

Réseau bayésien dynamique étiqueté : cadre et apprentissage de structure pour application aux réseaux écologiques

Étienne Auclair

Thèse encadrée par : Nathalie Peyrard, Régis Sabbadin



24 Janvier 2019

1/49

Contexte et objectif

Réseau bayésien

Réseau Bayésien Étiqueté

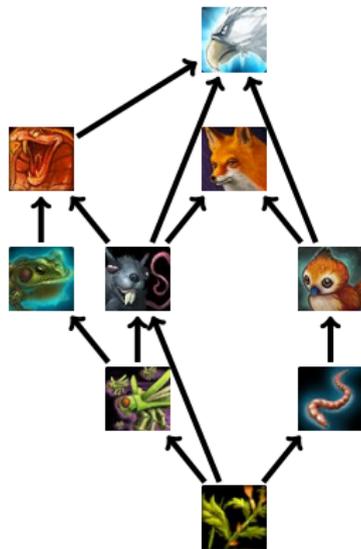
Apprentissage de structure

Cas d'étude

Conclusion

Réseau écologique

- ▶ Relations proie/prédateur : *Réseau trophique*
- ▶ Connaissance d'un milieu, gestion de la biodiversité, impact des espèces invasives...
- ▶ D'autres interactions existent (parasitisme, compétition...) :
Réseau écologique
- ▶ Difficile à identifier
- ▶ Enjeu : Reconstruire un réseau écologique mal connu



Apprendre un réseau écologique

Observation directe

- ▶ Observation sur le terrain, recherche de marqueurs trophiques
- ▶ Coûteux, échantillons difficiles à obtenir, nombre d'interactions très grand

Inférence

- ▶ Indicateurs biologiques (abondances, présence d'espèces...)
- ▶ Reconstruction de réseau par modèles mathématiques
- ▶ Peu coûteux, moins fiable

Objectif

Contraintes

- ▶ Peu de données / Données peu informatives :
présence/absence d'espèces au cours du temps
- ▶ Ajout de connaissances expertes
 - ▶ Interactions connues ou probables
 - ▶ Hypothèses sur la forme du modèle

Objectif

Développer une méthode permettant d'apprendre la structure d'un réseau écologique à partir de données temporelles de présence/absence.

Cadre méthodologique

Techniques d'inférence de réseaux écologiques

- ▶ Programmation logique¹
- ▶ Régression linéaire²
- ▶ Modèle Graphique Gaussien³
- ▶ Réseaux bayésiens⁴

Choix du cadre : **Réseaux bayésiens**

- ▶ Facilité de manipulation de données binaires
- ▶ Il existe un cadre dynamique

¹Tamaddoni-Nezhad et al. 2013.

²Oevelen et al. 2010.

³Kurtz et al. 2015.

⁴Faisal et al. 2010.

Contexte et objectif

Réseau bayésien

Réseau Bayésien Étiqueté

Apprentissage de structure

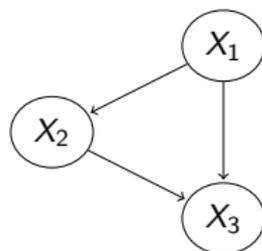
Cas d'étude

Conclusion

Réseau bayésien

Définition

- ▶ Ensemble de variables aléatoires discrètes
- ▶ Représentation : graphe orienté sans circuit
- ▶ Table de probabilités conditionnelles (TCP) entre chaque variable et ses parents



Probabilité jointe

$X_i (i \in 1, \dots, n)$: Variable aléatoire

π_i : Parents de i

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | X_{\pi_i})$$

Utilisations

Modélisation, simulations, estimation de TPC ou inférence de structure

Inférence de la structure d'un réseau bayésien

Principe

- ▶ Données : réalisation de variables aléatoires
- ▶ Apprentissage du graphe correspondant : recherche des meilleurs parents de chaque variable

Méthodes

- ▶ Tests d'indépendance conditionnelle
- ▶ Apprentissage par score

Inférence de réseaux bayésiens par tests d'indépendance conditionnelle

Idée générale

- ▶ Tests d'indépendance du χ^2
- ▶ Recherche de toutes les indépendances conditionnelles

Principes algorithmiques

- ▶ Recherche de structures particulières
 - ▶ V-structure⁵ $i \rightarrow j \leftarrow k$
 - ▶ Couverture de Markov⁶
- ▶ Orientation du graphe en respectant ces structures

⁵Spirtes, Glymour, and Scheines 2000.

