

<http://larcenciel.be/spip.php?article440>



Le partage des données nous sauvera-t-il ?

- TRACES DU FUTUR -



Date de mise en ligne : mercredi 23 janvier 2013

Copyright © LARCENCIEL - site de Michel Simonis - Tous droits réservés

Pour le meilleur ou pour le pire ?

L'enjeu n'est pas une fatalité, il dépend, en partie, de nous...

De notre vigilance, et de notre capacité à réagir.

Première étape, savoir et comprendre de qui se passe...

Sommaire

- [Bien sûr, l'avalanche de données n'est pas sans danger.](#)

Bienvenue dans l'ère de l'avalanche de données, ou big data, en anglais. Les nouveaux milliardaires de la Silicon Valley, chez Google puis chez Facebook, sont devenus les rois de la collecte de données sur Internet grâce à la publicité en ligne. Au Forum économique mondial de Davos (Suisse) en janvier 2012, ce phénomène était au cœur des débats. Selon un rapport intitulé "Big Data, Big Impact", les données sont un nouveau type de biens économiques, tout comme les devises ou l'or.

En quoi consiste cette avalanche de données ? Selon IDC, société spécialisée dans l'analyse des informations sur Internet, la quantité de données en circulation croît de 50 % chaque année. Et il ne s'agit pas seulement d'un flot d'informations toujours plus important, mais de flux entièrement nouveaux.

Dans le monde entier, il existe par exemple un nombre incalculable de capteurs numériques installés sur des équipements industriels, des automobiles, des compteurs électriques ou des conteneurs. Ces capteurs peuvent mesurer et transmettre des informations sur la localisation, les mouvements, les vibrations, la température, l'humidité et même les variations chimiques de l'air. En les reliant à des ordinateurs, on assiste à l'essor de ce qu'on appelle "l'Internet des choses" ou "l'Internet industriel".

C'est également parce qu'il est plus facile d'accéder à l'information que la tendance au déferlement de données se poursuit. Par exemple, aux Etats-Unis, des données gouvernementales - notamment les chiffres liés à l'emploi - sont régulièrement publiées sur le web. Non seulement il devient plus facile d'accéder aux informations, mais ces dernières sont aussi plus facilement analysées par les ordinateurs. Les techniques liées à l'intelligence artificielle - comme le traitement du langage naturel, la reconnaissance des formes et l'apprentissage automatique - évoluent très rapidement.

Ces formes d'intelligence artificielle peuvent servir dans nombre de domaines. Par exemple, chez Google, la branche publicitaire tout comme les voitures robots expérimentales, qui parcourent des milliers de kilomètres sur les routes californiennes, utilisent toute une flopée de petits outils qui s'appuient sur l'intelligence artificielle.

Ces deux exemples, parce qu'ils impliquent l'analyse de quantités considérables d'informations et une prise de décision instantanée, constituent le défi titanesque du phénomène "big data".

La richesse des informations disponibles contribue ensuite à accélérer les avancées dans le domaine informatique, dans une sorte de cercle vertueux. Les algorithmes d'apprentissage automatique, par exemple, se nourrissent des données : ainsi, plus il y a de données, plus les machines sont savantes. Prenons comme exemple Siri, l'application pour iPhone capable de répondre à diverses questions, lancée par Apple à l'automne 2011. C'était à l'origine un

projet développé par le Pentagone, qui est finalement devenu une start-UP de la Silicon Valley. Apple a acheté Siri en 2010 et a continué à l'alimenter en informations. Aujourd'hui, comme les gens lui fournissent des millions de questions, Siri est en train de devenir une assistante de plus en plus compétente : elle fait office de pense-bête, connaît la météo, suggère des restaurants et sait répondre à des questions de plus en plus variées.

Nombreuses sont les anecdotes qui montrent qu'un raisonnement fondé sur les chiffres peut rapporter gros. Le meilleur exemple reste Moneyball, un livre de Michael Lewis publié en 2003 qui raconte comment la modeste équipe de base-ball des Oakland Athletics a brassé les chiffres et les statistiques impénétrables de ce sport pour repérer des joueurs sous-estimés.

Les entreprises de distribution comme Walmart et Kohl's analysent les ventes, les prix, la météo ainsi que des données économiques et démographiques pour adapter au mieux les sélections de produits selon les magasins et déterminer à quelles périodes proposer des soldes. Quant aux entreprises postales comme UPS, elles exploitent les informations liées aux délais de livraison de leurs camions et au trafic, afin d'affiner les itinéraires.

De leur côté, les sites de rencontres en ligne, comme Match.com, passent constamment au crible les listes de caractéristiques personnelles, les réactions et les commentaires, pour perfectionner les algorithmes permettant de proposer des rencontres. Aux Etats-Unis, à commencer par New York, les services de police cartographient et analysent des variables telles que les historiques des arrestations, les jours de paie, les événements sportifs, les précipitations et les jours fériés pour tenter de prédire où des infractions sont susceptibles d'être commises et déployer des agents dans ces zones.

D'après les recherches publiées en 2011 par des économistes de la Sloan School of Management du MIT [Massachusetts Institute of Technology], il semble que les méthodes de gestion fondées sur les données se répandent dans le monde des affaires et commencent à porter leurs fruits. Après avoir étudié 179 grandes entreprises américaines, les économistes se sont rendu compte que les sociétés qui prenaient des décisions en fonction de l'analyse de données connaissaient une productivité de 5 à 6 % plus importante que les autres, impossible à expliquer par d'autres facteurs.

Bien sûr, l'avalanche de données n'est pas sans danger.

Les informations sont collectées et analysées grâce à des modèles informatiques et mathématiques. Ces modèles, comme les métaphores en littérature, sont des "simplifications explicatives". Ce sont des outils qui nous aident à mieux comprendre, mais ils ont leurs limites. Un modèle peut tout à fait repérer une corrélation et en tirer une conclusion statistique injuste ou discriminatoire fondée sur des requêtes sur le web. Cela est susceptible d'avoir des répercussions sur les produits, les emprunts bancaires et l'assurance-maladie qu'une personne se voit proposer, préviennent les défenseurs de la confidentialité [des données].

En dépit des mises en garde, l'expansion du phénomène semble irrésistible et ce sont les chiffres qui ont le contrôle de la situation. "Culturellement, quelque chose a changé", conclut Andrew Gelman, qui enseigne les statistiques et les sciences politiques à l'université Columbia.

Steve Lohr

Le Courrier International, n° 1130, juin-juillet 2012

Le partage des données nous sauvera-t-il ?

"Les sociétés qui prennent des décisions en fonction de l'analyse de données connaissent une productivité de 5 à 6 % plus importante."