Des succès des ignorances en informatique

Serge Abiteboul
ARCEP, Inria, ENS, Paris

Merci

- Aux organisateurs du Séminaire sur *L'ignorance dans les sciences*
- Cela m'a permis de réfléchir à un problème essentiel pour les sciences – il était temps

Morale : on ne réfléchit jamais assez sur les sciences, quand on est un scientifique

Plan

1. Calculabilité et complexité
2. L’analyse de données massives
3. Intelligence artificielle
4. Distribution massive
5. Conclusion

1. Calculabilité et complexité

Un mythe fondateur de l'informatique : la calculabilité

Théorème (Turing et d'autres) : Il y a des problèmes qu'il est impossible de résoudre avec des algorithmes

Quels problèmes ?

- On ne parle pas de : est-ce que Dieu existe ?
- Mais de, par exemple : est-ce que ce programme s'arrêtera un jour ?
- Ou de : est-ce qu'il est optimal ?

- Cette incalculabilité immortalise notre ignorance
- Ça ouvre des questions : est-ce qu'un problème particulier peut être résolu par un algorithme ou pas

La complexité des problèmes

- Certains problèmes peuvent être résolus mais avec des ressources irréalistes : un milliard de processeurs réalisant chacun un milliard de milliards d'opérations
- Ça ouvre des questions : est-ce qu'un problème particulier peut être résolu par un algorithme avec des ressources spécifiées, ou pas
- Pas des problèmes obscures mais des problèmes hyper pratiques : voyageur de commerce, satisfiabilité d'une formule boolénne…

L’ignorance moteur de l’informatique

- Une source de questions fantastiques avec beaucoup de résultats négatifs

- Prenez une question simple : tester si une formule booléenne a une solution
  - Trouver un algo, comprendre sa complexité dans le pire des cas, en moyenne, suivant la tête de la formule…
  - Des questions essentielles en pratique
  - Apporte une meilleure compréhension des algos et des problèmes
  - Des masses de résultats, une grande compétition internationale 3SAT

Nombreux résultats négatifs

- On bosse depuis des années comme des bœufs sur ces sujets et ça rame

- On ne peut pas calculer A en temps polynomial
- Si on savait calculer B en temps polynomial, on saurait aussi calculer A

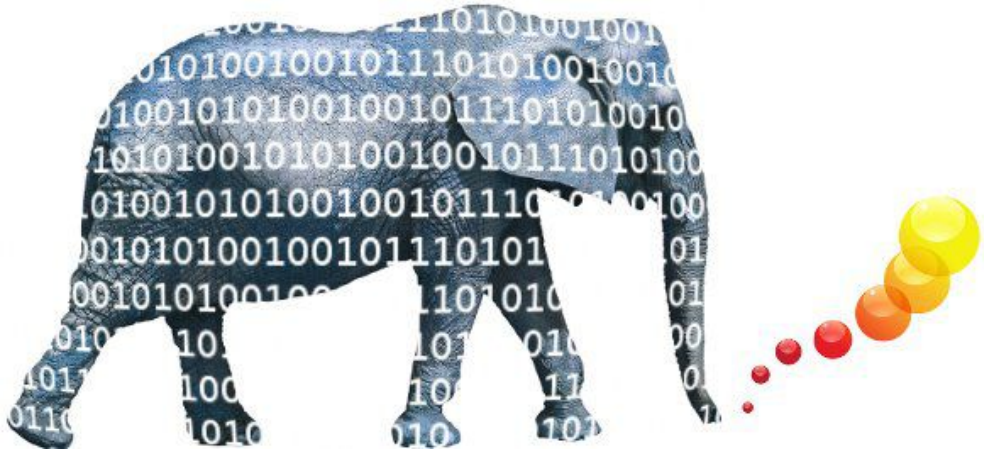
- Illustration personnelle : un des résultats dont je suis le plus fier, le Théorème Abiteboul-Vianu – du *foutage* de gueule ?

