Pourquoi cette vidéo ?

Les systèmes de recommandation et la bulle de filtres

Mise en situation. Sur une plateforme de vidéos, un réseau social ou un site de musique, la vidéo suivante démarre toute seule, et le fil d'actualité semble deviner nos goûts. Ce choix est fait par une IA : un **système de recommandation**.
Dans cette activité, vous allez découvrir comment une IA choisit ce qu'elle vous montre, programmer une petite recommandation, puis voir comment ces choix peuvent, peu à peu, enfermer chacun dans ses propres goûts.

Remarque

En 2018, le directeur des produits de YouTube a indiqué que plus de 70 % du temps passé à regarder des vidéos sur la plateforme venait de ses recommandations, et non des recherches des utilisateurs.

1. Des utilisateurs qui se ressemblent

Définition

Un **système de recommandation** propose à un utilisateur des contenus qu'il n'a pas encore vus. Une méthode très utilisée, le **filtrage collaboratif**, repose sur une idée simple : « si quelqu'un a les mêmes goûts que vous, ce qu'il a aimé vous plaira sans doute ».
Pour mesurer la ressemblance entre deux utilisateurs, on calcule leur **similarité** : le nombre de vidéos qu'ils ont **tous les deux** aimées.

Voici les goûts de cinq utilisateurs d'une plateforme :

	Foot	Jeux vidéo	Cuisine	Science	Musique	Humour
Léa	✓	✓	?	?	?	✓
Hugo	✓	✓		✓		✓
Tom	✓	✓		✓	✓	
Inès			✓		✓	✓
Zoé			✓	✓		

✓ : vidéo aimée ; case vide : vidéo pas aimée ou pas regardée ; ? : vidéo que Léa n'a pas encore vue.

Exemple

Léa et Tom ont tous les deux aimé « Foot » et « Jeux vidéo » : leur similarité vaut 2.

▶ À vous de jouer !

1. Calculer la similarité de Léa avec Hugo, avec Inès, puis avec Zoé.
2. Quel utilisateur ressemble le plus à Léa ?
3. Quelles vidéos cet utilisateur a-t-il aimées, que Léa n'a pas encore vues ? Laquelle la plateforme va-t-elle recommander à Léa ?

Coup de pouce

Pour Hugo, regardez colonne par colonne : y a-t-il une coche à la fois sur la ligne de Léa et sur celle d'Hugo ?

2. Un score pour chaque vidéo

Définition

Se fier à un seul utilisateur est risqué. On calcule plutôt un **score** pour chaque vidéo que Léa n'a pas vue : on additionne les similarités de tous les utilisateurs qui ont aimé cette vidéo. La vidéo de plus grand score est recommandée en premier.

▶ À vous de jouer !

1. Calculer le score de la vidéo « Cuisine », puis de « Science » et de « Musique ».
2. Dans quel ordre les trois vidéos seront-elles proposées à Léa ?
3. Pourquoi « Musique » obtient-elle un meilleur score que « Cuisine », alors que chacune a été aimée par deux personnes ?

Coup de pouce

« Cuisine » a été aimée par Inès (similarité 1) et Zoé (similarité 0) : son score vaut $1 + 0$.

3. Programmer la recommandation en Python

Exemple

On code les goûts de chaque utilisateur par une liste de 1 (vidéo aimée) et de 0, dans l'ordre Foot, Jeux vidéo, Cuisine, Science, Musique, Humour.

```
lea  = [1, 1, 0, 0, 0, 1]
hugo = [1, 1, 0, 1, 0, 1]
tom  = [1, 1, 0, 1, 1, 0]
ines = [0, 0, 1, 0, 1, 1]
zoe  = [0, 0, 1, 1, 0, 0]

def similarite(u, v):
    n = 0
    for i in range(len(u)):
        if u[i] == 1 and v[i] == 1:
            n = n + 1
    return n
```

▶ À vous de jouer !

1. Que renvoient `similarite(lea, hugo)` et `similarite(lea, zoe)` ?
2. Donner les valeurs successives de `n` lors de l'appel `similarite(lea, tom)`.
3. Écrire une fonction `score(j)` qui renvoie le score de la vidéo d'indice `j` pour Léa. Par exemple, `score(3)` doit renvoyer 5.

