

# 1 L'Histoire

Un système de recommandation est un moyen mis en place pour proposer un contenu spécifique à un utilisateur. Ce contenu est souvent associé aux goûts de l'utilisateur afin de faciliter ce dernier dans ses recherches ou de lui proposer un contenu inconnu en adéquation avec ses préférences.

L'essor du numérique lors de la seconde moitié du XXème siècle a favorisé le développement de ces systèmes de recommandation. Au départ ces principaux systèmes utilisaient le principe d'indexation (Van Rijsbergen, 1979) qui consistait à associer chaque objet à un label. Il était alors aisé de trouver le contenu associé à un label. 1989 vit la création du Web et avec lui une multitude d'algorithmes permettant de répondre aux demandes des utilisateurs, les moteurs de recherche. Néanmoins ces derniers utilisaient le fameux principe d'indexation et avec la multiplication des sites web, la qualité du service rendu par les moteurs de recherche était discutable, tant que sa quantité que sur sa pertinence .

Il fallut attendre 1992 et les travaux de Paul Resnick et John Riedl pour voir apparaître le premier système de recommandation collaboratif. Ce système utilisait un principe simple mais fondamental de notation. L'idée principale est que l'utilisateur va associer des notes au contenu. Par exemple, si une personne A et une personne B ont mis les mêmes notes à des contenus similaires, alors il est fortement probable que la personne A aimera tout ce que la personne B a aussi aimé. On peut donc recommander efficacement la personne A sur ces futurs choix.

Les années suivantes, plusieurs systèmes de recommandation se sont différenciés. Ils apportaient chacun des avantages et des inconvénients, tant dans la pertinence des résultats que dans le temps d'exécution ou dans la place mémoire. La nuance entre les différents systèmes de recommandation réside dans la manière qu'ils ont d'utiliser ces informations. Ainsi certains se focaliseront uniquement sur les goûts de l'utilisateur tandis que d'autres compareront les goûts des utilisateurs entre eux. D'autres systèmes de recommandation ne demanderont pas explicitement l'avis de l'utilisateur mais récolteront des informations sur l'activité de ce dernier. Il est donc primordiale d'étudier et de comparer entre eux chacun de ces systèmes pour pouvoir ensuite proposer le système de recommandation le plus pertinent dans la réalisation de notre projet.

## 2 Recommandation personnalisée et objet

### 2.1 Principe

Afin de d'illustrer nos propos, nous nous placerons dans le contexte où le système de recommandation propose un contenu cinématographique à l'utilisateur.

La première étape d'un système de recommandation personnalisée (et même un système de recommandation en général) est la collecte d'informations. Ces informations sont récoltées directement auprès de l'utilisateur. Le fait de noter un film ou de laisser un commentaire est une information active. L'utilisateur exprimant explicitement ses choix. Les recherches que peut faire l'utilisateur, ainsi que le temps passé sur une page web sont des informations passives. Salton et Buckley ont même soutenu en 1988 que l'absence de données peut aussi, dans une certaine mesure, permettre d'appréhender les goûts de l'utilisateur. Les informations peuvent être mises sous forme matricielle afin de faciliter la compréhension.

	Film1	Film2	Film3	Film4
Note	5	2		
Visionné	Oui	Oui	Oui	Non
Pourcentage visionné	100	75	25	
Ignoré	Non	Non	Non	Oui

Une fois la collecte d'informations achevée, les systèmes de recommandation personnalisée vont traiter les traiter. Ils utilisent alors les informations récoltées pour un utilisateur en particulier et les mettent en corrélation avec du contenu que l'utilisateur n'a pas encore vu. Ce contenu est lui même détaillé afin de pouvoir être comparé. Les films peuvent par exemple renvoyer à un réalisateur, à des acteurs ou à des genres. Encore une fois un modèle matriciel permet un représentation claire.

	Film1	Film2	Film3	Film4
Auteur	Nom1	Nom2	Nom3	Nom1
Policier	Oui	Non	Non	Oui
Comique	Non	Oui	Non	Oui
Romance	Non	Oui	Oui	Non
Drama	Oui	Non	Non	Oui

Salton et Buckley ont proposé en 1988 de pondérer les caractéristiques de chaque contenu. Ainsi certaines caractéristiques sont jugées plus importantes que d'autres. Ces pondérations proviennent d'une approche subjective de la personne qui met au point le système de recommandation. On peut donc penser qu'un utilisateur privilégiera le genre d'un film qu'un acteur en particulier.

