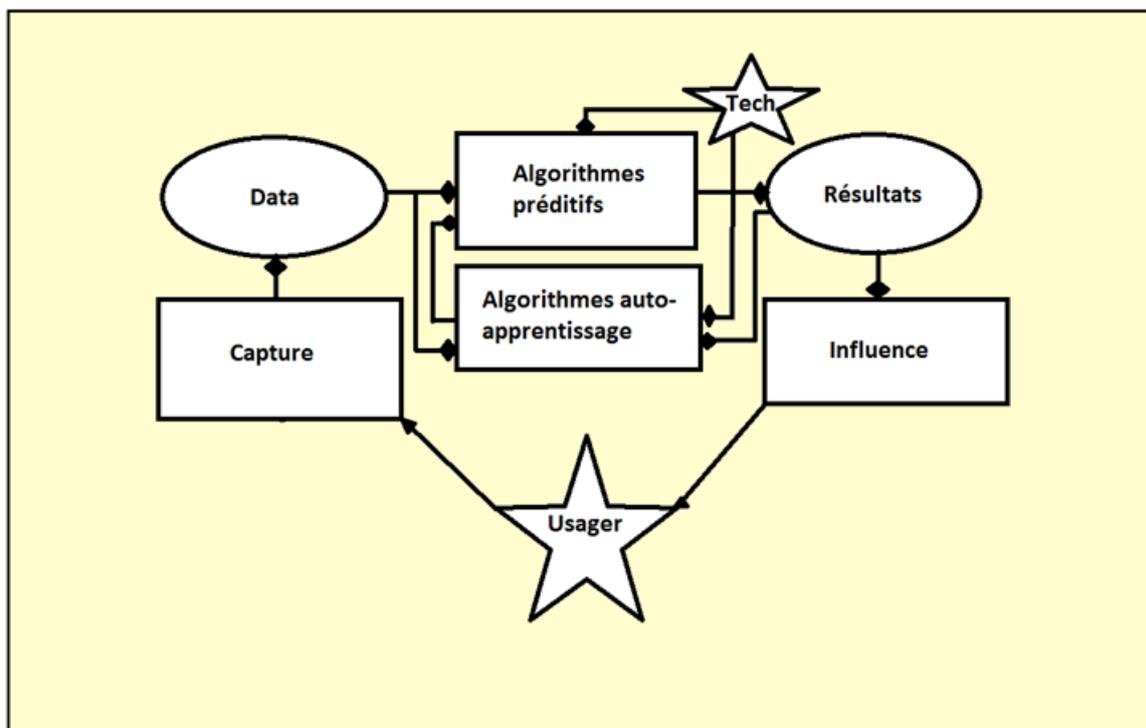


L'impact réel du big-data

Marc Halévy
Le 27 janvier 2017

L

Avant d'en mesurer les efficacités, les biais et les déficiences, analysons la "mécanique" du big-data sur base du schéma ci-dessous.



Les composantes du système.

Ce schéma met en présence deux acteurs : l'utilisateur (vous et moi lorsque nous utilisons des outils connectés) et le technicien qui programme le système (car il ne s'agit que de programmation d'un système totalement inintelligent, fait pour de ordinateurs qui ne savent faire qu'une seule chose particulièrement stupide : additionner des "zéro" et des "un").

Pour comprendre un système, quel qu'il soit, il faut en comprendre d'abord la finalité. La finalité du big-data est de prévoir les comportements futurs des usagers, sur la base de ses comportements passés, aux fins de les influencer dans le sens voulu. Quel est ce sens voulu ? Massivement, aujourd'hui, il s'agit de faire du big-data une énorme machine à faire acheter tout et n'importe quoi, à n'importe qui. C'est la plus puissante machinerie (machination ?) marketing et publicitaire de tous les temps. Tout ce qui vous concerne et qui passe par la Toile est capté, analysé et mouliné de façon à détecter vos "préférences" et vous fourguer, à longueur de temps, tout ce qui est censé correspondre à ces "préférences".

