

INFORMATIQUE // A partir d'archives de journaux et de statistiques, deux chercheurs ont réussi à prédire des épidémies ou des émeutes. Le « big data » va-t-il permettre de connaître le futur ?

Quand le « big data » prédira l'avenir

Frank Niedercorn
fniedercorn@lesechos.fr

L'an dernier, Cuba a été touché par une épidémie de choléra. Une première en cinquante ans qui a surpris le gouvernement et la communauté médicale. Pourtant, cette épidémie avait été prévue par deux scientifiques, Eric Horvitz, codirecteur de la recherche chez Microsoft, et Kira Radinsky, chercheuse à l'institut de recherche israélien Technion. On est encore très loin de l'univers du film « Minority Report » ou de la « psychohistoire », cette discipline inventée par l'auteur de science-fiction Isaac Asimov et capable de prédire l'avenir de l'humanité à partir des connaissances de la psychologie humaine, des phénomènes sociaux et des statistiques. Eric Horvitz et Kira Radinsky étaient beaucoup plus modestes : leur objectif était de prouver que l'on pouvait prédire certains événements très spécifiques, comme des épidémies, des crises de mortalité ou des émeutes en se basant sur le passé et les probabilités.

« Même si, dans le futur, chaque événement survient dans des circonstances particulières, il obéit tout de même à un modèle déjà observé dans le passé », résume Kira Radinsky. Les deux chercheurs ont compilé l'intégralité des articles du « New York Times » et largement utilisé les ressources de LinkedIn, un projet collaboratif du W3C (Consortium World Wide Web) qui met automatiquement en relation des centaines de sources d'informations et trouve des connexions entre des milliards de données. Ces données ont ensuite été traitées pendant un mois sur une trentaine d'ordinateurs.

L'objectif était de trouver des corrélations entre certains événements, puis de développer des modèles mathématiques permettant de prédire d'autres événements. Avec des résultats étonnants. Ils ont par exemple mis en évidence qu'un territoire enclavé avait de plus fortes chances d'être touché par une épidémie de choléra après des tempêtes, notamment s'il a d'abord été frappé par la sécheresse. Certes, les épidémiologistes suivent la même logique, mais « un système totalement informatisé traite un nombre de

données infiniment plus important, provenant de sources plus diversifiées sur une durée plus longue. En outre, il peut faire une veille en temps réel et produire une alerte si le risque d'un événement se précise », précisent les chercheurs dans un article publié à l'occasion de la conférence Web Search and Data Mining, en février dernier. Ils y affirment que l'épidémie de choléra qui a frappé l'Angola début 2007 était prévisible plusieurs mois auparavant. Et estiment que les manifestations survenues à New York en 1999 après la mort d'un jeune homme tué par la police étaient également prévisibles.

Un marché très juteux

« Cette approche consistant à limiter la recherche en se focalisant sur des événements précis et à explorer des ressources très variées est novatrice », explique Daniel Gatica-Perez, chercheur au sein de l'Idiap, un groupe de recherche dépendant de l'École polytechnique de Lausanne (EPFL). Et l'on ira sans doute beaucoup plus loin. D'abord parce que la puissance de traitement informatique ne coûte presque plus rien. Ensuite en raison de l'abondance de données, notamment publiques, disponibles sur Internet. « Nous vivons dans un monde qui se numérise. Nos faits et gestes, ainsi que la plupart des objets que nous manipulons, produisent des données que l'on sait désormais traiter à moindre coût », résume Patrice Poiraud, responsable de l'activité « big data » chez IBM.

Enfin, parce qu'il s'agit d'un marché très juteux. Depuis des années, les consommateurs ont de moins en moins de secrets pour les spécialistes du marketing. Parti d'Internet et du e-commerce, avec les algorithmes de Google ou d'Amazon, le phénomène va désormais bien au-delà. Avec l'aide d'IBM, le zoo de Cincinnati suit ses visiteurs à la trace et adapte son offre en fonction des publics et des parcours. A l'arrivée, une augmentation des ventes dans l'enceinte du zoo de 25 %. Kira Radinsky, qui fait partie de la liste des 35 innovateurs de moins de 35 ans les plus prometteurs du MIT, a déjà créé une start-up, SalesPredict.