2. L’analyse de données massives

Le mythe

- En simplifiant
- Approche classique : on propose une théorie qui tient sur quelques diapos et on cherche à la vérifier expérimentalement
- Approche « analyse de données massives »
  - On accumule des masses inouïes d'observations
  - On les laisse parler
  - Une théorie va sortir du chapeau (ou pas)

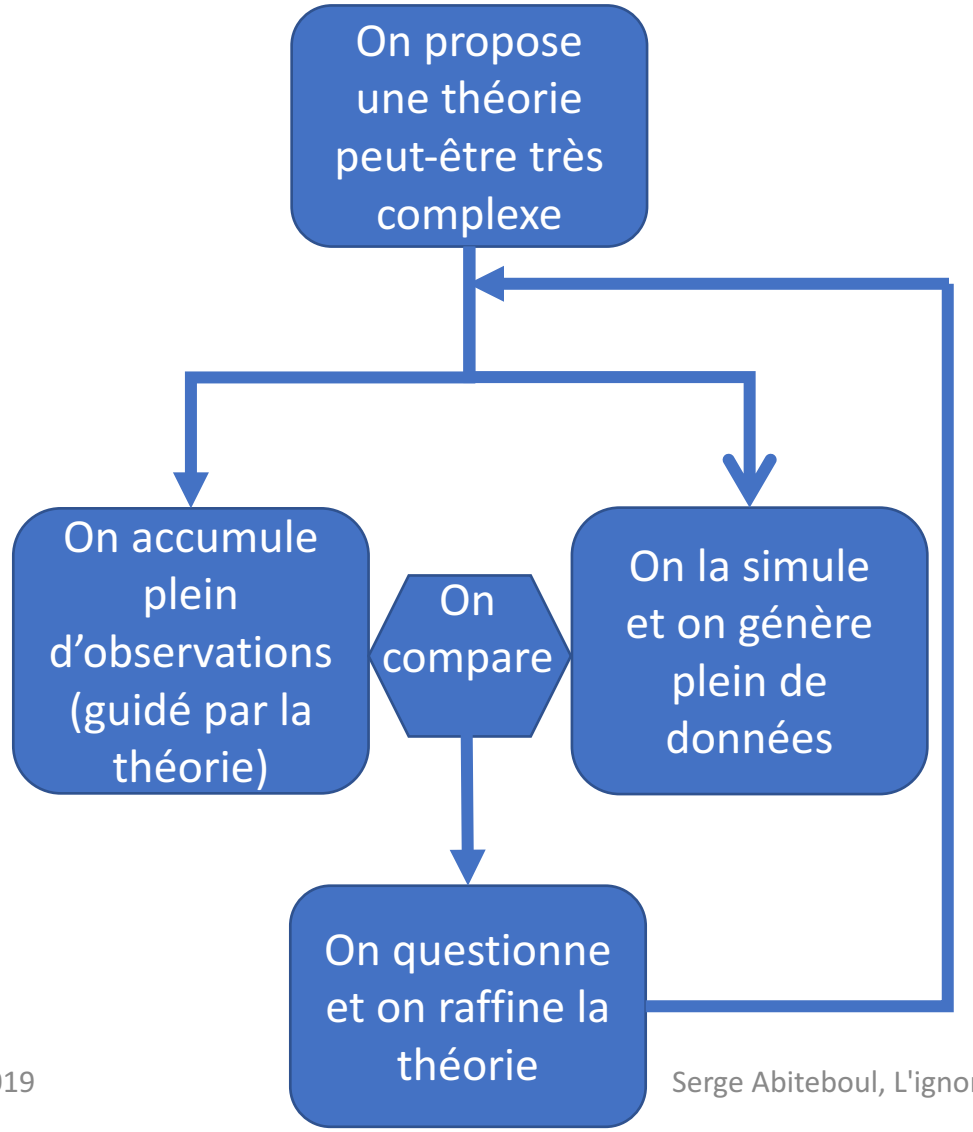
Les sciences questionnées



- L’explication du monde
  1. D’abord, par la religion
  2. Puis, par les sciences
  3. Maintenant, par l’analyse de données massives ?

- Evidemment, non: c’est juste un nouvel outil pour les scientifiques

Une vision un peu moins simpliste



- Exemple: le climat, le cerveau…
- Difficulté: on considère des systèmes physiques, biologiques, ou humain de plus en plus complexes
  - La théorie ne tient plus en quelques diapos mais en de gros programmes
  - Les techniques: gestion de données, calcul numérique, analyse de données, machine learning… sont partageables entre des sciences très différentes
  - Mais pas les connaissances fondamentales, et les idées théoriques

Les questionnements

Ce n’est pas simple

- Inventer une théorie
- Inventer des simulations réalistes mais réalisables
- Recueillir des données d’observations suffisantes
  - Vérifier l’absence de biais dans le recueil
  - Vérifier l’absence de biais dans les calculs

Les sciences en ressortent affaiblies ?

