

Le numérique

Gilles Dowek

Quelques remarques sur la notion de culture scientifique

Une expérience de pensée : supposons que

- ▶ la connaissance scientifique soit constituée d'un certain nombre d'éléments que nous puissions compter (livres, articles...)
- ▶ une personne A dont la science est le métier en connaisse 0.1%
- ▶ une personne B dont la science n'est pas le métier en connaisse 0.01%

La différence de connaissance est énorme : **un rapport 1 à 10**

Mais la différence d'ignorance est minime : A ignore 99.9% des connaissances, B en ignore 99.99% : **un rapport de 1 à 1.0009**

À la fin de cette intervention

Vous aurez (peut-être beaucoup) augmenté vos connaissances
Mais (certainement) très peu diminué votre ignorance

Le but de cette intervention

- ▶ n'est pas de vous apprendre des choses (une goutte dans l'océan)
- ▶ mais de vous convaincre que le numérique / l'informatique, c'est facile et que chaque jour vous pouvez en apprendre une chose nouvelle
- ▶ il ne faut pas s'en remettre aux experts, mais essayer de comprendre par soi-même (ce à quoi les experts peuvent nous aider)

Comprendre par soi-même ?

- ▶ Les experts disent que le vaccin est efficace
- ▶ Vaccin testé sur 43 000 personnes réparties en deux groupes de même taille. Dans le premier, placebo, 162 cas. Dans le second, vaccin, 8 cas, 154 cas en moins. Efficacité = $154/162 = 0.95$

Encore beaucoup de choses à comprendre : intervalle de confiance, risques, persistance de la protection...

Mais chaque jour, un peu

Pas d'illumination, des éclaircissements

C'est cela la vraie culture scientifique

Les quatre concepts de l'informatique / du numérique

L'informatique est née quand nous avons commencé à utiliser des **machines** pour exécuter des **algorithmes**

Qu'est-ce qu'un algorithme ?

Une réponse décevante :

une méthode systématique qui permet de résoudre un problème

Par ma foi, il y a plus de quarante ans que j'utilise des algorithmes, sans que j'en susse rien ; et je vous suis le plus obligé du monde, de m'avoir appris cela.

Exemples

Les recettes de cuisine (systématique : il n'y a qu'à suivre la recette)

Chercher un mot dans un dictionnaire

- ▶ comparer le mot recherché au premier mot du dictionnaire, si ce n'est pas le même passer au deuxième, puis au troisième... (60 000 comparaisons)
- ▶ ouvrir le dictionnaire au milieu et comparer le mot recherché au mot médian, si ce n'est pas le même éliminer un demi-dictionnaire, et recommencer avec le demi-dictionnaire restant (16 comparaisons)

Des algorithmes qui pétrissent de l'eau et la farine vs. des symboles écrits

(la nuit des temps, il y a 5 300 ans)

Systematique

Un algorithme ne nous demande pas de prendre d'initiatives
Nous devons l'exécuter **mécaniquement**
Dès lors, pourquoi ne pas le faire exécuter par une machine ?

Le plus grand mystère de l'histoire des techniques

Des algorithmes et des machines depuis la haute Antiquité
Mais des **machines à exécuter des algorithmes** depuis le XVII^e
siècle seulement (Pascaline)

Les quatre concepts de l'informatique

Les ordinateurs sont des **machines** à exécuter des **algorithmes**

Cela demande d'exprimer les algorithmes dans un **langage**

Ces algorithmes s'appliquent à des objets : les **données**

Le concept de langage

Comparer le mot recherché au mot médian, si ce n'est pas le même éliminer un demi-dictionnaire, et recommencer avec le demi-dictionnaire restant

```
i = 0
j = 59999
while i < j:
    k = (i + j) // 2
    if s == nom[k]:
        i = k
        j = k
    elif s < nom[k]:
        j = k - 1
    else:
        i = k + 1
```

Langues et langages

Les langages sont créés **de toutes pièces**, par un petit nombre de personnes, avec un plan d'ensemble

Lexique réduit

Grammaire simple

Rétrospectivement :

XVII, 17...



$$x^3 + 4x^2 = 20$$

CH₄

OD : -1,25 (-0,50)180° OG : -1,00 (-0,25)180°

1.e4 d6 2.d4 Cf6 3.Cc3 g6 4.Fe3 Fg7 5.Dd2 c6 6.f3 b5

Le concept de donnée

L'algorithme de recherche d'un mot dans un dictionnaire opère sur un dictionnaire

L'algorithme de l'addition

$$\begin{array}{r} 129 \\ 706 \\ \hline 835 \end{array}$$

opère sur des nombres (129, 706...)

