

Systèmes de recommandation

Jill-Jênn Vie

26 août 2016

Recommandation d'articles

Inspirés par les tendances générales de vos achats



Préparez l'été avec Nivea [Voir plus](#)



Bjorg: nos meilleures ventes [Découvrir](#)



Jonak

Meilleures ventes de sièges auto



[Voir plus](#)

Meilleures ventes de Mobiles pour bébé

Recommandation d'articles (fenêtre privée)

Les plus demandés dans DVD & Blu-ray [Voir plus](#)



Notre sélection de suspensions [Voir plus](#)

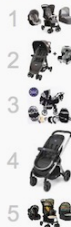


Notre sélection de tabourets [Voir plus](#)



Fourches de vélo
notre sélection

Meilleures ventes de poussettes Trio



[Voir plus](#)

Meilleures ventes de Balancelles



Recommandation d'articles (fenêtre privée)

Les plus demandés dans DVD & Blu-ray [Voir plus](#)



Notre sélection de suspensions [Voir plus](#)

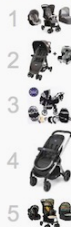


Notre sélection de tabourets [Voir plus](#)



Fourches de vélo
notre sélection

Meilleures ventes de poussettes Trio



[Voir plus](#)

Meilleures ventes de Balancelles



Filtrage collaboratif



| | | | | |
|--------|---|---|---|---|
| Sacha | ? | 5 | 2 | ? |
| Ondine | 4 | 1 | ? | 5 |
| Pierre | 3 | 3 | 1 | 4 |
| Joëlle | 5 | ? | 2 | ? |

Filtrage collaboratif



| | | | | |
|--------|---|---|---|---|
| Sacha | 3 | 5 | 2 | 2 |
| Ondine | 4 | 1 | 4 | 5 |
| Pierre | 3 | 3 | 1 | 4 |
| Joëlle | 5 | 2 | 2 | 5 |

Algorithme des plus proches voisins

Pour recommander des films à quelqu'un :

- ▶ On introduit un **score de similarité** entre personnes
- ▶ On détermine les 10 personnes **les plus proches** de lui
- ▶ On lui recommande ce qu'ils ont aimé qu'il n'a pas vu

Nos données

| | 007 | Batman 1 | Shrek 2 | Toy Story 3 | Star Wars 4 | Twilight 5 |
|---------|-----|----------|---------|-------------|-------------|------------|
| Alice | + | — | 0 | + | 0 | — |
| Bob | — | 0 | + | — | + | + |
| Charles | + | + | + | + | — | — |
| Daisy | + | + | 0 | 0 | + | — |
| Everett | + | — | + | + | — | 0 |

Quel score de similarité entre utilisateurs choisir ?

Calcul du score

| | 007 | Batman 1 | Shrek 2 | Toy Story 3 | Star Wars 4 | Twilight 5 |
|---------|-----|----------|---------|-------------|-------------|------------|
| Alice | + | - | 0 | + | 0 | - |
| Charles | + | + | + | + | - | - |
| Score | +1 | -1 | | +1 | | +1 |

$$\text{score}(\text{Alice}, \text{Charles}) = 3 + (-1) = 2$$

| | 007 | Batman 1 | Shrek 2 | Toy Story 3 | Star Wars 4 | Twilight 5 |
|-------|-----|----------|---------|-------------|-------------|------------|
| Alice | + | - | 0 | + | 0 | - |
| Bob | - | 0 | + | - | + | + |
| Score | -1 | | | -1 | | -1 |

$$\text{score}(\text{Alice}, \text{Bob}) = -3$$

Alice est **plus proche** de Charles que de Bob

Score de similarité entre personnes

| | Alice | Bob | Charles | Daisy | JJ |
|---------|-------|-----|---------|-------|----|
| Alice | 4 | -3 | 2 | 1 | 3 |
| Bob | -3 | 5 | -3 | -1 | -2 |
| Charles | 2 | -3 | 6 | 2 | 3 |
| Daisy | 1 | -1 | 2 | 4 | -1 |
| Everett | 3 | -2 | 3 | -1 | 5 |

Qui sont les 2 plus proches voisins d'Alice ?

Calcul des prédictions

| | 007 | Batman 1 | Shrek 2 | Toy Story 3 | Star Wars 4 | Twilight 5 |
|---------|-----|----------|---------|-------------|-------------|------------|
| Alice | + | — | ? | + | ? | — |
| Charles | + | + | + | + | — | — |
| Daisy | + | + | 0 | 0 | + | — |
| Everett | + | — | + | + | — | 0 |

Connaissant ses voisins, quelles sont les chances d'Alice d'apprécier ces films ?

Calcul des prédictions

| | 007 | Batman 1 | Shrek 2 | Toy Story 3 | Star Wars 4 | Twilight 5 |
|---------|-----|----------|---------|-------------|-------------|------------|
| Alice | + | — | + | + | — | — |
| Charles | + | + | + | + | — | — |
| Daisy | + | + | 0 | 0 | + | — |
| Everett | + | — | + | + | — | 0 |

On peut calculer la moyenne :
 $prediction(Alice, Star Wars 4) = 0,333...$

Place au code !

- ▶ `calculer_score(i, j)`
- ▶ `calculer_tous_scores()`
- ▶ `plus_proches_voisins(i)`
- ▶ `calculer_prediction(i, i_{film})`
- ▶ `calculer_toutes_predictions(i)`

Points importants

Écrire du code générique

- ▶ Le même code pour le petit exemple et pour la grosse base de données

Ne calculer que lorsque c'est nécessaire

- ▶ Calculer tous les scores, c'est long, j'ai juste besoin de connaître les voisins d'Alice
- ▶ Recalculer les voisins à chaque fois, c'est idiot

Petites subtilités

- ▶ Division par zéro lorsqu'aucun voisin n'a vu le film

Une petite anecdote

- ▶ Le 2 octobre 2006, Netflix a lancé un concours :
Le premier qui bat notre algorithme de plus de 10 % remportera 1 million de dollars.
et ont filé des données anonymisées
- ▶ La moitié de la communauté en IA s'est jetée sur le problème
- ▶ Le 8 octobre, quelqu'un a battu Cinematch
- ▶ Le 15 octobre, 3 équipes l'avaient battu, dont 1 de 1,06 %
- ▶ Le 26 juin 2009, une équipe n°1 bat Cinematch de 10,05 %
→ **last call** : plus qu'un mois pour gagner
- ▶ Le 25 juillet 2009, une **équipe n°2** bat Cinematch de 10,09 %
- ▶ L'équipe n°1 fait 10,09 % aussi
- ▶ 20 minutes plus tard **l'équipe n°2** fait 10,10 %
- ▶ ... En fait, les deux équipes étaient ex æquo sur le sous-ensemble de validation
- ▶ ... Du coup c'est la première équipe à envoyer ses résultats qui a gagné (équipe 1, 10,09 %)

Confidentialité des utilisateurs

- ▶ Août 2009, Netflix annonce une saison 2
- ▶ Entre-temps, en 2007 deux chercheurs de l'université du Texas ont été capables d'**identifier** les utilisateurs du jeu de données anonymisées en croisant les données avec IMDb
- ▶ (année approximative de naissance, code postal, films vus)
- ▶ En décembre 2009, 4 utilisateurs de Netflix ont attaqué Netflix en justice
- ▶ Mars 2010, arrangement à l'amiable, la plainte est close

Sujet de recherche

Faire de la recommandation qui respecte la **confidentialité** des utilisateurs

Merci de votre attention !

Et bon appétit :-)