Coup de pouce

Pour la question 3, on peut utiliser la liste `autres = [hugo, tom, ines, zoe]` et une boucle `for u in autres:`. L'utilisateur `u` a aimé la vidéo `j` si `u[j] == 1`.

4. La bulle de filtres

Définition

Une plateforme cherche à garder ses utilisateurs le plus longtemps possible, car elle gagne de l'argent avec la publicité. Elle montre donc surtout ce sur quoi l'utilisateur a déjà cliqué. Peu à peu, il ne voit plus que des contenus qui confirment ses goûts et ses opinions : il est enfermé dans une **bulle de filtres** (expression popularisée en 2011 par Eli Pariser).

Exemple

Un jour donné, la plateforme montre 100 vidéos à Léa, dont 20 de foot. Léa clique sur 80 % des vidéos de foot et sur 40 % des autres vidéos. Le lendemain, la plateforme montre la même part de foot que dans les clics de la veille.

▶ À vous de jouer !

1. Le premier jour, combien de vidéos de foot Léa regarde-t-elle ? Combien d'autres vidéos ? Quelle est la part du foot dans ses clics ?
2. On admet que si la part du foot affichée un jour vaut p, elle vaut $\dfrac{2p}{1+p}$ le lendemain. Vérifier ce résultat pour $p = \dfrac{1}{5}$.
3. Calculer la part du foot les jours suivants, sous forme de fractions, jusqu'au jour 5. À partir de quel jour dépasse-t-elle 85 % ?
4. Que devient le fil de Léa ? Pourquoi est-ce un problème ?

Coup de pouce

Pour $p = \dfrac{1}{3}$: $\dfrac{2p}{1+p} = \dfrac{2/3}{4/3} = \dfrac{2}{3} \times \dfrac{3}{4} = \dfrac{1}{2}$.

Exemple

En Python, on peut simuler plusieurs jours :

```
def bulle(p, jours):
    for j in range(jours):
        p = 2 * p / (1 + p)
    return p
```

`bulle(0.2, 4)` renvoie 0.8 : après 4 jours, 80 % des vidéos affichées sont du foot.

5. Les enjeux

Remarque

Depuis 2023, le règlement européen sur les services numériques (*Digital Services Act*) oblige les très grandes plateformes à expliquer les grands principes de leurs recommandations, et à proposer au moins un fil qui ne repose pas sur le profil de l'utilisateur (par exemple, un fil chronologique).

▶ À vous de jouer !

1. Pourquoi une plateforme gratuite a-t-elle intérêt à ce que vous restiez le plus longtemps possible ?
2. Proposer trois gestes pour sortir de sa bulle de filtres.

6. Pour aller plus loin

★ Question étoilée

Pour limiter la bulle, la plateforme décide que 20 % des vidéos affichées seront choisies au hasard, avec une part de foot de 20 %. Les 80 % restants suivent les clics de la veille. La part du foot passe alors de p à

$$0{,}8 \times \frac{2p}{1+p} + 0{,}2 \times 0{,}2.$$

1. Partant de $p = 0{,}2$, calculer la part du foot aux jours 1, 2 et 3 (arrondir au millième).
2. Modifier la fonction `bulle` pour tenir compte de cette règle. Avec un ordinateur, on trouve environ 0,693 au jour 10 et 0,697 au jour 30. Comparer avec la situation sans hasard, et conclure.

À retenir

- Un **système de recommandation** choisit ce qu'une plateforme vous montre. Le **filtrage collaboratif** recommande ce qu'ont aimé les utilisateurs qui vous ressemblent.
- La ressemblance se mesure par une **similarité** ; chaque contenu reçoit un **score**, et les meilleurs scores sont proposés en premier.
- En montrant surtout ce sur quoi on a déjà cliqué, les recommandations peuvent enfermer dans une **bulle de filtres**.
- Les plateformes gratuites vivent de la publicité : leur but est de capter notre attention. On peut sortir de sa bulle en variant ses sources et en choisissant des fils non personnalisés.