Biais et conformismes des traitements algorithmiques

François Pellegrini

université de BORDEAUX

Biais et conformismes des traitements algorithmiques

François Pellegrini
Professeur, Université de Bordeaux
francois.pellegrini@u-bordeaux.fr

Algorithmes et traitements (1)

- Le terme « algorithme » est majoritairement utilisé de façon inappropriée
  - Victime d'un regrettable effet de mode
- Un algorithme est la description d'une suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée
  - Recette de cuisine
  - Archétype d'histoire
  - Méthode mathématique

Algorithmes et traitements (2)

- Confusion entre trois objets techniquement et juridiquement distincts :
  - Algorithme
    - Objet mathématique de libre parcours
    - Ni « loyal », ni « éthique »
      - Mais tout projet scientifique suscite des questions éthiques
  - Logiciel
    - Création de forme exprimant un ou plusieurs algorithmes
    - Œuvre de l'esprit soumis au droit d'auteur adapté
  - Traitement de données
    - Mis en œuvre par un « responsable de traitement »

« Code is law »

- Un logiciel et ses algorithmes sous-jacents, comme tout artefact, s'inscrivent dans un environnement social, économique et culturel
  - Incorporent des biais humains
- « Code is law »
  - On ne peut faire que ce que la machine permet que l'on fasse

Principes des mégadonnées (1)

- Modélisation d'un système ouvert et non pas fermé
  - Impossibilité d'obtenir des certitudes
  - Il ne s'agit plus de modéliser, mais d'abstraire
- Raisonnement inductif et non pas déductif
  - Déduction : s'appuie sur une règle préétablie
    - Je suis vivant donc je suis né
  - Induction : phénomène généralement observé
    - On suppose que je vais mourir

Principes des mégadonnées (2)

- Conserver plus de données n'apporte pas plus de précision
  - Obsolescence des données comportementales
  - Il faut être capable d'oublier pour continuer à agir

Principes des mégadonnées (3)

- En statistique, un modèle est efficace si le résidu est un bruit blanc (aléatoire)
  - Permet de traiter les cas fréquents
  - Ne permet pas de capturer les cas différenciateurs
- « Les mégadonnées commencent là où la statistique s'arrête »
  - Détection de « signaux faibles », de « nouvelles tendances »
  - Personnalisation de l'information

Principes des mégadonnées (4)

- Les mégadonnées s'appuient sur :
  - L'abstraction : remplacer des informations par des informations plus compactes
  - La connexité : existence de liens entre les données
    - « Inter-legere »

Personnalisation

- Prise en compte des individualités
  - Modèles différenciateurs captant les « signaux faibles »
  - Prévoir comment chacun va raisonner par rapport à son référentiel de pensée
    - Rationalité et biais cognitifs
- Mieux connaître un individu, c'est :
  - Mieux le servir
  - Jouer sur ses sensibilités
  - Anticiper ses réactions

Mégadonnées et gouvernance (1)

- Les sociétés humaines sont déjà gérées selon les principes des mégadonnées
  - Les êtres humains sont capables de décider par induction à partir de données incomplètes
- La technologie se rapproche de plus en plus du modèle humain, mais peut-elle l'assister de façon utile ?
  - Au niveau individuel : service à l'usager
  - Au niveau collectif : influence sur la définition et la mise en œuvre des politiques publiques

Mégadonnées et gouvernance (2)

- Parmi les domaines dans lesquels on pense que les mégadonnées auront un fort bénéfice à s'appliquer, on identifie plusieurs secteurs stratégiques dont :
  - Santé
  - Régalien (police-justice)

Intelligence artificielle

- Projet scientifique issu de la conférence de Bournemouth de 1956
  - IA « forte » : intelligence synthétique généraliste
  - IA « faible » : assistance à des tâches spécialisées
- Actuellement, seules des IA « faibles » sont mises en œuvre
  - Traitements opérant dans un contexte borné en termes de données fournies et de réponses attendues
  - Bien loin de la « singularité »