⁶Yu et al. 2013.

Inférence de structure de réseaux bayésiens par score

Idée générale

- ▶ Score : évaluation d'un graphe par rapport aux données
- ▶ Recherche du "meilleur" graphe \mathcal{G}

Propriétés d'un score

- ▶ Score décomposable par variable

$$\log P(X|\mathcal{G}, \theta) = \sum_{i=1}^n \log(P(X_i|X_{\pi_i}))$$

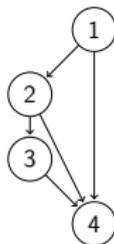
- ▶ Principe de parcimonie : pénalisation de structures complexes
 - ▶ Score séparable en évaluation + pénalisation

$$BIC(x|\mathcal{G}, \theta) = \log P(x|\mathcal{G}, \theta) - \frac{\log n \mathcal{B}}{2}$$

Inférence de structure de réseaux bayésiens par score

Principes algorithmiques

- ▶ Recherche par score des parents de chaque variable
- ▶ Réduction de l'espace de recherche
 - ▶ Par élagage, en utilisant le paramètre de pénalisation⁷
 - ▶ Par ordonnancement des variables⁸

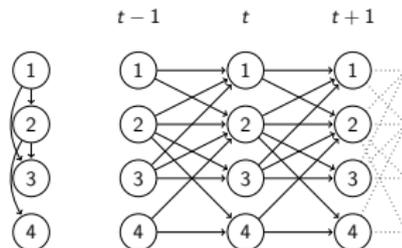


⁷Campos and Ji 2011.

⁸Silander and Myllymaki 2012.

Réseaux bayésiens dynamiques

- ▶ Phénomène dynamique
 X^0, X^1, \dots, X^T
- ▶ Structure initiale + structure de transition
- ▶ Phénomène markovien : L'état d'une variable à t ne dépend que de ses parents à $t - 1$ ou à t
- ▶ Stationnarité : même structure pour tout t



Résumé

Objectif

Apprendre la structure d'un réseau écologique à partir de données dynamique de présence/absence

Choix

- ▶ Réseaux bayésiens dynamiques : cadre efficace pour ces données
- ▶ Méthode par score : utilisation du score pour optimisation multiobjectif

Démarche

- ▶ Définition d'un cadre général
- ▶ Développement d'un algorithme d'apprentissage

Contexte et objectif

Réseau bayésien

Réseau Bayésien Étiqueté

Apprentissage de structure

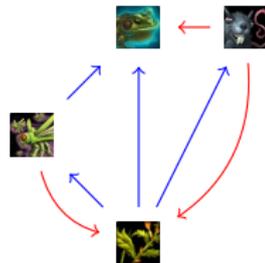
Cas d'étude

Conclusion

Hypothèses écologiques pour la modélisation

Hypothèses

- ▶ Distinction entre influences positives et négatives
- ▶ Modèle de dynamique : "Une espèce survit si elle réussit à se nourrir, et qu'elle ne se fait pas manger"
- ▶ La survie d'une espèce dépend du nombre de ses proies/prédateurs



Modèle

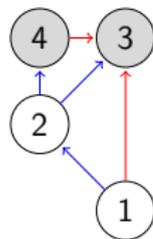
- ▶ Contraintes sur les probabilités conditionnelles⁹
- ▶ Forme particulière des contraintes

⁹Niculescu, Mitchell, and Rao 2006; Oniśko, Druzdzel, and Wasyluk 2001

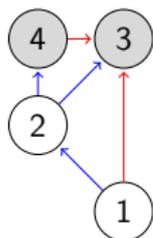
Réseau bayésien étiqueté

Idée générale

- ▶ Chaque arc est étiqueté
- ▶ Voisinage similaire → même probabilité
- ▶ Distribution de probabilité conditionnelle : fonction d'un vecteur de paramètres θ .



