

Algorithmes de recommandation et libre arbitre

A l'heure du numérique, l'explosion de la quantité d'information accessible a conduit à la nécessité d'implémentation d'algorithmes de recommandation qui filtrent le contenu proposé à chacun. Cependant, ces algorithmes soulèvent des enjeux impliquant individu comme communauté dans son rapport au numérique : manque de contraste lié à une information ciblée et création de comportements culturels grégaires et normalisés.

I Fonctionnement des algorithmes de recommandation

I.1 Collection de données

Comment les algorithmes de recommandation filtrent et collectent nos données ?
Avant de pouvoir nous proposer une liste de recommandations, un algorithme cherche à collecter un certain nombre d'informations pour caractériser et donner un profil à l'utilisateur.
Deux grands type de collecte de données se démarquent :
La collecte de données explicite : Ce type de collecte va directement recueillir les informations à l'utilisateur, par exemple en prenant en compte ce qu'il a pu "liker" sur certains réseaux, s'il a partagé tel ou tel contenu... Ce type de collecte, contrairement au second cité ci-dessous, a l'avantage d'identifier clairement l'utilisateur, de pouvoir faire la différence entre plusieurs personnes qui utiliseraient des applications à partir du même téléphone ou ordinateur.
//on peut citer le biais de déclaration comme inconvenient , encore quelque recherche a faire la dessus//
La collecte de données implicites : Les données de l'utilisateur vont être collectées sans même que celui ne s'en rende compte, de manière implicite. Cette collecte se base sur l'analyse des positions GPS pour certaines applications comme Snapchat, Instagram ou encore Tinder, sur les listes de musique écoutées par l'utilisateur pour des plateformes musicales comme Deezer ou Spotify, ou encore sur l'analyse de la fréquence à laquelle un utilisateur se connecte sur une page. Bien que ces données soient facilement collectées, il n'en est pas moins facile d'attribuer ces données à un utilisateur comme cité plus haut.

I.2 Stratégies de recommandation

Les données une fois collectées, il vient de les exploiter afin de recommander le meilleur contenu. On distingue trois méthodes de recommandation, se démarquant par les catégories de données utilisées principalement.

I.2.1 Basée sur le contenu

Ici, on s'intéresse essentiellement au contenu consommé par l'utilisateur. Il s'agit de proposer à l'utilisateur des produits similaires à ceux précédemment consommés.
L'algorithme s'intéresse à des informations pouvant déterminer à quel point le produit a plu. Par exemple, le temps passé sur une vidéo Youtube, le nombre d'écoutes d'une musique sur Spotify, ou encore un commentaire laissé sur un produit acheté sur Amazon. Sur les plateformes ou ceci est possible, des informations importantes se trouvent également directement dans le profil de l'utilisateur. C'est une des raisons participant au fait que de plus en plus de plateformes demandent ou proposent d'avoir un profil pour profiter pleinement de son expérience d'utilisateur.

L'algorithme tente, en intersectant grand nombre de ces données, de proposer un produit susceptible de plaire. C'est la politique de "Vous avez aimé ceci, vous aimerez sûrement cela..."

Une telle stratégie présente des nombreux avantages. Par exemple, elle tient compte uniquement de l'expérience l'utilisateur seul, c'est-à-dire que le profil qu'elle construit dépend uniquement de celui qui consomme. Elle n'aura donc pas de mal à proposer des produits assez atypiques, comme le cas de quelqu'un écoutant un style de musique peu populaire sur Spotify.

En revanche, la complexité à recommander augmente avec le nombre de produits consommés, ce qui n'est pas une bonne chose. Les profils deviennent de plus en plus difficiles à dresser, à mesure que les centres d'intérêts se multiplient.

De plus, il est très difficile de recommander un nouveau profil, n'ayant pas encore consommé. Cette stratégie est extrêmement dépendante du temps investi par l'utilisateur sur la plateforme.

I.2.2 basée sur le filtrage collaboratif

Ce type d'algorithme s'intéresse à regrouper des utilisateurs proches. Il y a deux manières de rapprocher deux utilisateurs. D'une part, il est possible de rapprocher deux utilisateurs par leur consommation : de manières similaire à la recommandation basée sur le contenu, deux utilisateurs seront proches s'ils consomment les mêmes produits.