Ce dictionnaire, ces nombres, sont des **données** sur lesquelles les algorithmes opèrent

Des mots, des nombres, mais aussi des images, des sons (représentation : mais déjà 129 ou CXXIX?)...

Les données voyagent dans l'espace et le temps

I. À bas les algorithmes, vive les données

Qu'est-ce qui compte le plus : le dictionnaire ou l'algorithme pour chercher un mot dedans ?

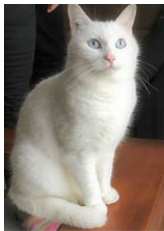
7 questions

3 portent sur les données (2 sur les algorithmes, 1 sur les langages, 1 sur les machines)

Pourquoi le concept de donnée gagne-t-il en importance ?

Un paradoxe : est-il difficile

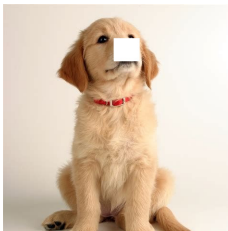
- ▶ de multiplier deux nombres de mille chiffres ? (oui / non)
- ▶ de reconnaître un chat d'un chien ? (non / oui)



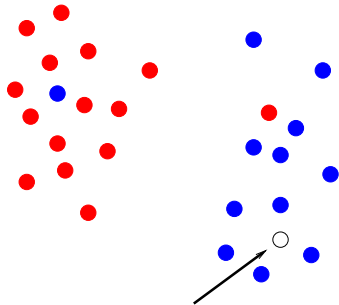
Pourquoi le concept de donnée gagne-t-il en importance ?

Un paradoxe : est-il difficile

- ▶ de multiplier deux nombres de mille chiffres ? (oui / non)
- ▶ de reconnaître un chat d'un chien ? (non / oui)

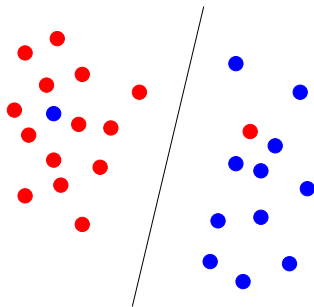


Puis une solution



Distance moyenne aux points bleus vs. distance moyenne aux points rouges

Mais on ne calcule pas la distance moyenne à chaque nouveau point



« Apprendre » la frontière (pas forcément une droite) à partir des données une fois pour toutes

Situer le nouveau point par rapport à la frontière

Reconnaissance d'images, mais aussi traduction automatique :
« *grenade* » ou « *pomegranate* » ? Un grand corpus de textes traduits suffit

Mais aussi des limites

Tout ce qui ressemble à un chien n'est pas un chien

Statistiquement correct

Véhicules sans conducteurs : tout est simple (trains, avions...) sauf l'analyse de la scène (voitures)

Faire confiance à un algorithme d'apprentissage ?

Mais aussi : tout ce qui ressemble à un récidiviste n'est pas un récidiviste

Pas de justice imitative (quoique : Hamurabi, jurisprudence...)

Comment s'appellent ces algorithmes imitatifs ?

Fondés sur l'utilisation d'une grande quantité de données

« Apprentissage automatique », « apprentissage machine »,
« apprentissage statistique »

Mais aussi **Intelligence Artificielle** (convocation d'un mythe : Galatée, le Golem, la créature de Frankenstein, les robots de Rossum, HAL 9000...)

Une autre révolution dans les données

Les chaînes de blocs

Registres (bases de données) **répartis et infalsifiables**

Une bonne métaphore : des notaires qui s'échangent des actes au lieu de les consigner au service de conservation des hypothèques

Toutes les activités consistant à tenir des registres automatisés (notariat, conservation des hypothèques)

Mais aussi banque (monnaie scripturale), émission de monnaie...