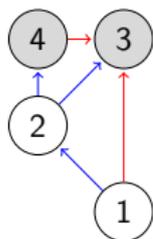
RBE : modèle de base



- ▶ $X_i \in \{0, 1\}$: Variable aléatoire associée au nœud i . i est dit "présent" si $X_i = 1$
- ▶ ρ : proba de réussite d'un impulseur (influence positive)
- ▶ τ : proba de réussite d'un inhibiteur (influence négative)
- ▶ π_i : Parents de i et leurs étiquettes.
- ▶ $\overset{l}{N}_i$: Nombre de parents de i d'étiquette $l \in \{q, r\}$ présents.

$$P(X_i = 1 | X_{\pi_i}) = \left(1 - (1 - \rho)^{\overset{q}{N}_i} \right) \cdot (1 - \tau)^{\overset{r}{N}_i}$$

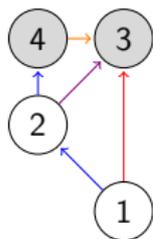
RBE : modèle de base



- ε : proba de présence indépendante des valeurs des autres variables

$$P(X_i = 1 | X_{\pi_i}) = \left(\varepsilon + (1 - \varepsilon) \cdot \left(1 - (1 - \rho)^{N_i^q} \right) \right) \cdot (1 - \tau)^{N_i^r}$$

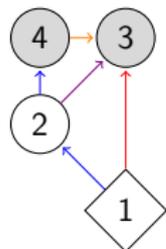
RBE : étiquettes multiples



- ▶ q_1, q_2 : étiquettes d'impulseurs.
- ▶ r_1, r_2 : étiquettes d'inhibiteurs.

$$P(X_i = 1 | X_{\pi_i}) = \left(\varepsilon + (1 - \varepsilon) \cdot \prod_{q=q_1}^{q_{max}} \left(1 - (1 - \rho_q)^{N_i^q} \right) \right) \cdot \prod_{r=r_1}^{r_{max}} (1 - \tau_r)^{N_i^r}$$

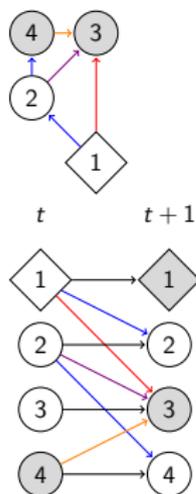
RBE : probabilités indépendantes multiples



- ▶ u_i : comportement inhérent à i .
- ▶ ε_{u_i} : Probabilité indépendante associée à u_i .

$$P(X_i = 1 | X_{\pi_i}, u_i) = \left(\varepsilon_{u_i} + (1 - \varepsilon_{u_i}) \cdot \prod_{q=q_1}^{q_{max}} \left(1 - (1 - \rho_q)^{N_i^q} \right) \right) \cdot \prod_{r=r_1}^{r_{max}} (1 - \tau_r)^{N_i^r}$$

RBDE : Réseau Bayésien Dynamique étiqueté



- ▶ Modèle dynamique, markovien, stationnaire.
- ▶ Différentiation entre apparition et survie.
- ▶ Impact différent selon apparition ou survie : paramètres doublés

$$P(X_i^{t+1} = 1 | X_i^t, X_j^t \forall j \in \pi_i)$$

RBDE : Probabilités de transition

Apparition

$$P(X_i^{t+1} = 1 | X_i^t = 0, X_j^t \forall j \in \pi_i) = \left(\varepsilon_{u_i}^{app} + (1 - \varepsilon_{u_i}^{app}) \cdot \left(1 - \prod_{q=q_1}^{q_{max}} (1 - \rho_q^{app})^{N_i^t} \right) \right) \cdot \prod_{r=r_1}^{r_{max}} (1 - \tau_r^{app})^{N_i^t}$$