A l'écoute des réseaux sociaux

Aujourd'hui, le phénomène dépasse le cadre du marketing. Ainsi, plusieurs villes améri-

caines organisent désormais leurs patrouilles de police en fonction des prévisions de la criminalité basées sur des statistiques. L'analyse des réseaux sociaux peut aussi aider à la prédiction. « Les débordements de supporters au Trocadéro après le titre du PSG étaient parfaitement prévisibles. Nous les avons suivis sur Twitter plus de vingt-quatre heures avant », affirme Joël Robino, fondateur de la société Apicube, une start-up qui analyse automatiquement 1,5 milliard de conversations issues des réseaux sociaux.

Même nos comportements individuels semblent prévisibles : dans le cadre d'un concours organisé par Nokia, les scientifiques de l'EPFL ont réussi à prédire le déplacement des participants grâce à l'analyse de toutes les données de leur téléphone mobile, avec un taux de réussite moyen de 60 %, allant jusqu'à 90 % pour certains individus.

Est-ce pour autant la promesse d'un monde devenant très prévisible ? Certains restent sceptiques. David L. Donoho, professeur de statistiques à Stanford, souligne l'incapacité de la plupart des experts à anticiper les grandes crises que le monde a connues, de la chute de l'Empire soviétique au printemps arabe en passant par la crise financière de ces dernières années. Et, même bardés de statistiques, les experts se sont parfois lourdement trompés : en 2009, la grippe H1N1 a finalement provoqué moins de morts qu'une épidémie de grippe banale.

Rien ne sera sans doute jamais prévisible à 100 %. « C'est impossible et même arrogant d'y songer. Il suffit de regarder la météorologie, dont les prévisions, pourtant extrêmement utiles, ne sont jamais exactes à 100 % et ne le seront sans doute jamais », estime Cameron Alverson, spécialiste de l'analyse de données numériques chez Dell. Nate Silver, l'un des spécialistes les plus reconnus du monde des statistiques (1), prône d'ailleurs la modestie : « On ne peut pas affirmer que l'humanité va devenir plus prédictible. L'inverse est peut-être vrai aussi. La science, qui a permis de découvrir une partie des lois de la nature, a en parallèle rendu l'organisation de nos sociétés plus complexe. »

(1) « The Signal and the Noise », Penguin Press, 2012.

Petite histoire des prédictions

1936. Le mathématicien britannique Alan Turing théorise le fonctionnement des futurs ordinateurs et définit ce qu'est un algorithme. Il emploie pour la première fois le terme « intelligence artificielle » dans un article en 1950.

1942. Isaac Asimov publie une série de nouvelles sous le titre « Fondation ». Il y imagine une science fictive permettant de prévoir l'avenir de l'humanité grâce à l'application des statistiques aux phénomènes sociaux, la « psychohistoire ».

1950. L'ordinateur américain Eniac effectue la première prévision météorologique numérique. Les résultats sont bons jusqu'à 24 heures... mais catastrophiques au-delà.

1980. Les prévisions météorologiques nationales sont fiables jusqu'à 48 heures maximum. Il faut attendre 2003 pour avoir de bons résultats à 7 jours.

2008. Gartner Group utilise le terme « big data » en référence à l'explosion des données numériques. Le cabinet s'appuie sur des recherches du cabinet Meta Group, qu'il a racheté en 2004.

« Même si chaque événement survient dans des circonstances particulières, il obéit à un modèle déjà observé dans le passé. »

KIRA RADINSKY
Chercheur à l'Institut Technion



Dans « Minority Report », film de Steven Spielberg inspiré d'une nouvelle de Philip K. Dick, la police est capable de visualiser les crimes avant même qu'ils n'aient lieu. Photo Twentieth Century Fox / Dreamworks