Mais au-delà de cette machinerie marketing, le big data peut avoir, pourra avoir et a déjà d'autres applications : tout ce que le système accumule vous concernant peut être revendu ou

pillé, et être utilisé contre vous à des fins policières, électorales, financières, fiscales, médicales, assurancielles, ...

Comment fonctionne le système du big-data ? Le mot central à bien comprendre est le mot "algorithme" qui, en grec, signifie simplement "calcul pénible". Un algorithme est un schéma logique, conçu et programmé par des hommes (il n'y a pas d'intelligence artificielle) pour tirer le meilleur profit statistique d'un immense ensemble de données mémorisées.

Ainsi, au premier niveau, l'algorithme pourra vous dire : "les autres clients qui ont acheté ou consulté le produit X ont aussi acheté ou consulté les produits Y et Z". C'est de la statistique de base, du niveau de l'école primaire.

Dans le schéma ci-dessus, ces algorithmes-là sont symbolisés sous l'étiquette "algorithme prédictif". Mais ils peuvent être beaucoup plus sophistiqués dès lors que l'on recourt à des outils statistiques sophistiqués comme les probabilités de corrélations, les analyses multicritères, les structures internes des distributions, etc ...

Quoiqu'il en soit, dès lors qu'il s'agit d'analyses statistiques, il faut se rappeler de leurs limites intrinsèques comme les effets d'ailes, la règle du cygne noir et, surtout, l'impossibilité de réduire les comportements humains à des mécanismes déterministes : mes "préférences" d'aujourd'hui ne sont que le reflet de mes humeurs d'aujourd'hui. Cela signifie, donc, en termes scientifiques que le comportement humain est un processus complexe, loin de l'équilibre, qui n'est que rarement prédictible.

Mais il y a une autre "couche" d'algorithmes que le schéma indique comme "algorithme d'autoapprentissage". De quoi s'agit-il ?

Il s'agit d'algorithmes qui exploitent les divers historiques du système pour faire évoluer les paramètres des algorithmes de base vers plus de précision ou d'efficacité ou de pertinence. Bien sûr, ce n'est pas l'ordinateur qui devient intelligent "en apprenant" comme le laisse parfois entendre la littérature de vulgarisation. L'ordinateur reste et restera à tout jamais une machine rudimentaire d'une effroyable bêtise. En fait, les programmes d'exploitation du système sont doublés de méta-programmes (conçus et programmés par de l'intelligence humaine) qui les font évoluer en fonction des résultats atteints lors des itérations précédentes. Autrement dit, grâce à ces algorithmes dits d'autoapprentissage ou méta-programmes, le système s'enrichit de ses propres travaux antérieurs.

Ce qu'il faut retenir, c'est que ces méta-programmes sont le fruit de la pensée humaine et que l'ordinateur n'y joue aucun rôle. Un ordinateur n'est jamais créatif. Il est une mécanique purement réactive qui ne fait que ce qu'on lui dit de faire au travers de programmes. Mais ces programmes font illusion tant ils parviennent à *simuler* un processus d'intelligence ou d'innovation.

Les biais et limites du système.

Tout le jeu est une partie de "je te tiens et tu me tiens par la barbichette" entre le technicien (l'inventeur de algorithmes) et l'utilisateur (le "bénéficiaire" du système).

Le système possède quatre fonctions : la capture des données, les algorithmes prédictifs, les algorithmes d'autoapprentissage et l'action d'influence.

Chacune de ces quatre fonctions est sujette à divers biais et limites.

La fonction "capture des données".

Outre l'épineuse question éthique et juridique du "droit" à capturer (et à exploiter et à revendre) des données, souvent personnelles, sans que leur propriétaire en soit vraiment

conscient, les mécanismes de capture sont incroyablement puissants. Dès que vous êtes connecté, vous êtes repéré et enregistré et pisté et épié. Tout est mémorisé.