Survie

$$P(X_i^{t+1} = 1 | X_i^t = 1, X_j^t \forall j \in \pi_i) = \left(\varepsilon_{u_i}^{sur} + (1 - \varepsilon_{u_i}^{sur}) \cdot \left(1 - \prod_{q=q_1}^{q_{max}} (1 - \rho_q^{sur})^{N_i^t} \right) \right) \cdot \prod_{r=r_1}^{r_{max}} (1 - \tau_r^{sur})^{N_i^t}$$

Modèle RBDE pour réseau écologique

Apparition

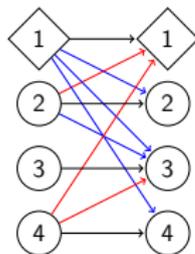
Spontanée : aucun impact des interactions

$$P(X_i^{t+1} = 1 | X_i^t = 0) = \varepsilon^{app}$$

Survie

Facilitateurs ou proies nécessaires pour survivre (sauf pour les espèces basales)

$$P(X_i^{t+1} = 1 | X_i^t = 1, u_i, x_j^t \forall j \in \pi_i) = \left(\varepsilon_{u_i} + (1 - \varepsilon_{u_i}) \cdot \left(1 - (1 - \rho)^{N_i^t} \right) \right) \cdot (1 - \tau)^{N_i^t}$$



Bilan

Réseau bayésien dynamique étiqueté

- ▶ Réseau bayésien dynamique avec graphe étiqueté
- ▶ Une étiquette caractérise un arc : influence positive ou négative
- ▶ Complexification : plusieurs étiquettes, étiquettes de nœuds

Caractéristiques

- ▶ Applicable à plusieurs modèles connus
 - ▶ Modèles de propagation par contact
- ▶ Nombre de paramètres indépendant de la structure
- ▶ Peu de paramètres : facilite l'apprentissage

Contexte et objectif

Réseau bayésien

Réseau Bayésien Étiqueté

Apprentissage de structure

Cas d'étude

Conclusion

Apprendre un RBDE

Ce qu'il faut apprendre

- ▶ Arcs
- ▶ Étiquettes
- ▶ Étiquettes de nœuds (probabilités indépendantes)
- ▶ Paramètres

Différences avec un réseau bayésien dynamique classique

- ▶ Probabilités conditionnelles comme fonction de θ : pas de table de probabilité conditionnelles
- ▶ Nombre de paramètres indépendant de la structure : pas de pénalité nécessaire
- ▶ Dérivée de l'expression de la vraisemblance non explicite

Apprentissage de structure de RBDE par algorithme "Estimation-Restauration"

Apprentissage par score

- ▶ Score : vraisemblance
- ▶ Algorithme glouton ER¹⁰
 - ▶ Étape 1 (Estimation) : Estimation des paramètres par maximum de vraisemblance à structure de graphe connue
 - ▶ Étape 2 (Restauration) : Apprentissage de la structure de grahe maximisant la vraisemblance à paramètres connus
 - ▶ Retour à l'étape 1 jusqu'à convergence
- ▶ Obtention d'un maximum local

¹⁰Auclair, Peyrard, and Sabbadin 2017.

Étape d'estimation

- ▶ Structure de graphe connue
- ▶ Estimation des paramètres par maximum de vraisemblance
- ▶ $\theta = \{\varepsilon_1, \dots, \varepsilon_{u_{max}}, \rho_1, \dots, \rho_{q_{max}}, \tau_1, \dots, \tau_{r_{max}}\}$
- ▶ Nombre de paramètres limité : étape peu critique

Probabilités de transition

$$P(X_i^{t+1} | X_i^t = 0) = \left(\varepsilon_{u_i}^{app} + (1 - \varepsilon_{u_i}^{app}) \cdot \left(1 - \prod_{q=q_1}^{q_{max}} (1 - \rho_q^{app})^{N_i^t} \right) \right) \cdot \prod_{r=r_1}^{r_{max}} (1 - \tau_r^{app})^{N_i^t}$$

$$P(X_i^{t+1} | X_i^t = 1) = \left(\varepsilon_{u_i}^{sur} + (1 - \varepsilon_{u_i}^{sur}) \cdot \left(1 - \prod_{q=q_1}^{q_{max}} (1 - \rho_q^{sur})^{N_i^t} \right) \right) \cdot \prod_{r=r_1}^{r_{max}} (1 - \tau_r^{sur})^{N_i^t}$$