Traitements auto-apprenants (1)

- L'apprentissage « profond » (« *deep learning* ») est une amélioration de l'apprentissage par réseaux de neurones
  - Structuration en couches des neurones
- Permet l'extraction de caractéristiques de plus en plus « abstraites »
  - Jusqu'à « capturer » les éléments stylistiques d'un tableau pour les transposer dans un autre

Traitements auto-apprenants (2)

- Problématique de la détection des biais
  - Filtrage lors de la modélisation des jeux de données
    - « M. / Mme » dans les réponses
  - Sélection des jeux de données
    - « Tous les Anatole sont des tueurs en série »
  - Convergence de l'algorithme
    - Cas de la détection des rideaux et non du lit dans les chambres à coucher
- Problématique du re-jeu
  - Preuve a posteriori d'existence d'un biais ?

Le cas PredPol (1)

- Logiciel permettant de « faire baisser la criminalité » dans les lieux où il est mis en œuvre
- Tentative de reproduction de l'expérience
  - Par Ismaël Benslimane, Master Physique U. Grenoble et membre du CorteX (E&R en esprit critique)

Le cas PredPol (2)

- Faits :
  - Biais méthodologiques et d'usage
    - Logiciel servant à la prédiction tant qu'à la déclaration
    - Logiciel vu comme bénéfique que le crime ait eu lieu ou pas
  - Retrouve une évidence statistique : loi de Pareto
    - 80 % des délits sont concentrés sur 20 % du territoire
    - Il suffit de prédire toujours les mêmes lieux « à risque » pour être aussi performant que PredPol
  - Retrouve une évidence « métier » :
    - Le crime est contagieux spatio-temporellement
    - Une victime non protégée se fera cambrioler à nouveau

Le cas PredPol (3)

- Analyse :
  - Expose des biais
    - Met en évidence le contexte social sous-jacent sans le traiter
    - Si on les traite, on déplace le problème, mais où ?
  - Modèle issu de la sismologie
    - Ne sait pas prédire les séismes, mais efficace pour prédire les répliques
    - Principe de localité : se concentre sur la répétition des victimisations
    - Modèle « stationnaire » entre deux « catastrophes »
    - Identique pour l'étude des flux Twitter

université de BORDEAUX

Conformisme de la « recommandation » (1)

- Les traitements de recommandation opèrent selon la logique inductive
  - Détection corrélations entre jeux d'entrée et jeux de sortie
- Exposent les structures sous-jacentes
  - Utile en médecine, où les relations de cause à effet sont bordées par la physiologie
  - Utile pour détecter les motifs répétitifs
    - Fraude (s'insérer dans les failles), comportements de consommation (biais culturels et éducatifs)

Conformisme de la « recommandation » (2)

- Ne peuvent que travailler en régime stationnaire
  - Sédimentation de la norme sociale
  - Cas de la « justice prédictive »
    - Pression sur les juges tant de la part des parties que de la hiérarchie judiciaire

Conclusion (1)

- Les mécanismes d'analyse algorithmique sont utiles pour :
  - Identifier des motifs au sein de masses de données
  - Potentiellement, évaluer l'évolution desdits motifs en réaction à un changement de l'environnement
    - Tel qu'un changement de politique publique
- Pouvoir modifier l'environnement lui-même dans une direction donnée suppose la capacité de « sortir » du modèle
  - Les personnes ne sont pas réductibles à leurs données, aussi précises fussent-elles

Conclusion (2)

- Existence d'une asymétrie majeure entre :
  - Les capacités et actes des responsables de traitement
  - Les niveaux d'information et de compréhension des personnes concernées
- Rapport de force difficile à renverser
  - Inertie des comportements individuels
  - Nécessité d'un encadrement législatif
    - Droits des personnes concernées (art. 10 loi I&L)
    - Loyauté et éthique des responsables de traitement