On peut également rapprocher deux utilisateurs en effectuant des similitudes entre leur profil, comme la situation géographique par exemple, ou tout simplement par une liste d'amis ou d'abonnés lorsque la plateforme y a accès, Facebook et Twitter étant de bons exemples. Cela pousse donc certaines structures ne se présentant pas réellement comme un réseau social à se développer comme tel, il est par exemple possible de s'abonner au profil des utilisateurs sur Spotify.

Ces rapprochements faits, l'algorithme va alors pouvoir proposer des contenus consommés par des personnes similaires à l'utilisateur. on peut résumer cela par la phrase "un tel a aimé ceci, vous devriez aimer cela..."

A nouveau, cette méthode a ses propres avantages et inconvénients. Elle se base sur une plus grande diversité de données, et donc peut proposer facilement plusieurs type de contenus. Cependant, elle peut proposer des contenus pas toujours adaptés. En effet, il n'est pas garanti que deux utilisateurs ayant les mêmes goûts sur un sujet s'entendent sur d'autres, et pourrait donc proposer des produits moins susceptibles de plaire.

I.2.3 Recommandation de contenu populaire

Cette méthode consiste à recommander les contenus les plus populaires, la définition de ce terme changeant en fonction de chaque algorithme.

Par exemple, pour l'algorithme PageRank de Google, il s'agit de proposer les pages webs les plus référencées par d'autres. Concrètement, la page se voit attribuer un score, en fonction du nombre de fois qu'un utilisateur se déplacement de manière aléatoire sur le web rencontrerait la page en question.

Sur Youtube, un bon indicateur de la popularité d'une vidéo peut simplement être perçue comme le nombre de vues et de " mentions j'aime" laissées sur la vidéo.

On peut supposer que la fiabilité à être apprécié et la qualité d'un contenu augmentent avec sa popularité, mais cette méthode est moins efficace pour recommander un utilisateur ayant des préférences de contenus moins populaires.

I.2.4 Méthode hybride

La plupart des algorithmes actuels n'utilisent pas que une de ces stratégies mais les combinent, afin d'accroître leur efficacité. Il s'agit de palier aux inconvénients de chacune pour optimiser la recommandation en proposant un produit toujours plus propice de plaire. Ainsi, dans certaines situations il sera plus judicieux de se baser sur le filtrage collaboratif, dans le cas d'un utilisateur ayant peu consommé, mais présentant de fortes similitudes avec d'autres profils. De même, une personne consommant uniquement un certain type de produit n'aura pas forcément envie de se faire recommander un produit ne correspondant pas à ses goûts. La lecture automatique sur Youtube en est un très bon exemple : elle propose en même temps des contenus similaires aux précédents, mais tend à converger tout de même vers les contenus les plus populaires. De plus étant un choix par défaut à enlever, elle "nudg" l'utilisateur en le poussant à visionner un nouvelle vidéo.

2 Conséquences

2.1 Désinformation/Accès à une information variable selon l'individu

ALEX

2.2 Désir, Stiegler

Un des problèmes posés par le processus d'automatisation et d'algorithmisation de notre rapport au numérique est celui de l'individualisation : en nous donnant la voie vers un monde non seulement plus vaste mais aussi plus facile d'accès et dont la navigation est plus aisée, est-il plus facile de s'émanciper avec le numérique ?

Effectivement, l'internet nous donne accès au monde mais aussi donne au monde l'accès à nous : par le biais des notifications par exemple, il est bien plus facile d'attirer l'attention d'un individu et c'est une conséquence de la numérisation du monde que les entreprises ont transformé en atout, comme en atteste cette citation fameuse du directeur de TFI, Patrick Lelay : « Soyons réalistes : le business de TFI, c'est de vendre du temps de cerveau disponible à Coca Cola ».