Étape de restauration

Programmation linéaire en nombres entiers 0/1 (PLNE)

- ▶ Optimisation d'une fonction linéaire
 $f(X) = c_1X_1 + c_2X_2 + \dots + c_nX_n$
- ▶ Matrice A, vecteur B : contraintes linéaires
- ▶ Vecteur c : fonction objectif.

$$\begin{cases} \min c^T x \\ Ax \leq B \end{cases}$$

Utilisation dans l'étape R

- ▶ Vraisemblance comme fonction linéaire de la présence des arcs du graphe G_{ij}^l
- ▶ Optimisation de cette fonction \rightarrow graphe appris

Étape de restauration

Programmation linéaire en nombres entiers 0/1 (PLNE)

- ▶ Linéarisation du problème : ajout de variables binaires définies par des contraintes linéaires sur les variables G_{ij}^l

- ▶ $n \cdot L + U + T \left(L \cdot k + \sum_{h=0}^k C_{h+L-1}^L \right)$ variables pour les n problèmes (avec n nœuds, T pas de temps, L étiquettes différentes, U comportements indépendants différents et k parents au maximum)

$$\begin{array}{l} n = 5 \\ T = 5 \\ k = 5 \rightarrow n \times 332 \\ L = 2 \\ U = 1 \end{array}$$

$$\begin{array}{l} n = 6 \\ T = 5 \\ k = 5 \rightarrow n \times 334 \\ L = 2 \\ U = 1 \end{array}$$

$$\begin{array}{l} n = 5 \\ T = 5 \\ k = 6 \rightarrow n \times 403 \\ L = 2 \\ U = 1 \end{array}$$

$$\begin{array}{l} n = 5 \\ T = 5 \\ k = 5 \rightarrow n \times 620 \\ L = 3 \\ U = 1 \end{array}$$

Ajout de connaissances expertes

Connaissances sous forme de contraintes linéaires

- ▶ Contraintes de structure : un seul arc par paire de nœuds, aucune boucle
- ▶ Connaissances "dures" : arcs ou absences d'arc connues
- ▶ Structures particulières : degré maximum, type de probabilités indépendantes...

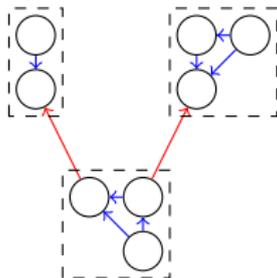
Autres formes de connaissances

- ▶ Connaissance sur les paramètres (nombre de paramètres, paramètres égaux, paramètres connus)
- ▶ Prior sur certains arcs modifiant les coefficients dans la fonction linéaire

Modèle à bloc stochastique (SBM) pour un prior sur la structure du graphe

Définition

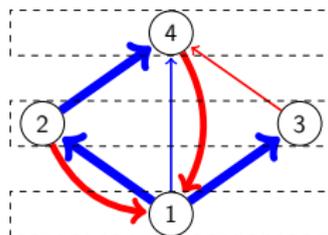
- ▶ Des "blocs" dans le graphe sont connus (communautés)
- ▶ La probabilité de présence d'un arc étiqueté $i \rightarrow j$ est fonction de son étiquette et des blocs de i et de j .
- ▶ Probabilités décrites par un ensemble de paramètres



Exemple écologique : SBM pour un prior sur les niveaux trophiques

Hypothèses sachant les niveaux trophiques :

- ▶ Aucun arc + vers un niveau trophique inférieur
- ▶ Arcs + plus probables vers des niveaux trophiques proches
- ▶ Arcs - plus probables vers les niveaux trophiques inférieurs