Votre position géographique, vos courriels, vos photos, vos vidéos, vos achats, vos recherches, vos consultations, votre relationnel, vos "amis", vos consommations à domicile, vos rendez-vous, vos restaurants, vos opérations domotiques ou de télésurveillance, vos visites médicales ou hospitalières, vos médicaments, vos paiements, vos programmes télévisuels, vos horaires de lever et de coucher, vos amendes, vos condamnations, vos affiliations et *memberships*, vos déplacements, vos médias privilégiés, vos pétitions, vos commentaires de presse, etc ... La liste est infinie. Le "nuage" a une empreinte très détaillée et très intime de vous. Et cette empreinte, répétons-le, peut être exploitée (dans quel but ? commercial, policier, politique, assurantiel, fiscal, ... ?) et (re)vendue à quelqu'un qui l'exploitera à ses propres fins (lesquelles ?).

Il n'est guère difficile, à partir de cette montagne de données, de vous caser dans une quelconque typologie sociologique, idéologique ou économique.

Bien sûr - et c'est le premier biais - ces typologies sont terriblement simplifiantes et idéalisantes, elles sont incroyablement "pauvres" par rapport à la richesse qui est réellement en vous. Personne ne peut réellement être fiché dans une "case" de quelque typologie que ce soit. Mais ceux qui exploitent vos données personnelles, s'en fichent ... pourvu que vous soyez "casé". Cela seul importe puisque, on le verra, les exploitations algorithmiques sont toujours statistiques ; vous n'êtes qu'un item. Oui, mais ... si, au-delà des exploitations statistiques de masse, pour une raison quelconque, "on" vise une exploitation individuelle précise de votre profil numérique, les données de détail sont toujours bien là et le petit curieux peut en faire un usage ultra-personnalisé.

On peut partiellement se protéger contre ces captures de données personnelles. Trois grandes tactiques sont possibles.

La première : ne pas se connecter ... ou le moins possible ... et surtout pas aux systèmes prédateurs comme ceux des GAFAs californiens qui sont des pirates voyous, hors la loi, sans foi ni loi. Boycoter systématiquement les mafieux du type Google ou FaceBook. Les éradiquer de vos ordinateurs et de vos ordiphones qui seront tout ce que vous voulez *sauf* des Apple.

La deuxième : passer du temps, lors de vos connections, à brouiller les pistes et à parcourir la Toile sur des chemins et au travers de sites ou applications qui ne vous intéressent pas le moins du monde. Ainsi votre "portrait" dans le "nuage" sera falsifié.

La troisième : utiliser plein d'ordinateurs différents avec plein de noms, d'avatars, de pseudos, d'adresses, de mots de passe différents. Ainsi vous n'êtes plus vous ; vous devenez plusieurs. Mais bien sûr, on est là dans le jeu de dupe du garde-chasse et du braconnier : par recoupement de certaines informations, le système finira par "comprendre" vos tactiques de brouillage et finira par vous débusquer.

Aussi, la seule tactique pérenne est la première : connectez-vous aussi peu que possible et *jamais* en passant par les pirates californiens ou assimilés.

La fonction "algorithmes prédictifs".

L'idée de base est, après vous avoir "casé" dans une typologie statistique, et tout en continuant à suivre de près votre évolution, que l'algorithme prédictif va vous "attaquer" en assimilant votre comportement avec le comportement des autres humains enfermés dans la même "case" que vous.

Cela donne : "Les autres internautes qui ont fait cet achat, ont aussi regardé/acheté ceci ...".

Ou, ce symétrique : "Puisque vous avez acheté ceci, alors cela vous intéressera aussi ...".

C'est la logique des "préférences" ou des "favoris".

L'hypothèse centrale de ces algorithmes prédictifs, c'est que l'animal humain se comporte de façon hautement mécanique et déterministe. Le système vise donc à stimuler un comportement qui, statistiquement, correspond à votre profil. Et ça marche ...

