

Le profilage culturel après le profilage publicitaire

CHARLES CUVELLIEZ

Ecole polytechnique de l'ULB

La réforme européenne sur la protection de la vie privée, annoncée le 18 décembre, évoque les risques de profilage sur internet, que tout le monde a expérimenté avec les publicités ciblées. Il en est un autre, bien plus intrusif, dont on parle peu: le profilage culturel, celui que nous laissons au hasard des musiques, jeux, vidéos que nous écoutons sur YouTube, Netflix et autres Steam.

Or ces choix culturels participent à la manière dont la pensée se forme chez l'individu. Ils touchent à notre identité et ils participent à l'image que nous voulons renvoyer (et contrôler), par exemple via les réseaux sociaux, pour manifester une «appartenance». Or ce profil culturel peut échapper à notre contrôle, dit la CNIL, la commission française pour la vie privée. La même CNIL avait déjà mis en garde contre le big data des capteurs qu'on porte sur soi: connaître le nombre de pas de Monsieur X telle journée est insignifiant, mais, étalé sur plusieurs années, il permettra d'établir son emploi du temps. Le profilage qui a commencé par un «si vous avez aimé ceci, vous aimerez...» n'est pas mauvais en soi. L'offre (et sa qualité ou diversité) a considérablement augmenté grâce à la recommandation. Elle évite le syndrome de ne plus savoir par où commencer quand l'abondance est là. Les playlists sont une étape plus loin, mais présentent un risque, celui de ne plus jamais nous faire découvrir autre chose que ce que nous avons déjà aimé.

Livre numérique

Mais il n'y a pas que l'individu qui est impacté dans sa diversité culturelle: le secteur culturel aussi. La création littéraire peut s'affaiblir et sa diversité affectée. Sait-on qu'Amazon rémunère les auteurs de livres numériques à la page lue depuis juillet dernier? Les éditeurs numériques auront les moyens de prédire si une page, un texte sera lu en agrégeant les données de lecture des internautes. On sait déjà que les best-sellers obéissent à de tels principes, hors internet. Internet et big data rendront ces prédictions de succès plus fiables via les statistiques de lecture: la vitesse à laquelle le livre est lu (ou abandonné), en combien de temps il est lu. Fini l'avant-garde et les révélations

La musique sur internet

Les playlists sont familières sur Spotify, Deezer ou Soundcloud, devant les dizaines de millions de titres disponibles. Les algorithmes sont capables de donner le génome des morceaux de musique. Pandora, le leader des chaînes radio par internet, a établi les 450 caractéristiques de ce génome. Ces mêmes algorithmes ont montré que c'est l'humeur de l'auditeur (fête, sport, détente, voyage, météo...) qui détermine la bonne playlist, pas le genre musical! Pour la CNIL, la compréhension des comportements et des activités de l'utilisateur prendra le dessus sur le contenu musical lui-même. Avec un tel profilage, ces plateformes vont connaître l'individu mieux que lui-même. Elles construiront une connaissance fine des utilisateurs qui n'est même pas partagée avec ces derniers, mais vendue à des tiers. Partager cette connaissance avec l'utilisateur final serait une manière de l'aider à vivre ce profilage intrusif et lui permettre de comprendre ses goûts.

Le profilage «culturel» est celui que nous laissons au hasard des musiques, jeux, vidéos que nous écoutons sur YouTube, Netflix et autres Steam.

L'audiovisuel

Ce qui est vrai pour la musique l'est pour la vidéo: Netflix a établi, pour son algorithme, pas moins de 77.000 microgenres (par exemple: «film d'époque, émouvant avec personnage féminin de 30 à 40 ans...»). Les critiques de cinéma sont recrutés par Canal Plus pour les transformer en prescripteurs. Et le profilage permet de déterminer combien d'épisodes d'une série le spectateur regarde par soirée pour éviter d'y placer la même musique par exemple.

Les jeux vidéo

Curieusement, c'est sur le jeu vidéo que le profilage est le moins évolué. Sur Steam, les créateurs de jeu déposent leur application et les internautes la téléchargent. Le jeu vidéo a pourtant un grand besoin de profilage: la plupart des jeux sont gratuits, mais on pousse les joueurs à l'addiction pour les forcer à acheter des accessoires payants pour aller plus loin ou plus vite dans le jeu. C'est le modèle dit du «freemium». L'interaction humain/jeu dépasse le stade du clavier et de la souris, c'est aussi un enjeu de profilage (via les capteurs de mouvement et autres Kinect): on pourrait

analyser le comportement des joueurs, vérifier ce qui plaît pour optimiser l'addiction. Plus le joueur reste connecté, plus l'éditeur peut le monétiser, ne fût-ce qu'avec ses accessoires ou passe-droits vers le niveau supérieur. Ces techniques de profilage sont peu pratiquées. Le meilleur ou le pire est donc encore à venir dans ce quatrième pilier «culturel».

La recommandation, outil fondamental pour la culture

La recommandation est essentielle à l'offre culturelle digitale: plus le choix est vaste, moins l'utilisateur choisit. La recommandation compense les défauts inhérents à la culture: on ne peut prévoir si un produit culturel plaira et on ne peut exiger un remboursement après. Mais un bien déjà apprécié par ses amis a plus de chances d'être aimé par soi, c'est une manière pour nous de minimiser le risque, un autre bienfait (une externalité) des réseaux. Enfin, la recommandation allonge la durée de vie commerciale d'une œuvre au-delà de son lancement. Jusqu'à présent, il y avait 4 types de recommandations: la plus classique, les critiques «humaines» d'avant internet, ensuite par analogie de contenu («tous les films de guerre de 40-45»), par analyse de l'utilisateur (recommander ce que les amis sur Facebook aiment) et le filtrage collaboratif: si deux utilisateurs ont aimé les trois mêmes films, le quatrième vu par l'un plaira à l'autre. Une nouvelle méthode a vu le jour: le «learning to rank» qui combine l'historique de l'utilisateur, le filtrage collaboratif et sa manière de consommer (quand, comment, compulsif...), tout comme les publicités ciblées analysent le comportement de l'internaute pour lui envoyer la bonne pub.

La personnalisation de l'œuvre est l'étape d'après: imaginez des œuvres musicales inachevées qui se complèteraient d'elles-mêmes en fonction du moment ou de l'état d'esprit avec lequel vous l'écoutez ou du livre numérique dont l'histoire s'auto-complèterait. Et on a déjà vu de telles expérimentations avec les webseries TV qui offrent des histoires à tiroir et interactives.

C'est l'opacité des algorithmes qui pose problème, pas leur valeur ajoutée: à ne pas savoir comment ces algorithmes nous classent, nous profilent et pré-déterminent ce que nous allons aimer ou pas, c'est nous enfermer dans des bulles qui résultent des filtres que ces algorithmes tissent autour de nous. Sans opacité, cela devient un choix en connaissance de cause et il se trouvera toujours des start-ups qui vous proposeront un nouvel algorithme du type: «Vous n'avez jamais testé cela, vous devriez.»

Pour en savoir plus: dans les «Cahiers IP et Prospectives», n° 3, CNIL, «Les données, muses et frontières de la création.»