Prior SBM comme connaissance dans la fonction objectif

Changements dans l'algorithme ER

- ▶ Étape E : estimation de paramètres spécifiques au SBM
 - ▶ Étape peu complexe
 - ▶ Quelques paramètres en plus à estimer
- ▶ Étape R : Prior SBM comme coefficients de la fonction objectif
 - ▶ Pas d'ajout de variables ni de contraintes
 - ▶ Complexifie peu cette étape

Évaluation de l'algorithme E-R

Expériences

- ▶ Estimation des paramètres à graphe connu (Étape E)
- ▶ Apprentissage de structure à paramètres connu (Étape R)
- ▶ Apprentissage de structure par algorithme E-R
- ▶ Apprentissage de structure par algorithme E-R avec prior SBM
 - ▶ Comparaison avec algorithme E-R sans prior
 - ▶ Comparaison avec algorithme de la littérature¹¹ : étiquetage à postériori¹²

¹¹Vinh et al. 2011.

¹²Wellman 1990.

Résultats de l'évaluation

Résultats

- ▶ Algorithme de la littérature peu efficace pour apprendre RBDE
- ▶ Estimation des paramètres proche des paramètres réels
- ▶ Apprentissage de structure difficile
 - ▶ Peu d'interactions "réelles" retrouvées
 - ▶ L'agrégation de graphes améliore l'apprentissage
- ▶ Prior SBM efficace : apprentissage plus précis

Expérience supplémentaire

- ▶ Pas d'agrégation de graphe : maximisation de vraisemblance globale sur plusieurs jeux de données
- ▶ 15 graphes synthétiques, paramètres aléatoires, 10 simulations par graphe. $n = 20$, $T = 16$

Résultats sur réseaux synthétiques

Différences entre paramètres réels et estimés

Expérimentation	Moyenne	Écart-type	Médiane
Pas de prior	0,105	0,17	0,021
PriorSBM	0,090	0,17	0,016

Différences entre réseaux

Moyenne des rappels et précisions

Expérimentation	Impulseurs		Inhibiteurs	
	Précision	Rappel	Précision	Rappel
Pas de prior	43,63%	23,91%	32,49%	4,08%
Prior SBM	53,12%	35,81%	47,49%	15,86%

- ▶ Apprentissage amélioré par l'ajout de connaissances

Contexte et objectif

Réseau bayésien

Réseau Bayésien Étiqueté

Apprentissage de structure

Cas d'étude

Conclusion

Cas d'étude : échantillons d'arthropodes

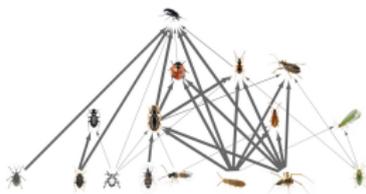


Figure: Bohan, Vacher, et al. 2017

Contexte

Arthropodes piégés dans des champs expérimentaux¹³

- ▶ 2 dates d'échantillonnage : avant récolte et après récolte
- ▶ Données d'abondance
- ▶ Apprentissage de réseau trophique par programmation logique

¹³Bohan, Caron-Lormier, et al. 2011.

Données "Arthropodes"

Description des données

- ▶ Espèces d'arthropodes piégés dans différentes parcelles de 4 cultures différentes

Culture	Betterave	Maïs	Colza d'hiver	Colza d'été
Espèces	41	29	40	29
Parcelles	66	59	67	65

Questions

- ▶ Les relations écologiques sont-elles les mêmes pour toutes les cultures ?
- ▶ Apprend-on les mêmes interactions par RBDE et par programmation logique ?

Apprentissage de structure de réseau écologique "Arthropodes"

Méthode

- ▶ Conversion abondance en présence/absence
- ▶ Apprentissage par algorithme E-R
- ▶ Niveau trophique représenté par la taille (5 classes de taille)

Réseaux appris

- ▶ Un réseau par culture
- ▶ Un réseau "global" sans distinction de culture : 57 espèces

Les relations écologiques sont elles les mêmes pour toutes les cultures ?

