Comment notre ordinateur nous manipule

Le Monde | 10.04.2014 à 17h00 • Mis à jour le 13.04.2014 à 14h05 |

Par **Yves Eudes** (/journaliste/yves-eudes/)

Tout a commencé avec une géniale invention, celle du cookie. Une simple ligne de code – par exemple MC1:UID = 6daa554691bd4 f9089dc9d92e5cdadf4 – déposée sur votre navigateur par les sites Web que vous visitez, et des publicités ciblées s'affichent sur votre écran.

Appelé ainsi en référence aux biscuits que les restaurants offrent au moment de l'addition, le cookie apparaît dès 1994, l'année où le Web s'ouvre au public. Vingt ans plus tard, il reste le socle de la publicité en ligne, une industrie qui a réalisé en 2013 un chiffre d'affaires mondial de 102 milliards de dollars.

FICHÉ 108 FOIS EN 3 CLICS

Les cookies sont gérés par des sociétés spécialisées qui les déposent, les récoltent, les classent, les analysent, les agrègent et les revendent. Ils servent à vous identifier, à vous pister de site en site, à retenir vos mots de passe, à gérer vos paniers

d'achat, à déterminer si votre navigation est lente ou rapide, hésitante ou déterminée, systématique ou superficielle...

L'objectif est de vous « profiler », c'est-à-dire de créer des fichiers personnalisés, stockés dans des bases de données. En d'autres termes, de mieux vous connaître afin de vous présenter le bon message publicitaire au bon moment et dans le bon format. Vous pouvez effacer les cookies, mais de nouveaux arriveront dès que vous reprendrez la navigation. Et si vous les bloquez, la plupart des sites ne fonctionneront plus. Certains cookies ont la vie dure : ceux que dépose Amazon aujourd'hui sont conçus pour durer jusqu'en 2037.

Un exemple : dès la page d'accueil du site de e-commerce Priceminister, votre navigateur reçoit d'un coup 44 cookies provenant de 14 agences spécialisées – telles que RichRelevance, Doubleclick, Exelator... Rendez-vous à la rubrique « Téléphonie mobile », vous récoltez 22 nouveaux cookies. Et en cliquant sur la photo d'un smartphone Samsung, vous déclenchez une nouvelle rafale de 42 cookies provenant de 28 sources : en trois clics, vous voilà fiché 108 fois par une quarantaine de bases de données. Si vous commencez à acheter le téléphone mais abandonnez en cours de route, vous serez repéré par la société française Criteo, spécialisée dans le « reciblage ». Des publicités pour le produit que vous avez failli acheter s'affichent sur votre écran pendant des jours, et vous suivront sur tous les sites que vous visiterez.

Criteo a besoin de machines puissantes : pour identifier un internaute, contacter la plate-forme qui gère ses espaces publicitaires, proposer un prix, conclure l'affaire et lancer l'affichage de la bannière, elle dispose de 13 centièmes de seconde – faute de quoi l'espace sera vendu à un concurrent.

Entreprise de taille moyenne, Criteo reçoit 20 teraoctets (vingt mille milliards) de données par jour, et touche 850 millions d'internautes par mois, certains des centaines de fois. On ignore le volume de données traitées par un géant comme Google, mais on sait qu'il possède plus d'un million de serveurs dans le monde. Pour rendre compte de cette réalité, les mathématiciens ont

inventé une nouvelle unité de compte, le zettaoctet (mille milliards de milliards).

PROFILS DÉTAILLÉS

Afin d'affiner le ciblage, les publicitaires croisent les cookies avec d'autres données récoltées sur Internet : votre adresse IP (Internet Protocol, qui identifie et localise votre ordinateur), votre langue usuelle, vos requêtes sur les moteurs de recherche, le modèle de votre ordinateur et de votre navigateur, le type de votre carte de crédit… Si vous avez livré des informations nominatives – en faisant un achat ou en remplissant un questionnaire –, elles seront également exploitées. Parfois, vos données Internet seront croisées avec d'autres, provenant du monde réel – relevés de cartes bancaires, tickets de caisse, déplacements de votre téléphone…