Plus le niveau d'éducation est bas, plus l'esprit critique est faible, plus la résistance à la tentation est basse, et plus les résultats confirment que l'algorithme a raison de vous prendre pour un ... crétin.

Le jeu des préférences a aussi des conséquences bien plus graves que d'exacerber les propensions d'achat. Lorsque vous faites des recherches sur Google, par exemple, le système enregistre non seulement vos requêtes, mais aussi les propositions de réponse que vous sélectionnez. Donc, de proche en proche, il connaîtra le genre de réponses qui vous intéresse et que vous souhaitez recevoir. Aussi, progressivement, ils ne vous proposera plus que les réponses qui confirment ce que vous pensez, ce que vous recherchez, ce que vous préférez. Ainsi, vous êtes, au fil du temps, conforté dans vos convictions ... en dépit de tout esprit critique et de toute confrontation à la contradiction.

Cela s'appelle de la manipulation, du bidouillage de convictions et d'opinions. En gros : "Vous avez bien raison de penser ce que vous pensez, et on vous en offre le 'preuve' permanente". Quelle est la parade ? Déstabiliser le jeu des préférences et vous intéresser souvent à ce qui sort du cadre de vos habitudes, de vos goûts et de vos convictions.

La fonction "algorithmes d'autoapprentissage".

On entre ici dans les tactiques de manipulation au "second degré". Il s'agit moins de vous "caser" dans une typologie que d'apprendre à connaître vos logiques (changeantes) de choix et d'évolution. On passe d'une statistique statique à une statistique dynamique. Comment les gens qui ont réagi à A en faisant X, vont ils réagir à B ? Comment déduire des divers mouvements observés (et il y en a des millions de milliards en tout dont quelques centaines de milliers vous concernant), des modèles de logiques comportementales ?

Les algorithmes prédictifs supputent des corrélations de choix entre vous et les habitants de votre "case" typologique. Les algorithmes d'autoapprentissage modélisent les évolutions de ces corrélations de choix et vos trajectoires possibles au sein de la typologie même.

Comment déjouer tout cela ? En adoptant sciemment des stratégies d'évolution illogiques, irrationnelles, surréalistes, incohérentes, afin d'affoler les algorithmes jusqu'à les amener à vous mettre dans la seule case bénie de leurs maudites typologies comportementales : celles des "fous" imprévisibles qu'il faut oublier.

Je crois que Michel Foucault doit mourir de rire dans sa tombe.

La fonction "influence".

La fonction "influence" est bien sûr le résultat de tout ce qui précède. Le système vous envoie des stimuli, mais ceux-ci sont-ils efficaces ? Entraînent-ils, de votre part, le comportement escompté ? Êtes-vous dociles et obéissants ou rebelles et réfractaires aux influences ?

Personnellement, je fais très attention à cela. Dans mes comportement d'achat de livres, par exemple : lorsque le système tente de me fourguer un livre "corrélé" à l'achat que je viens de faire, il peut arriver - bien qu'assez rarement - qu'il me propose un livre qui, effectivement, m'intéresse. Dans ces cas, jamais je ne l'achète là et je fais mon achat par un autre canal.

Conclusion.

Une fois que l'on a bien compris le fonctionnement du big-data - et j'espère que cet article y aura contribué -, il faut en tirer trois conclusions.

Primo : de par la réalité grégaire et moutonnaire de l'espèce humaine, le big-data a de beaux jours devant lui, surtout en sachant que la science algorithmique fait d'impressionnants progrès tous les jours.

Secundo : en étant prudent et un tantinet rusé, il est possible de déjouer les algorithmes, mais ces ruses font en général long feu : c'est le jeu, disais-je, du braconnier et du garde-chasse.

Tertio : la seule tactique vraiment efficace est la déconnexion maximale et le boycott systématique des voyous californiens.

*

* *