- ▶ Extrait de graphe : 20 espèces communes à toutes les cultures
- ▶ Nombre d'arcs en commun entre chaque réseau

Même étiquette (Étiquettes différentes)

Graphe	B	M	SR	WR	Global
B	79 (0)	10 (14)	12 (8)	7 (13)	52 (0)
M	10 (14)	46 (0)	8 (3)	10 (7)	12 (24)
SR	12 (8)	8 (3)	29 (0)	9 (1)	13 (10)
WR	7 (13)	10 (7)	9 (1)	55 (0)	9 (21)
Global	52 (0)	12 (24)	13 (10)	9 (21)	83 (0)

- ▶ Betterave peu différent du réseau global
- ▶ Beaucoup de différences sur les autres réseaux

Bilan

- ▶ Chaque culture a son propre réseau
- ▶ Peu d'interactions communes avec programmation logique
- ▶ L'apprentissage par programmation logique :
 - ▶ Ne contient que des relations trophiques
 - ▶ Utilise des données d'abondances
 - ▶ Prend en compte d'autres informations (régime alimentaire, groupe fonctionnel des espèces...)
 - ▶ N'utilise pas la dynamique
- ▶ RBDE serait un graphe "théorique", renseignant d'interactions probables

Contexte et objectif

Réseau bayésien

Réseau Bayésien Étiqueté

Apprentissage de structure

Cas d'étude

Conclusion

Résumé général

Objectif

- ▶ Apprentissage de réseau écologique
- ▶ Données temporelles de présence/absence
- ▶ Ajout de connaissances expertes

Contributions

- ▶ Cadre général : réseau bayésien dynamique étiqueté
- ▶ Inférence : Algorithme "Estimation-Restauration" pour l'apprentissage
- ▶ Évaluation sur donnée simulées, puis réelles

Modèle : Réseau bayésien dynamique étiqueté

Bilan

- ▶ Modélise plusieurs types de phénomènes
- ▶ Restriction du nombre de paramètres
- ▶ Connaissance experte sous forme de paramètres

Perspectives

- ▶ Confirmer l'utilité du modèle : utilisation dans d'autres domaines
- ▶ Ajout de types d'étiquettes autre de négatif/positif
- ▶ Enlever la restriction aux variables binaires

Apprentissage de structure

Bilan

- ▶ Connaissance experte facilement intégrable
- ▶ Graphe obtenu à interpréter à posteriori
- ▶ Étape de restauration complexe à calculer

Perspectives

- ▶ Réduire la complexité de l'algorithme (ex : relaxation ILP)
- ▶ Score avec pénalisé par la structure du graphe
- ▶ Ajout de connaissances dans le processus d'apprentissage
- ▶ Application à la gestion adaptative : compromis entre qualité d'apprentissage et gestion

Références

Réseaux écologiques



Ali Faisal et al. "Inferring species interaction networks from species abundance data: A comparative evaluation of various statistical and machine learning methods". In: *Ecological Informatics* 5.6 (2010), pp. 451–464



Nabil Majdi et al. "There's no harm in having too much: A comprehensive toolbox of methods in trophic ecology". In: *Elsevier - Food Webs Journal* (Sept. 2018), e00100. DOI: 10.1016/j.foodweb.2018.e00100

Apprentissage de réseaux bayésiens



Norbert Dojer. "Learning Bayesian networks does not have to be NP-hard". In: *International Symposium on Mathematical Foundations of Computer Science*. Springer. 2006, pp. 305–314



Mark Bartlett and James Cussens. "Integer linear programming for the Bayesian network structure learning problem". In: *Artificial Intelligence* 244 (2017), pp. 258–271



Etienne Auclair, Nathalie Peyrard, and Régis Sabbadin. "Labeled DBN learning with community structure knowledge". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2017, pp. 158–174

Cas d'étude "arthropodes"



David A Bohan, Corinne Vacher, et al. "Next-generation global biomonitoring: large-scale, automated reconstruction of ecological networks". In: *Trends in ecology & evolution* 32.7 (2017), pp. 477–487