Un autre célèbre système de recommandation est la méthode de retour de pertinence de Rocchio. Il permet à l'utilisateur de juger lui même si les recommandations qui lui sont faites sont pertinentes. Ainsi cette méthode prendra en compte l'avis de l'utilisateur pour adapter ses recommandations.

## 2.2 Exemples

- L'un des projets pionniers en matière de recommandation individuelle a été INFOSCOPE (Fischer and Stevens, 1991). Ce système permettait la création de groupes de discussion qui correspondraient au profil des utilisateurs. Ce système se décompose en quatre étapes :

- L'interaction de l'utilisateur avec les groupes de discussion existants
- L'utilisateur aime un groupe s'il lit un certain nombre de messages de ce dernier.
- Lorsque la base de donnée est suffisamment importante l'utilisateur se voit proposer des groupes
- Si un groupe recommandé ne lui plaît pas, il peut le signaler et le système s'adaptera.

- Les annonces publicitaires d'internet (Goolge, Facebook,...) utilisent les actions actives et passives de l'utilisateur. Se sont des sytèmes de recommandation personnalisés

- Certaines structures multimédia utilisent des facettes de la recommandation personnalisée. Cette utilisation n'est cependant jamais le coeur du système de recommandation de ces sites.

## 2.3 Synthèse

Les systèmes de recommandation personnalisée ont l'avantage de proposer des résultats pertinents qui ne demande pas aux systèmes de connaître les particularité du contenu qu'ils traitent. Ils ont cependant besoin de récolter des informations auprès de l'utilisateur. Un nouvel utilisateur ne pourra donc pas se voir recommander un contenu. Les systèmes de recommandation personnalisés sont actuellement très peu utilisés, on leur préférera d'autres systèmes jugés plus pertinents et qui répondent plus facilement au problème de temps et de place mémoire.

Ils ont cependant posé les bases de tous les autres systèmes de recommandation et continuent d'être utilisés dans de rares cas.

# 3 Recommandation coopérative

## 3.1 Principe

Les systèmes de recommandation coopérative, ou coopération sociale, recueillent les préférences des utilisateur (au travers de notes par exemples) et corrélent les goûts des utilisateurs entre eux. L'idée centrale de ces systèmes est que si deux personnes aiment le même contenu dans le passé, elles aimeront le même dans le future. Il existe deux différents type de recommandation sociale :

- memory-based
- model-based

La recommandation coopérative utilise la notions de **profils utilisateur**. Un profil utilisateur correspond à une entité censée représenter un ensemble d'individus ayant les mêmes goûts. Ainsi le système de recommandation associe chaque utilisateur à un profil utilisateur. Dès lors le système proposera les mêmes films à tous ceux associés à ce profil. Cette méthode a pour avantage de ne pas nécessiter de connaissance précise sur le contenu traité, mais uniquement sur les goûts/personnalités des utilisateurs.

## 3.2 Les différentes méthodes de recommandation coopérative

### 3.2.1 La méthode memory-based

La méthode memory-based compare les notes (ou préférences) des utilisateurs entre eux. Pour cela les notes de chaque utilisateur vont tout d'abord être stockées d'comme dans le modèle ci-dessous :

	Film1	Film2	Film3	Film4
Utilisateur1	5		2	4
Utilisateur2	2	4	4	2
Utilisateur3	2	4		1
Utilisateur4	4	2		5
Utilisateur5	2	2	2	

Si l'on cherche à savoir si l'utilisateur 4 peut aimer le film 3, le système va faire une moyenne pondérée avec les différentes notes des autres utilisateurs. Cette pondération provient de la corrélation des goûts entre les différents individus. Deux méthodes de pondérations sont actuellement largement utilisées :

*Soit  $V_{ij}$  la note donnée à l'utilisateur  $i$  pour le film  $j$  ( $V_{ij} = ?$  si l'utilisateur n'a pas noté le film). Soit  $U_{ij}$  la corrélation entre les utilisateurs  $i$  et  $j$ .*

- Le coefficient de corrélation de Pearson renvoie un résultat entre -1 et 1. Plus le résultat est proche de 1, plus la corrélation entre les utilisateurs  $i$  et  $k$  sera importante. Le coefficient de corrélation de Pearson se calcule de la manière suivante :

$$U_{ij} = \frac{\sum_{k=0}^m (V_{ik} - V_i)(V_{jk} - V_j)}{\sqrt{\sum_{k=0}^m (V_{ik} - V_i)^2 \sum_{k=0}^m (V_{jk} - V_j)^2}} \quad (1)$$

