 94:2655.
(<http://www.esapubs.org/archive/ecol/E094/245/>)

Encadrement

Marie-Josée Cros et Nathalie Peyrard
(marie-josee.cros@inra.fr, nathalie.peyrard@inra.fr)
Equipe Modélisation des Agro-écosystèmes et Décision (MAD),
Unité de Mathématiques et Informatique Appliquées, INRA Toulouse.
Web : <https://mia.toulouse.inra.fr/Accueil>