La société Acxiom, spécialisée dans le croisement des cookies et des données obtenues par d'autres moyens, vend aux annonceurs des profils triés selon 150 critères, parmi lesquels « fait de la couture », « héberge un parent âgé » ou « possède un chat ». Un fichier de mille personnes contenant des données de base est vendu en moyenne 60 centimes, mais le prix peut grimper à 250 euros pour des profils détaillés – comme, pour un laboratoire pharmaceutique, une liste d'adultes obèses ayant déjà acheté des produits amincissants. Ces données sont « anonymisées » car, pour vous cibler, les ordinateurs n'ont pas besoin de votre nom – il leur suffit de connaître vos revenus, vos envies, vos besoins, votre sexe, votre âge, votre métier, vos loisirs, votre origine ethnique, votre code postal, vos maladies, votre situation de famille, votre logement, votre voiture, votre religion, vos voyages…

Le ciblage va jusqu'à modifier le prix d'un produit en fonction du profil. Quand un site de voyage voit que vous venez de consulter un comparateur de prix, il baisse ses prix pour s'aligner sur ceux de ses concurrents, quitte à se rattraper sur les « frais de dossier ». Si vous vous connectez avec un ordinateur à 3 900 euros, le site affichera des chambres d'hôtel plus chères que si vous utilisez un portable à 300 euros. Le libre choix du consommateur, apparemment décuplé par la puissance de l'informatique, semble en fait amoindri. L'homme numérique aurait-il accepté de se laisser dominer par les machines dans le seul but de faire ses courses de façon moins fatigante ?

« NOUS LAISSONS LES TECHNOLOGIES NOUS FAÇONNER »

Le philosophe William Bates, professeur d'histoire des technologies à l'université de Californie, à Berkeley, note que les professionnels du big data exploitent une caractéristique essentielle de l'être humain : *« Ils nous font comprendre que nous ne contrôlons pas notre propre comportement et, en un sens, ils ont raison. Statistiquement, les humains réagissent de façon très prévisible à certaines situations, selon des mécanismes cérébraux qu'ils ne maîtrisent pas. »*

Une situation qu'amplifie l'accélération de l'innovation : *« D'un côté, les technologies de l'information influent sur notre cerveau : nous ne pensons plus de la même façon que les*

générations précédentes, souligne William Bates. *Mais par ailleurs, très peu d'entre nous, y compris les jeunes, comprennent comment fonctionne un ordinateur. Nous laissons les technologies nous façonner, mais nous n'avons pas encore créé les outils intellectuels pour nous aider à comprendre ce qui nous arrive. »*

Le philosophe relativise cependant cette angoisse existentielle : *« Depuis les débuts de la civilisation, les humains sont construits par la culture dans laquelle ils évoluent, et donc par la technologie qui en fait partie. C'est ce qui nous distingue des animaux. Il serait naïf de croire que le cerveau humain était jadis plus "naturel" ou plus "libre", et que les nouvelles technologies nous auraient dépossédé de quelque chose. »*

Pour influer sur nos cerveaux, les publicitaires disposent d'une énorme puissance de calcul. Ils emploient des mathématiciens pour concevoir les algorithmes, des développeurs pour les traduire en langage informatique, des ingénieurs pour construire l'architecture des bases de données, des analystes pour capter et exploiter des données…

ALGORITHMES AUTO-APPRENANTS

Or, le secteur vit déjà une nouvelle révolution grâce aux constants progrès techniques. La dernière mode est à « l'analyse prédictive » : au lieu de réagir au comportement des internautes, les publicitaires veulent les prévoir afin d'agir par anticipation. Pour cela, ils font appel à une discipline encore expérimentale : le *machine learning*, ou apprentissage automatique, une branche de l'intelligence artificielle. Il s'agit de doter les ordinateurs de la capacité à améliorer leurs performances sans intervention humaine. Selon Franck Le Ouay, directeur scientifique de Criteo, *« le "machine learning" désigne la capacité d'un programme à s'adapter à une nouvelle situation. Nous mettons au point des algorithmes auto-régulés et auto-apprenants »*.