- Un peu : on sait mal expliquer les résultats
- Un peu : La théorie n‘est jamais parfaite/finie
- Mais les sciences sont habituées à ne pas savoir répondre à tout
- Surtout cela s’attaque à des questions auxquelles les sciences « classiques » ne savaient pas répondre

3. L’intelligence artificielle

Un autre mythe fondateur : l’intelligence artificielle

- Le point de vue de Turing
  - comme je ne sais pas définir « l’intelligence », je propose un test : est-on capable de construire un logiciel qui nous escroque et arrive à se faire passer pour un humain ?
- Le niveau ultime de l’ignorance : on ignore même ce qu’on cherche à faire
- Ca fait plus de cinquante ans qu’on suit pourtant ce programme
  - Trouver une tâche que les humains savent réaliser
  - Poser la question : et avec un algo ?
  - Reconnaissance d’image, traduction automatique, vérification de preuve…

On a développé plein de super techniques

- Le raisonnement automatique (système expert)
- Reconnaissance de forme
- Bases de données et bases de connaissances
- Apprentissage automatique (machine learning, deep learning…)
- …

- Correspond à différentes « fonctionnalités » humaines : raisonner, savoir, voir/entendre, apprendre…

L’apprentissage automatique

- Idée : imiter une base d’exemples
- Le point de départ
  - Par ex, un réseau neuronal + des données d’apprentissage
- L'algorithme « tune » les paramètres du réseau
- Super résultats sur certains problèmes
  - Analyse d’image, traduction automatique, etc.
  - Magique : on ne comprends pas bien les sciences derrière
- Problèmes
  - Mauvaises données d’apprentissage ➜ mauvais résultats

Garbage in – Garbage out

  - Nous ne savons pas comment mesurer l'erreur
  - Nous ne pouvons pas expliquer les résultats

Résultat collatéral : questionnements sur ce que c’est qu’être humain, sur l’humanisme

4. La distribution massive

On perd la maitrise de l’informatique ?

- Pour un programme donné, un algorithme précis
  - On peut l’analyser, en apprendre quelque chose
  - On peut le faire tourner et analyser ses traces d’exécution
- Mais quand on en a des millions, des milliards ?
- Ce n’est pas simple

- On finit par regarder cela comme ces grands systèmes physiques qu’on ne peut étudier que de l’extérieur : le climat, le cerveau…
- Un autre affaiblissement d’un point de vue scientifique ?

On perd la maitrise de la société ?

- C’est encore plus compliqué que ça
- On a un système avec
  - Des milliards d’humains connectés
  - Des milliards de machines avec leurs algorithmes qui, par exemple, choisissent les contenus à pousser
  - Des interactions complexes

Quelle société va-t-on construire avec des toutes ces machines et tous ces humains qui interagissent ? Comment on contrôle ça ?

On ne manque pas de questions…

5. Conclusion

L’ignorance est le moteur de la recherche

Rien de bien nouveau sous le soleil

Mais finalement, après 70 ans de progrès considérables, en informatique

- La plupart des problèmes étudiés aujourd’hui n’existaient pas alors
- On a accumulé plus d’ignorance que quand on a commencé
- On a encore plus de challenges que quand on a commencé

(Bonne nouvelle : les chercheurs ne sont pas près d’être au chômage)

Une mission : expliquer

- Le monde est devenu numérique
- Nombre de problèmes sociétaux actuels ont des liens étroits avec l'informatique
  - Réseaux sociaux : *hate speech, fake news, bullying, information bubble…*
  - Médecine, droit, police, usine... numérique
  - Vote électronique, voitures autonomes, cryptomonnaies
  - Communications cryptéees et anonymisation
  - Blockchain et cruptomonnaie
  - Disparition du travail avec les machines...
- Il faut aider la population à monter en compétence pour faire les bons choix sociétaux/politiques

C’est pour ça que je passe du temps

- A écrire des livres de vulgarisation, le blog binaire.blog.lemonde.fr
- A écrire des articles dans Le Monde, La Recherche…
- A faire des conférences grand publique, dans des lycées…

- Parce qu’il faut expliquer aux jeunes, aux décideurs, à tous…
- Parce qu’il faut tordre le cou aux angoisses… et aux visions bisounours

- **Et aussi parce qu’on peut y trouver beaucoup de plaisir**

binaire.blog.lemonde.fr

- Chronique dans Le Monde
- Chronique dans La Recherche