La gratuité

Nous conservons les données lorsque nous les donnons

Si je mange une pomme, vous ne pouvez pas la manger aussi : un bien rival

Si j'écoute la radio, vous pouvez l'écouter aussi : un bien non rival

Prix des données (= coût marginal) : 0

Une nouvelle économie de la gratuité et de l'abondance

Logiciel libre (et souvent gratuit)

Wikipédia : 30 millions d'articles (dans 280 langues)

Qu'est-ce qui est duplicable à coût nul dans un bâtiment ? Pas de ciment... Mais le plan de l'architecte, le calcul de structure, la planification des travaux... oui

Mais aussi une tendance inverse

Hadopi : penser le monde présent avec les outils du passé (la rémunération des créateurs à la copie, jadis permise par la rivalité du support)

Plutôt se poser la question : comment rémunérer les créateurs dans le monde présent

Une copie de la Joconde n'est pas la Joconde (perte de l'« aura »)

Les certificats non fongibles (**NFT**) permettent de distinguer l'original et de créer une rareté artificielle

Assuétude à la rareté

Mais aussi pièces et billets virtuels et non copiables

Les données publiques



Tout le monde produit des données en permanence

Mais l'État en particulier

Les citoyens ont le droit de savoir ce que l'État fait de leurs impôts

Ces données appartiennent collectivement aux citoyens

L'État a un **devoir** de les partager

Démocratie : participation mais aussi information des citoyens

Les données publiques

L'informaticienne : C'est mal de ne pas mettre en ligne le budget de la commune

Le Maire : Pourtant avant, nous ne le faisons pas. Était-ce mal ?

L'informaticienne : Non, car nous ne le pouvions pas. Mais **maintenant** c'est mal

Un devoir ?



Un devoir ?



Dès que la bouche d'incendie apparaît, l'action est un devoir

Le progrès technique **crée** un devoir éthique

Faut-il partager toutes les données ?

Non : tension entre deux valeurs : la **transparence** et le **respect de la vie privée**

Je ne veux pas que soit diffusé (ou diffusé à grande échelle) mon dossier de santé, les condamnations dont j'ai été l'objet, les prestations sociales dont je bénéficie, mon salaire...

Entre en conflit avec : je veux savoir ce que l'État fait des impôts que je paie

Une tentative de solution : l'anonymat

Anonymiser (littéralement : supprimer le nom)

Mais supprimer le nom ne suffit pas

Wolfgang Amadeus ***, compositeur né à Salzbourg le 27 janvier 1756 et mort à Vienne le 5 décembre 1791...

80% d'entre nous identifiables par leur date de naissance et leur code postal ($30\ 000 \times 100\ 000 = 3\ 000\ 000\ 000$ combinaisons)

Protéger la vie privée est une science (*differential privacy*, *k*-anonymat : il y a un xxx parmi les habitants de Paris vs. il y a un xxx parmi les habitants de La Bâtie-des-Fonds (2 habitants))

II. Algorithmes et décision

Utiliser des algorithmes pour décider

Exemples : APB, Parcoursup...

Les Jeunes communistes à l'offensive contre Parcoursup

Parcoursup : après les premières réponses, l'angoisse des listes d'attente

Parcoursup, machine à stress

(L'Humanité, Le Figaro, Le Monde)

Quel est le problème :

- ▶ l'utilisation d'un algorithme ?
- ▶ l'utilisation d'un algorithme exécuté par une machine ?
- ▶ l'utilisation de cet algorithme là ?

Un algorithme utilisé (sans machine) depuis le XVIII^e siècle

Les écoles ont des préférences (classement du concours), les candidats aussi

- ▶ Les écoles publient la liste des reçus
- ▶ Chaque candidat reçu dans une école démissionne des écoles qu'il souhaite moins
- ▶ À chaque fois qu'un candidat démissionne d'une école un nouveau candidat est reçu

Pas optimal

Alice préfère X à Y, reçue à Y et première non reçue X

Bob préfère Y à X, reçu à X et premier non reçu à Y

Aucun des deux ne démissionnera (un tiens vaut mieux...)

Si Alice démissionnait de Y, Bob y serait reçu et démissionnerait de X, où Alice serait reçue

Mais socialement accepté

Le problème ne semble pas être l'utilisation **en soi** d'un algorithme

Rhétorique de la **machine froide et inhumaine** « Un système totalement inhumain, un algorithme qui choisit votre vie. » (A. Hidalgo) a pratiquement disparu

Cet algorithme là ?