Si un soir, vers 18 heures, le programme détecte que les Parisiens utilisent moins Internet que d'habitude, il modulera les affichages publicitaires en conséquence. Mieux : bientôt, le système pourra constater, en consultant un site de trafic routier,

qu'un gros embouteillage bloque la capitale, et en déduire que les Parisiens arriveront chez eux en retard, ce qui décalera leur usage d'Internet... Criteo, qui emploie déjà 160 mathématiciens, va en embaucher une centaine en 2014. Franck Le Ouay place la barre très haut : *« Nous devons réussir l'exploit de faire plus de prédictions avec la même masse de données. »*

Les expériences se multiplient. Le mathématicien Erick Alphonse, de l'université Paris-XIII, met au point un système baptisé Predictive Mix. Il étudie d'abord deux échantillons d'internautes : l'un verra une bannière publicitaire s'afficher sur son écran, l'autre non. En comparant le taux d'achat du produit dans chaque groupe, il quantifie l'efficacité du message.

Armé de ces premiers résultats, Predictive Mix répartit les profils en quatre groupes. L'internaute qui achète un produit sans voir de publicité est un « captif » : inutile de gaspiller de l'argent pour le convaincre ; celui qui achète quand il ne voit pas de publicité, mais cesse d'acheter quand il en voit est un « réfractaire » : il faut le laisser tranquille ; celui qui n'achète jamais rien est à oublier – trop difficile à convaincre ; enfin, celui qui n'achète rien quand il ne voit pas de publicité mais se met à acheter quand il en voit est un « réceptif », la cible la plus digne d'intérêt.

Etape suivante, l'ordinateur compare l'échantillon des « réceptifs confirmés » avec la population générale contenue dans une base de données. Tous les internautes dont le profil est similaire à ceux des « réceptifs » seront à leur tour classés comme tels. Et l'annonceur éliminera de sa campagne les trois autres groupes. *« Grâce à ce ciblage fin,* explique Erick Alphonse, *l'annonceur économise de l'argent, puisqu'il achète moins d'espaces tout en réalisant un meilleur retour sur investissement. »*

TAXONOMIE DU WEB

Une autre société française, Weborama, s'est lancée dans une aventure encore plus lourde : exploiter le « Web des mots ». Grâce à des programmes robotisés, elle collecte les textes publiés sur un vaste choix de sites et de forums. A partir de ces données brutes, les linguistes de Weborama ont extrait un

lexique de six mille termes pertinents dans le contexte de la publicité.

Dans un second temps, les mathématiciens organisent le Web comme un « espace métrique » : ils calculent la distance relative entre les mots, selon qu'ils sont plus ou moins souvent associés dans la même phrase. Puis ils rassemblent ces mots associés en 177 groupes thématiques – assurances, jeux d'argent, nourriture, sport, animaux domestiques… Le patron de Weborama, Alain Levy, résume ainsi son approche : *« Cette taxonomie devient notre vision du Web. La référence n'est plus le site, mais le mot. »*

On passe alors à l'exploitation commerciale. Grâce à des accords avec des agences, Weborama place des cookies sur des millions de navigateurs. Puis elle les piste à travers le Web, et collecte les mots publiés sur tous les sites qu'ils vont visiter : *« Chaque profil se voit ainsi attribuer un nuage de mots qui lui est propre »*, explique Alain Levy. Les ordinateurs vont ensuite projeter ce « nuage » sur la base de données contenant les groupes de mots, et attribuer à chaque profil une note par catégorie.

En croisant les notes – par exemple 13 sur 14 (le maximum) pour les mots associés à la mode, 12 pour le design, mais seulement 2 pour le sport, 1 pour les voitures – Weborama va pouvoir dire à l'annonceur qui se cache derrière chaque cookie: *« Ce sera par exemple une femme de 34 à 49 ans, passionnée de mode, indifférente au sport,* détaille Alain Levy. *Elle sera sans intérêt pour certains annonceurs et très désirable pour d'autres. L'Oréal sera prêt à payer 2 euros pour afficher une bannière sur son écran. »* Weborama possède à ce jour 62 millions de profils pour la France – il y a des doublons, car une même personne peut utiliser plusieurs appareils (PC, smartphone, tablette). La mise à jour est permanente, chaque clic provoquant de nouveaux calculs.

