

Méthodes formelles et apprentissage ?

Gilles Dowek (outsider)

Outsider ?

Deducteam : une équipe de méthodes formelles

Notre cœur de métier : l'échange de preuves entre systèmes

Exemple : Bibliothèque arithmétique

- ▶ 340 lemmes dont le petit théorème de Fermat
- ▶ 1.5 Mo
- ▶ exportée dans cinq systèmes différents : Coq, Matita, HOL Light, Isabelle / HOL, HOL4

Intérêt pour l'apprentissage : Éducation, Cerna

Comme l'huile et l'eau ?

Méthodes formelles : démontrer qu'un algorithme vérifie certaines propriétés

Apprentissage :

- ▶ on ne sait pas (et on ne veut pas savoir) comment fonctionnent les algorithmes
- ▶ on ne sait pas spécifier ce qu'ils devraient faire (reconnaître un Lorenzetti d'un Basquiat)
- ▶ ce qu'ils font change avec le temps
- ▶ qu'il le fassent bien ou non dépend aussi de la qualité des données utilisées pour apprendre



C'est peut-être mieux comme ça

Pourtant...

Les domaines d'application sont les mêmes

Apprentissage utilisé dans les **domaines critiques** (Transports, Santé, Énergie)

Véhicules autonomes, analyse d'images médicales... : bug = mort

De plus...

L'importance des **explications**

Puis-je obtenir un prêt à la consommation ? Non

Suis-je admis dans le master xyz ? Non

Puis-je obtenir un prêt à la consommation ?

Non, parce que vos revenus ne vous permettraient pas de rembourser ce prêt

Non, parce que vous êtes blanc, anglophone et protestant

Qu'est-ce qu'une explication ?

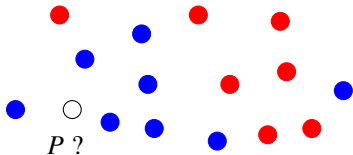
Une démonstration

Qu'est-ce qui distingue une bonne d'une mauvaise explication ? Les axiomes utilisés dans cette démonstration

Comment commencer ?

Petit : démontrer la correction d'un algorithme d'apprentissage sans se préoccuper de la qualité des données

Exemple : apprentissage supervisé



$$f(P) = \text{bleu} \Leftrightarrow \frac{\sum_i d(P, B_i)}{b} < \frac{\sum_i d(P, R_i)}{r}$$

Idem

Algorithme de clusterisation

Spécification : pas de meilleure clusterisation

Un peu plus ambitieux (et sans doute la question intéressante)

Les quatre concepts de l'informatique : algorithme, machine, langage, information (donnée)

Nous savons spécifier et prouver

- ▶ des algorithmes (recherche en table, tri...)
- ▶ des machines (circuits : addition, multiplication...)
- ▶ des langages (syntaxe, sémantique, compilation...)
- ▶ des données ?

Que signifie : spécifier et prouver des données ?

- ▶ Partir d'exemples de "bug" (Tay, le lion et la savane...)
- ▶ Comprendre ce qui ne va pas dans ces données (fréquence élevée de certains mots ? présence de mots déplacés ? manque de diversité des images ?)
- ▶ Comprendre en quoi ces propriétés impactent l'algorithme d'apprentissage utilisé
- ▶ Comprendre si on peut formuler ces propriétés mathématiquement (dans un langage logique)
- ▶ Comprendre si on peut établir ces propriétés automatiquement ou semi-automatiquement
- ▶ ...

Pour conclure

Trop peu de réflexion sur les liens entre méthodes formelles et apprentissage

Réflexions (très) préliminaires

Un sujet de recherche à créer

Commencer petit (exemples académiques)

Mais inspirés de vrais problèmes rencontrés par les algorithmes d'apprentissage