

# **Le succès des algorithmes prédictifs remet-il en cause notre notion de connaissance scientifique ?**

Anouk Barberousse  
Sciences, Normes, Démocratie, UMR 8011  
Sorbonne Université

***Une discussion sur la connaissance  
inductive et sa formalisation***

# Le point de départ

- Capacité de certains systèmes informatiques à faire des inférences statistiques (bien) plus performantes que les humains même aidés par des outils statistiques plus traditionnels
  - grâce à la taille des bases de données prises en compte
  - grâce aux algorithmes (nature de l'apprentissage)
  - utilisés pour des tâches d'aide à la décision (par exemple judiciaire)
  - qui donnent en sorties des prédictions probabilistes sur des cas individuels.
- Que fait cette capacité à la notion de connaissance scientifique ?
  - Doit-elle conduire à la modifier ?
  - Ou doit-on considérer qu'elle produit des connaissances de seconde zone ?

# Rendre la question plus précise

- S'agit-il de se demander s'il est plus important, dans les sciences empiriques, d'expliquer que de prédire ? (ou l'inverse ?)
  - Un vieux débat qui engage des positions fondamentalement opposées en philosophie des sciences
  - illustré récemment à propos du langage par le débat entre Chomsky et Norvig
  - qui traverse les pratiques des statisticiens.
- La question que je pose est différente.
  - Je considère que les algorithmes prédictifs peuvent, en droit, contribuer à la connaissance scientifique.
  - Je me demande comment caractériser cette contribution.

# Rendre la question plus précise, 2

- S'agit-il d'affirmer que la vraie connaissance est causale et de soutenir Judea Pearl (dans *The Book of Why*) dans sa croisade contre les statistiques et l'IA qui ne s'appuie pas sur les réseaux bayésiens ?
  - Pearl défend une conception anthropomorphique de la causalité, fondée sur des suppositions sur l'évolution cognitive de nos ancêtres.
  - Cette conception de la causalité a ses vertus, mais la notion de connaissance scientifique ne se réduit pas à la connaissance causale.
  - Certains continuent d'affirmer avec Pearson et Russell que, pour analyser correctement l'activité scientifique, il vaut mieux se passer entièrement du recours aux causes (la connaissance générale est portée par les énoncés de loi, exprimés sous forme d'équations— si on doute de la pertinence de la notion de loi, on peut la remplacer par celle de symétrie).
- L'idée est de faire droit aux conceptions de la connaissance scientifique qui ne la limitent pas à la connaissance causale.

# Le cœur de la question, d'un point de vue abstrait

- L'induction, source de connaissance empirique célèbre entre toutes, peut-elle être automatisée ?
- Les processus automatiques qui scrutent les bases de données de grande taille peuvent-ils donner lieu à la formation de justifications ?
  - de sorte que la propriété principale de la connaissance soit réalisée, car, en toute généralité, les connaissances peuvent être représentées par des énoncés vrais (ou plus exactement, dont on pense qu'ils sont vrais) dont on peut donner une justification acceptée par les spécialistes du domaine.

**Le cœur de la question,  
du point de vue de ses répercussions politiques et sociales**

- Peut-on affirmer que les résultats des algorithmes sont
  - evidenced-based?
  - knowledge-based?
- Les réponses à ces questions ont des répercussions sur l'acceptabilité des décisions que les algorithmes aident à prendre.
- Cet enjeu invite à s'interroger sur les propriétés des connaissances.

# Raisons de douter du caractère de connaissance des sorties des algorithmes

- Opacité du processus d'apprentissage
  - Opacité du calcul parallèle
  - Opacité du rapport entre calcul de fonction et réponse à une question dans une langue naturelle
- Démesure des bases de données relativement aux capacités cognitives humaines
- Il semble difficile de trouver, dans l'ensemble du processus, des *raisons* ouvertes à la discussion, comme le sont les justifications qui caractérisent les connaissances.

# L'induction sous toutes ses formes

- L'induction formalisée : un type de raisonnement non-monotone (la conclusion ne suit pas déductivement des prémisses).
  - Le saut logique entre les prémisses et la conclusion peut être représenté et quantifié grâce à de nombreux outils logiques et mathématiques — on n'a que l'embaras du choix, des logiques non-classiques à toutes les écoles de statistiques.
  - L'existence même de ce choix fait partie du paysage théorique : nous ne partons pas d'une situation dans laquelle l'induction recevrait un traitement formel unifié.
  - (Je mets de côté les inférences abductives pour ne prendre en considération que les raisonnements qui ne font pas appel au concept d'explications : généralisations, prédictions, classifications, ...)
- L'induction inconsciente, opérée par le système cognitif sans que le sujet s'en mêle (pour ainsi dire).
- L'induction automatique opérée par des machines.

# Vers une réponse de normand ?

## *Les vertus de la formalisation*

- D'un côté, c'est grâce à la formalisation que l'induction participe au processus de construction des connaissances scientifiques.
  - La formalisation permet l'unification conceptuelle des éléments qui entrent dans les prémisses.
  - Les règles du calcul fournissent le moyen de contrôler le caractère justifié de la conclusion.
  - L'ensemble est accessible à la discussion rationnelle.

## *Le socle des connaissances humaines*

- De l'autre, une grande partie de ce que nous savons, dans la vie quotidienne, provient des processus inconscients de notre système cognitif.
- Peut-on dire que les inductions automatiques participent des deux régimes ?

# Propriétés comparatives des inductions inconscientes et des inductions automatiques

- Le cerveau humain est très efficace pour faire des généralisations tout seul (de façon inconsciente, et donc opaque).
- Mais ce processus semble être affecté par des biais, c'est-à-dire des tendances de l'esprit humain à raisonner de façon incorrecte dans certaines circonstances.
  - Il existe un vaste débat en psychologie du raisonnement sur la question de savoir ce que sont vraiment les biais que certains chercheurs pensent avoir mis en évidence.
  - Cependant, il semble avéré que les capacités humaines de raisonnement sont affectées par des facteurs que les sciences formelles éliminent : pertinence, saillance, fréquence dans la vie quotidienne, ...
- Les algorithmes d'apprentissage profond sont construits de sorte à ne pas instancier (tous) les biais humains *et* à être plus efficace que les outils statistiques (plus) classiques.
  - En outre, ils sont plus rapides ; ils remplissent exactement la tâche qu'on leur confie, sans l'interpréter ; ils peuvent s'appuyer sur des nombres bien plus grands de cas.

# Conclusion

- Je ne vois pas de raison suffisante pour ne pas accepter comme des connaissances les sorties des systèmes informatiques.
  - Ce n'est pas de la connaissance bon marché en raison de la continuité scientifique entre statistiques et apprentissage machine.
- En revanche, la notion de connaissance scientifique me semble fondamentalement solidaire d'autres notions, comme celles de justification et de raison, qui dépendent trop du langage naturel pour pouvoir être affectées par les usages actuels et prévisibles dans un futur proche de l'apprentissage machine.
  - Les résultats des algorithmes prédictifs s'insèrent dans un ensemble de pratiques de connaissance qui, malgré toutes les entreprises de formalisation, sont sous le contrôle des capacités cognitives propres aux humains, dont le langage (naturel).