ESPOIR SUPRÊME

L'analyse prédictive envahit tous les secteurs d'activité. Des start-up se créent pour faciliter la migration de ces nouvelles compétences vers les industries traditionnelles. En France, la

société Dataiku a mis au point une suite logicielle qui permettra à des cadres sans formation informatique pointue de se lancer dans la gestion de bases de données et l'analyse prédictive : *« Nos clients potentiels,* affirme Florian Douetteau, patron de Dataiku, *sont les entreprises industrielles qui possèdent des stocks de données dont ils ne font rien, et qui veulent les exploiter pour résoudre des problèmes de façon innovante. »* Il cite comme exemple un gestionnaire de parcmètres souhaitant, à partir des données de stationnement, modéliser la circulation automobile dans des milliers de villes.

L'espoir suprême des chercheurs est que les ordinateurs donnent du sens à des données diffuses et chaotiques, livrées en vrac. En découvrant des modèles et des corrélations qu'aucun humain n'aurait imaginé, ils répondront à des questions que personne ne leur a posées.

De ce fait, le débat sur l'existence d'une « intelligence » chez ces ordinateurs auto-apprenants est déjà obsolète. Pour les professionnels du secteur, l’important n'est pas de savoir si la machine fonctionne comme un cerveau humain, mais de constater qu'elle obtient, par des voies différentes, des résultats égaux ou supérieurs à celui qu'aurait obtenu un humain.

Quant au rapport entre l'homme et la machine, divers penseurs américains l'ont résumé en une question que chacun devra bientôt se poser : votre savoir-faire est-il complémentaire du savoir-faire de votre ordinateur, ou votre ordinateur fait-il un meilleur travail sans vous ? De fait, les mathématiciens ont commencé à détruire certains métiers de la publicité – analystes, media-planners, etc. Reste à savoir si la montée en puissance de ces techniques de plus en plus invasives va entraîner une réaction des populations visées.

STRATÉGIE D’AUTO-DÉFENSE

Le philosophe William Bates rappelle que rien n'est jamais joué : *« Le fait même que notre cerveau soit malléable signifie que la technologie ne nous prédétermine pas entièrement. A certains moments historiques, nous pouvons décider de ce que nous allons devenir. Mais pour cela, il faut réfléchir et agir. Or, c'est*

peut-être ce qui nous effraie le plus. Si nous décidons que l'innovation est devenue incontrôlable, nous nous déchargeons de toute responsabilité, c'est plus confortable. »

De son côté, Melanie Swan, une créatrice de start-up californienne qui se définit comme une « philosophe de la technologie », remarque que de nombreux internautes commencent à déployer des stratégies d'auto-défense : *« Ils éparpillent leurs données sur plusieurs sites – leurs photos chez un prestataire, leurs e-mails chez un autre, leurs requêtes chez un troisième – dans l'espoir qu'aucune de ces sociétés ne pourra établir leur profil complet. Ce comportement est l'indice d'une "proto-sensibilité" à ce problème. Ils sentent que quelque chose ne va pas, mais restent impuissants. Mais c'est en train de changer. »*

Selon Melanie Swan, on assiste aux Etats-Unis à la naissance d'un mouvement intellectuel visant à inciter les internautes à devenir des sujets actifs dans cette affaire : *« Quand nous laissons une entreprise s'emparer de nos données personnelles, nous effectuons une transaction, nous livrons une matière première qui a de la valeur. Or, nous n'avons aucun pouvoir de négociation, nous acceptons les conditions imposées par l'industrie. »* La solution est évidente : *« Je pense que les internautes vont s'unir et s'organiser pour défendre leurs intérêts en tant que fournisseurs de données. Pour cela, ils vont s'inspirer des associations de défense des consommateurs, ou même des syndicats ouvriers. Seule une réponse collective et solidaire pourra rétablir l'équilibre. »* Si les internautes parviennent à changer le rapport de force avec l'industrie, ils pourront exiger d'être payés pour leurs données, ou imposer des conditions et des restrictions à leur usage. Pour les penseurs de la Silicon Valley, cette stratégie sera sans doute plus efficace que des lois imposées par les Etats, souvent en retard d'une guerre.

Yves Eudes (/journaliste/yves-eudes/)
Grand reporter
Suivre