Selon Bernard Stiegler, philosophe français, on assiste avec l'algorithmisation du numérique à une forme de « marketing de la libido » : les entreprises cherchent par tous les moyens à capter notre attention (par exemple par la télévision, par les publicités) afin de créer en nous le besoin de consommer le produit. Par ce biais, le désir de l'individu ayant accès à internet est façonné par les industries : on crée le désir de consommer en présentant le produit comme le moyen de satisfaire ce dernier. Et ce phénomène, antérieur à l'apparition du numérique, a été facilité par l'arrivée de ce dernier : par l'apparition de produits dématérialisés et par des biais algorithmiques, la consommation moderne repose sur des principes de recommandation par des pairs ; par exemple, lors de l'achat d'un livre sur internet : « Vous avez aimé ce livre, vous aimerez sans doute aussi celui-là ». Cela présente deux avantages : de un, les individus eux-mêmes sont les propres victimes de la société de consommation puisqu'ils sont actifs dans le processus de création de leur propre désir ! De plus, ces algorithmes conduisent à l'émergence d'un désir unifié puisque l'individu est redirigé vers des produits qu'il est censé aimer (ce que l'algorithme calcule et prévoit). Ainsi, affirme Bernard Stiegler, on assiste à un véritable processus de dé-individualisation par les algorithmes...

3 Que faire contre ?

3.1 Solutions globales

3.2 Alternatives individuelles

DuckDuckGo préserve la vie privée et ne stocke aucune information personnelle concernant les utilisateurs (adresse IP et traces numériques comme la signature de navigateur). Le principe de ce moteur de recherche et de ne pas limiter les utilisateurs à une "bulle" d'information qui serait définie de manière personnalisée pour chaque individu (comme le font la plus part des autres moteurs de recherche) mais bien de permettre à tous d'accéder au Web tout entier. DuckDuckGo a pour fonctionnalité de proposer les mêmes résultats non filtrés à tous les utilisateurs en leur laissant le choix de régler des paramètres explicites: spécialisation géographique des résultats, masquage ou non de résultat adulte etc... Cela permet une meilleure sérénité (le fait de réaliser une découverte scientifique de façon inattendue, très souvent dans le cadre d'une recherche sur un autre sujet).

« la représentation que se fait un individu de lui-même est construite sur une somme de contradictions et d'idéaux. Nous avons la possibilité de changer en permanence. Or ces algorithmes détruisent notre singularité en créant un double de nous-même qui n'est qu'une image statistique." Bernard Stiegler

Quelles sont les solutions pour faire valoir son "droit à l'information" et volontairement sortir de sa bulle de filtre ?
Rencontrer des points de vue différents (opposés) au siens.
Mark Zuckerberg a minimiser l'impact et la portée de ce phénomène sur Facebook mais l'entreprise commence tout de même à travailler sur des mesures pour limiter cet impact.
Il existe aussi des extensions sur différents moteurs de recherche qui permettent de mettre en lumière les bulles de filtres.
En France, le journal Le Drenche propose pour tous les sujets deux avis opposés (cependant cela pourrait tendre à privilégier les extrêmes au détriment des points de vue nuancés).

Critique de cette pensée (« le système de sélection de Facebook ne modifie que de 1 % l'exposition aux contenus politiques de camps opposés » André Gunther dans "Et si on arrêtait avec les bulles de filtre ?", et Dominique Cardon : « la bulle, c'est nous qui la créons. Par un mécanisme typique de reproduction sociale. Le vrai filtre, c'est le choix de nos amis, plus que l'algorithme de Facebook. ».

De nos jours, une solution pour pallier la concentration et au manque de diversité des contenus multimédia, bien trop souvent limités par des algorithmes de recommandation, serait de proposer des ressources nouvelles aux consommateurs, par le biais de site mettant en avant ces dernières. Un exemple concret et parlant serait la plateforme Forgotify, qui regroupe plus de 4 millions de titres oubliés et méconnus sur Spotify. Cette solution a pour but de sortir les auditeurs d'un enfermement autour d'un style musical, de playlist proposées par les algorithmes souvent redondantes. Finalement cette plateforme permet de découvrir du contenu original, tout en permettant aux auteurs de gagner en visibilité.
Dans une optique similaire, Soundcloud est une plateforme de musique qui se veut très facilement accessible, et permet à tout type d'artiste de proposer son contenu gratuitement. Une particularité de ce média social est qu'il est très facile d'y trouver un contenu nouveau, proposé par des artistes souvent inconnus, des remix. En contrepartie les musiques et albums les plus populaires y seront souvent absents.