- La similarité par cosinus renvoie elle aussi un coefficient entre -1 et 1. Son expression est la suivante :

$$U_{ij} = \cos(U_i, U_j) = \frac{\sum_{k=1}^m V_{ik} V_{jk}}{\sqrt{\sum_{k=1}^m V_{ik}^2 \sum_{k=1}^m V_{jk}^2}} \quad (2)$$

*Il est à noter que bien que moins utilisé, d'autres coefficient de corrélation existent, telle que la corrélation de Ringo ou le coefficient de corrélation de Spearman.*

La note théorique que l'utilisateur va donner à un film  $k$  qu'il n'a pas vu est alors estimée par la formule :

$$V_{ik}^* = K \sum_{V_{jk} \neq ?} U_{kj} V_{jk} \quad (3)$$

**L'exemple ci-dessous provient d'un article Steeve Huang publié sur Hackernoon.**

Le système de recommandation va au départ calculer la similarité entre l'utilisateur  $E$  et les autres individus. L'expression **NA** signifie que les utilisateurs n'ont pas assez de films en commun pour être comparés.

	The Avengers	Sherlock	Transformers	Matrix	Titanic	Me Before You	Similarity(i, E)
A	2		2	4	5		NA
B	5		4			1	
C			5		2		
D		1		5		4	
E			4			2	1
F	4	5		1			NA

L'étape suivante est de calculer une note estimative pour les films que l'utilisateur  $E$  n'a pas encore visionné.

	The Avengers	Sherlock	Transformers	Matrix	Titanic	Me Before You	Similarity(i, E)
A	2		2	4	5		NA
B	5		4			1	0.87
C			5		2		1
D		1		5		4	-1
E	3.51*	3.81*	4	2.42*	2.48*	2	1
F	4	5		1			NA

Le principal problème rencontré par cette méthode apparaît lorsqu'il faut traiter des millions d'individus. La méthode des "meilleurs voisins" répond à cette problématique. Elle consiste à sélectionner les individus ayant la plus forte corrélation avec l'utilisateur étudié. Un seuil peut tout simplement décidé afin

de délimiter les utilisateurs "proches" des utilisateurs "éloignés". Afin d'obtenir les résultats les plus probants, plusieurs études ont montré que le nombre de "meilleurs voisins" est optimal lorsqu'il se situe entre 20 et 50.

### 3.2.2 La méthode model-based

Contrairement à la méthode **memory-based** qui compare les utilisateurs entre eux, la méthode **model-based** compare le contenu. Cette méthode établie dans notre exemple un coefficient de corrélation entre les différents films. La formule de similarité par cosinus (définie ci-dessus) est largement employée pour cette méthode.

La méthode **model-based** permet d'éviter les problèmes mémoires de la méthode **memory-based**.

**En poursuivant l'exemple de l'article de Steeve Huang publié sur Hackernoon.**

Dans le cas où l'on souhaiterait estimer les notes que pourrait avoir *Me Before You*, le tableau précédent se remplirait comme ci-dessous :

	The Avengers	Sherlock	Transformers	Matrix	Titanic	Me Before You
A	2		2	4	5	2.94*
B	5		4			1
C			5		2	2.48*
D		1		5		4
E			4			2
F	4	5		1		1.12*
Similarity	-1	-1	0.86	1	1	

Si la méthode précédente remplissait notre tableau ligne par ligne (utilisateur par utilisateur), le **model-based** remplit le tableau colonne par colonne (film par film).

## 3.3 Synthèse

Les systèmes de recommandation sociale étudient les comportements des masses afin d'estimer les préférences des individus. Les profils d'utilisateurs permettent ainsi aux recommandations d'être plus optimales que celles des systèmes de recommandation personnelle. Les résultats probants ainsi que l'adaptabilité de ces méthodes sur les différentes plateformes ont permis aux systèmes de recommandation sociale de se développer grandement.

Certains problèmes persistent néanmoins. Un nombre trop important d'utilisateurs peut entraîner des problèmes de mémoire. Les nouveaux utilisateurs auront une

recommandation peut fiable et il est peu probable que tout le contenu soit noté. De plus, le système de note dépend des utilisateurs, or sur une grande quantité d'utilisateurs, seul une minorité note régulièrement le contenu qu'ils apprécient (ou déprécient).