
Les chercheurs en sciences sociales s'initient au langage des algorithmes

Description

Les applications informatiques au service de l'étude des comportements des individus se multiplient. Dans les centres de recherche en sciences sociales fleurissent des programmes consacrés à l'interprétation par des algorithmes des actions des individus.

Aux Pays-Bas, les chercheurs de l'université de Twente, en collaboration avec l'institut Meertens, spécialisé dans la langue et la culture néerlandaises, ont créé un programme permettant de découvrir le sexe et l'âge des utilisateurs de Twitter en fonction du contenu des messages qu'ils envoient sur le site de microblogging. Les textes des 200 derniers tweets de plus de 3 000 utilisateurs de Twitter ont ainsi été croisés avec des listes de mots ou suites de mots que les chercheurs avaient eux-mêmes établies par tranche d'âge et par sexe. Chaque néerlandophone peut en faire l'expérience sur le site tweetgenie.nl lancé en mai 2013.

Concernant le sexe, les résultats formulés correspondent à des stéréotypes, le mot « football » se retrouve fréquemment dans les messages écrits par des hommes et le mot « ongles » dans ceux des femmes. Les plus jeunes tweetent des émoticônes et les personnes plus âgées forment des phrases plus longues.

En revanche, la détermination de l'âge de l'utilisateur connaît une marge d'erreur de quatre ans en moyenne : les résultats sont plus pertinents pour les jeunes gens que pour les personnes plus âgées pour laquelle la marge d'erreur est supérieure à la moyenne.

Autre constat des chercheurs : au-delà de 35 ans, les utilisateurs de Twitter rédigent des messages dont la facture est assez semblable, tandis que les différences d'expression linguistique sont beaucoup plus marquées entre un utilisateur de 15 ans et un utilisateur de 20 ans. Ce programme devrait prochainement être étendu à d'autres langues que le néerlandais, ainsi qu'à d'autres réseaux sociaux.

Une étude britannique publiée en février 2013 dans les Actes de l'Académie américaine des sciences (PNAS) porte sur l'utilisation des algorithmes pour créer des profils de personnalité. Les chercheurs de l'université de Cambridge ont montré que les données numériques, facilement accessibles, relatives aux comportements des utilisateurs de Facebook peuvent servir à prédire, avec précision et de façon totalement automatisée, toute une gamme d'attributs personnels, notamment l'âge, le sexe, l'orientation politique, religieuse ou sexuelle, l'intelligence, le bonheur...

L'étude a été réalisée sur une population de plus de 58 000 volontaires utilisateurs de

Facebook, à partir de leur usage du bouton « Like », de leur profil démographique et de tests psychométriques (techniques de mesures pratiquées en psychologie). Les résultats obtenus sont exacts à 88 % pour prédire la sexualité des hommes, à 95 % pour identifier les Noirs et les Blancs et à 82 % les musulmans et les chrétiens. S'appuyant sur une grande quantité de données, les calculs mathématiques permettent d'aller loin dans les déductions ayant trait à la personnalité, en révélant l'orientation sexuelle, l'usage de drogue ou encore le fait d'avoir des parents divorcés. L'homosexualité est ainsi révélée, non par la fréquentation de sites spécialisés, mais par les musiques écoulées ou les programmes de télévision regardés. Lancé en 2009, la fonction « j'aime » est devenue omniprésente sur Internet, elle a engendré environ 2,7 milliards de clics par jour en 2012.

Seul projet impliquant les sciences sociales proposées dans le cadre du concours organisé par la Commission européenne pour déterminer les deux grands programmes phares de recherche dans le domaine des technologies futures émergentes (mais pas finaliste, voir infra), FuturICT a pour ambition de modéliser les évolutions sociales afin de pouvoir les anticiper. Plus de 50 établissements de recherche (grandes écoles et universités) établis dans une vingtaine de pays européens (dont la France) et extra-européens, notamment la Chine, participent à l'élaboration d'un simulateur sociologique planétaire (*Living Earth Simulator*). Utilisant la puissance de calcul des ordinateurs, cette simulation repose sur tous les aspects de la société, politique, économique, culturel, social et climatique. Basé sur la théorie des dominos selon laquelle une cascade de petits événements de moindre importance peut engendrer une crise mondiale, politique, financière, économique ou environnementale, FuturICT imagine un « système nerveux planétaire », captant en continu sur Internet des milliards de données en provenance des institutions ou des individus (*Big data*) et de l'environnement naturel (météorologie), pour déterminer des modèles sociaux. En jouant sur une alchimie des paramètres, le but est de tester l'efficacité des choix opérés. Grand observatoire de l'humanité, le simulateur sociologique du projet FuturICT permettrait ainsi au monde de se prémunir contre d'éventuelles catastrophes qui le menacent. Cependant, les algorithmes pourront-ils vraiment tout simuler, alors que « l'essentiel de la vie humaine, c'est précisément ce qui ne peut être calculé », selon la formule de l'écrivain et journaliste Jean-Claude Guillebaud ? Accessible à tous, la plate-forme Futur ICT sera opérationnelle en 2014.

Les chercheurs n'ont jamais eu autant d'instruments de travail à leur disposition pour tenter d'appréhender les phénomènes sociaux et d'étudier l'individu, lui-même désormais équipé et metteur de données sur lui et sur les autres. À l'heure où la manipulation de données avec l'avènement du Big data est annoncée comme un nouveau champ d'investigation, le risque encouru n'est-il pas, en privilégiant les corrélations plutôt que les hypothèses, de faire paraître comme vrai ce qui ne l'est pas ?

Sources :

- FuturICT, Global computing for our complex world, futurict.eu
- « Private traits and attributes are predictable from digital records of human behaviour », Michal

Kosinski David Stillwell and Thore Graepel, edited by Kenneth Wachter, University of California, Berkeley, CA, pnas.org, February 2013.

- « Le bouton « j'aime » sur Facebook vous révèle », AFP et AP, radio-canada.ca, 11 mars 2013.
- « Simulation numérique des conflits sociaux », Pablo Jensen, *Le Monde diplomatique*, avril 2013.
- « Pays-Bas : un programme évalue votre âge en fonction de vos tweets », AFP, LePoint.fr, 13 mai 2013

Categorie

1. Techniques

date création

21 mars 2013

Auteur

françoise