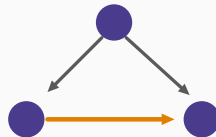


Inférence causale et machine learning

Chapitre 3 : Graphes causaux et variables à ne pas contrôler



Écrire ses hypothèses

Les trois configurations élémentaires

Les erreurs qui en découlent

Le critère de sélection

Ce que cela implique pour le machine learning

Ce qu'il faut retenir

Écrire ses hypothèses

Graphe orienté acyclique

Un **graphe causal** représente les variables par des nœuds et les effets directs par des flèches orientées. Aucun cycle n'est permis : rien ne peut être sa propre cause.

Ce que le graphe force à faire

Le dessiner oblige à **expliquer** ce qu'on croit du monde, avant de toucher aux données.

Il ne se déduit pas des données : deux graphes différents peuvent produire les mêmes corrélations. C'est un objet d'hypothèses — et c'est précisément sa valeur, puisqu'il rend ces hypothèses discutables.

Les trois configurations élémentaires

La chaîne

$$T \rightarrow M \rightarrow Y$$

M est un **médiateur** : le traitement agit à travers lui.

Le contrôler bloque une partie de l'effet qu'on cherche à mesurer.

La fourche

$$T \leftarrow X \rightarrow Y$$

X est un **confondant** : il cause les deux, et crée une association non causale entre eux.

Il faut le contrôler.

Définition

$$T \rightarrow C \leftarrow Y$$

C est un **collisionneur** : il est causé par les deux.

Le contrôler crée une association là où il n'y en avait aucune. C'est le résultat le plus contre-intuitif du chapitre.

Exemple

Deux causes indépendantes d'être **embauché** : avoir suivi la formation, et avoir un réseau personnel.

Parmi les seuls embauchés, celui qui n'a pas suivi la formation avait probablement un bon réseau — sinon il n'aurait pas été recruté.

En restreignant l'analyse aux embauchés, on crée une corrélation négative entre formation et réseau, alors qu'elles étaient indépendantes dans la population.

Les erreurs qui en découlent

Trois biais nommés

Le biais du médiateur

Contrôler une variable située **sur** le chemin causal retire une partie de l'effet.

Contrôler le salaire pour estimer l'effet de la formation sur la satisfaction, alors que la formation agit précisément **par** le salaire, sous-estime l'effet total.

Le biais de sélection

Conditionner sur un collisionneur — ou sur l'un de ses descendants — introduit une association artificielle.

C'est le mécanisme de nombreux biais d'échantillonnage : n'observer que les entreprises **survivantes** , que les patients **hospitalisés** , que les candidats **retenus** .

Un cas classique en finance

Le **biais du survivant** du chapitre 11 d'*Apprentissage profond* est exactement un collisionneur.

Rester coté dépend à la fois de la performance passée et de la stratégie de l'entreprise. Analyser les seules entreprises encore cotées, c'est conditionner sur cette survie — et fabriquer des relations qui n'existent pas dans la population complète.

Le critère de sélection

Comment choisir l'ensemble de contrôles

Le critère de la porte arrière

Un ensemble X identifie l'effet de T sur Y si :

1. il **bloque** tous les chemins de T vers Y qui commencent par une flèche **entrant** dans T ;
2. il ne contient **aucun descendant** de T .

La lecture, en une phrase

Il faut fermer les chemins par lesquels une **cause commune** fait passer de l'information, sans toucher aux chemins qui portent l'**effet** lui-même.

Contrôler n'est donc pas « ajouter des variables » : c'est **fermer les bonnes portes et laisser les autres ouvertes**.

Ce que cela change au chapitre 1

Le paradoxe de Simpson se résout ici. La qualification cause à la fois la participation à la formation et l'emploi : c'est une **fourche**, donc un confondant.

Il fallait la contrôler — et les taux par groupe donnaient la bonne réponse, tandis que le total donnait la mauvaise.

Le graphe dit pourquoi : là où l'intuition seule hésitait

Ce que cela implique pour le
machine learning

Une mise en garde décisive

Pourquoi on ne peut pas tout donner à l'algorithme

Le réflexe du praticien de l'apprentissage statistique est de fournir toutes les variables disponibles et de laisser la méthode trier.

En prévision, c'est raisonnable. **En causalité, c'est dangereux** : l'algorithme retiendra volontiers un médiateur ou un collisionneur, parce que tous deux améliorent la prévision.

Un lasso qui sélectionne des variables pour prévoir **ne sélectionne pas** un ensemble de contrôles valide.

Ce que le chapitre 10 pourra, et ne pourra pas

Le double machine learning utilisera des méthodes d'apprentissage pour estimer des fonctions de nuisance de grande dimension. C'est un progrès considérable.

Il suppose néanmoins que **l'ensemble des contrôles a été choisi correctement**, sur des bases théoriques. L'algorithme gère la dimension; il ne décide pas de l'identification.

Aucune méthode ne remplace le graphe.

Ce qu'il faut retenir

Cinq points

1. Le **graphe causal** explicite les hypothèses; il ne se déduit pas des données.
2. Trois motifs : **chaîne** (médiateur), **fourche** (confondant), **collisionneur**.
3. Contrôler un **confondant** est nécessaire; contrôler un **médiateur** ou un **collisionneur** nuit.
4. Règle sûre : ne contrôler que des variables **antérieures** au traitement.
5. Le **machine learning ne choisit pas** les contrôles : il gère la dimension, pas l'identification.

La suite

Il existe une situation où tous ces problèmes disparaissent d'un coup. Le chapitre 4 l'examine — et elle sert de référence à toutes les méthodes suivantes.