Mais en gros **le même** que celui utilisé depuis le XVIII^e siècle (agrégation des préférences des formations et des candidats)

Le problème est ailleurs

Dans la manière dont les formations déterminent leurs préférences

Normal que les candidats déterminent leurs préférences de manière opaque (choix individuel)

Les formations doivent déterminer leurs préférences selon des critères objectifs et transparents (choix collectif)

Or, c'est loin d'être le cas

« Algorithmes locaux » : parfois des algorithmes exécutés par des machines (moyenne des notes...), parfois des algorithmes utilisés par des humains, parfois ... nous n'en savons rien

Une inversion des valeurs

Au XX^e siècle on regrettait que la élèves soient réduits à une note (l'homme **multidimensionnel**)

Désormais on préfère des **critères objectifs et transparents** (comme la loi ? comme un algorithme ?)

Révolution culturelle où l'informatique a joué un certain rôle

III. Machines et effet de serre

La part des émissions de gaz à effet de serre dûes à *l'Information and Communication Technology*

Réévaluée récemment de 1.8-2.8% à **2.1-3.9%**

Agriculture 19%, bâtiments 18%, transport 16% (dont routier 12% et aérien 2%),

Ch. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G.S. Blair, A. Friday, The real climate and transformative impact of ICT : A critique of estimates, trends, and regulations, Patterns, 2, 9, 2021.

Mais ce n'est pas une raison pour ne pas faire mieux

La preuve de travail des chaînes de bloc

Des machines qui échangent des registres

Mais que faire si une machine est en désaccord ?

La majorité l'emporte

Pour éviter les attaques : importance d'éviter la multiplication des machines qui échangent ces registres : ticket d'entrée élevé : un potlatch énergétique (preuve de travail)

Chercher un autre ticket d'entrée

Mais ce n'est pas une raison pour ne pas faire mieux

Une part importante du trafic est de la vidéo

Si deux personnes en France téléchargent une même vidéo en Australie, elle est envoyée deux fois

Garder une copie en France au cas où quelqu'un d'autre veut la télécharger (mais si personne n'en veut...)

Mais ce n'est pas une raison pour ne pas faire mieux

À plus long terme

Calculer demande de transformer de l'électricité en chaleur

Mais on transforme beaucoup d'électricité en chaleur pour chauffer les bâtiments, sans même en profiter pour faire un calcul

Transformer les radiateurs en ordinateurs qui **calculent et chauffent en même temps**

IV. Les langages et le droit

La loi du 10 avril 1954

« Cette taxe est perçue à un taux ordinaire de 16,85 p. 100. »

Mais il faut ensuite écrire un programme qui calcule le montant toutes taxes comprises à partir du montant hors taxe

$$ttc = ht * 1.1685$$

Ces deux textes expriment la même chose

L'un dans la langue, l'autre dans un langage

Le code des impôts

détermine le montant de l'impôt de chaque contribuable en fonction de ses revenus, son capital, sa situation familiale...

C'est donc un algorithme... d'abord exprimé dans la langue, puis dans un langage de programmation

Mais la langue n'est pas si bien adaptée pour exprimer des algorithmes

Pourquoi, à (très ?) long terme, ne pas l'exprimer directement dans un langage de programmation ?

Ainsi, les députés voteraient directement le programme (le code)

À plus court terme, comment s'assurer que le code fait bien ce que dit la loi ?

Trouver un langage qui soit à la fois exécutable (compréhensible par un ordinateur) et d'assez haut niveau (compréhensible par un juriste) (Denis Merigoux)

Pour conclure

Les quatre concepts de l'informatique vous concernent

- ▶ **donnée** : les limites des méthodes d'apprentissage, les chaînes de blocs qui mettent en danger le monopole des États à émettre de la monnaie, la notion de propriété menacée par la gratuité, le devoir d'ouvrir les données publiques
- ▶ **algorithme** : décision publique
- ▶ **machine** : développement durable
- ▶ **langage** : l'expression de la loi

Mais cela n'est pas propre aux juristes

L'informatique nous concerne toutes et tous

Car nous sommes des animaux informationnels

L'essentiel de notre activité consiste à pétrir des symboles (et parfois nous pétrissons un peu de pâte à pain ou de ciment)

Le doux plus important pour nous que le dur

Les trois révolutions du doux : **l'écriture, l'imprimerie,**

l'informatique sont les trois grandes révolutions de notre histoire