


Atelier de Machine Learning Semaine 9

Recommender system


Jeudi 18 avril
2019

Jastry & conoel



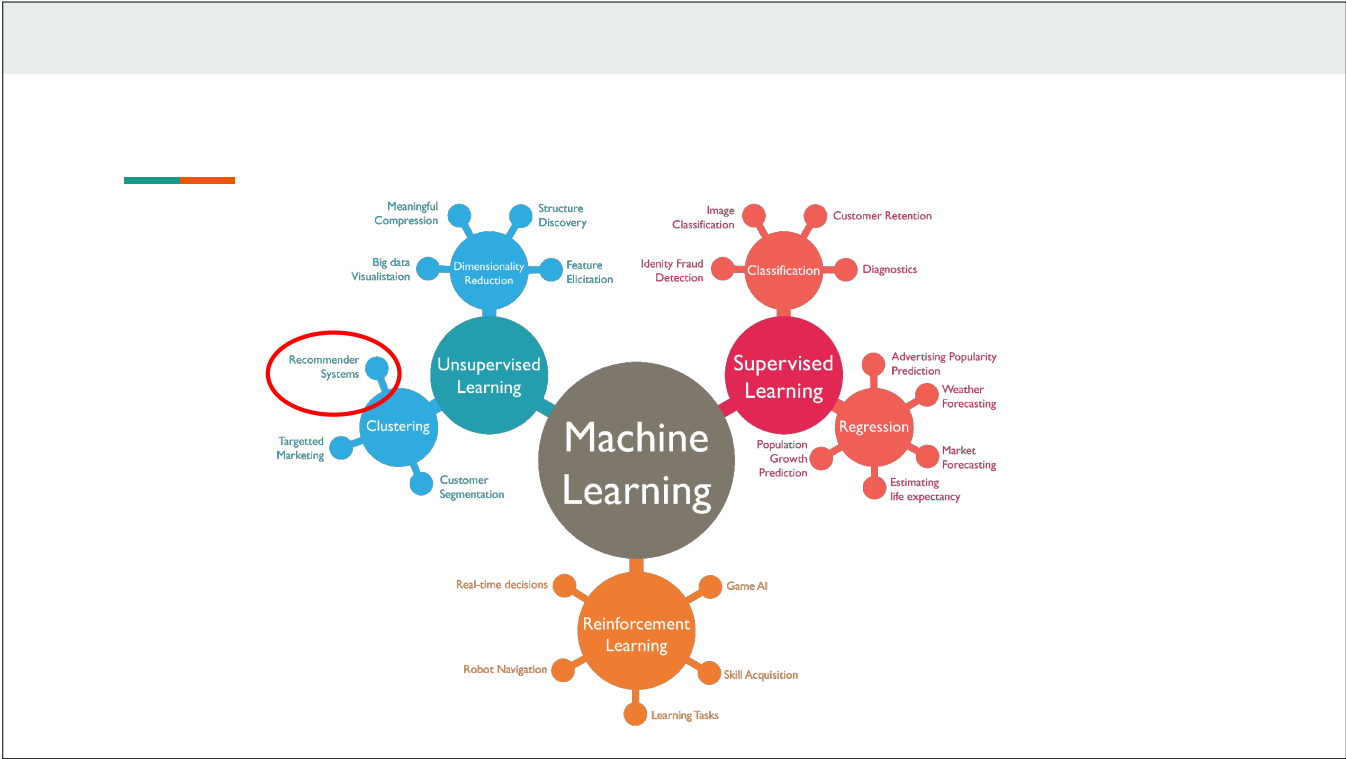
- Dire bonjour
- Ne pas toucher l'écran
-

Pour l'atelier du 2 mai 2019



11 avr	Unsupervised (PCA & k-means)
18 avr	Recommender systems
25 avr	Neural networks (intro)
2 mai	Convolutional neural networks (CNN)
9 mai	Recurrent neural networks (RNN)
16 mai	Reinforcement learning + Q-learning
23 mai	Deep Reinforcement learning (DQN + Policy Gradients)

On as besoin de gens c trop cool tout ça



On est ici

Liste de recommandations



amazon

Netflix Prize



Le prix Netflix, 1 millions de dollars

Frequently Bought Together



Price for all three: **\$74.20**

[Add all three to Cart](#) [Add all three to Wish List](#)

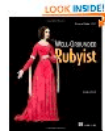
[Show availability and shipping details](#)

- ✓ **This item:** Beginning Ruby: From Novice to Professional (Expert's Voice in Open Source) by Peter Cooper Paperback **\$27.78**
- ✓ [Learn to Program, Second Edition \(The Facets of Ruby Series\)](#) by Chris Pine Paperback **\$16.94**
- ✓ [Ruby on Rails Tutorial: Learn Web Development with Rails \(2nd Edition\) \(Addison-Wesley Professional Ruby ...\)](#) by Michael Hartl Paperback **\$29.48**

Customers Who Bought This Item Also Bought



[Learn to Program, Second Edition \(The Facets of...\)](#)
Chris Pine
★★★★★ 42
Paperback
\$16.94 ✓Prime



[The Well-Grounded Rubyist](#)
› David A. Black
★★★★★ 39
Paperback
\$32.49 ✓Prime



[Ruby on Rails Tutorial: Learn Web Development...](#)
› Michael Hartl
★★★★★ 70
Paperback
\$29.48 ✓Prime



[The Ruby Programming Language](#)
› David Flanagan
★★★★★ 74
Paperback
\$26.35 ✓Prime



[The Well-Grounded Rubyist](#)
› David A. Black
★★★★★ 19
#1 Best Seller in Ruby Programming Computer
Paperback
\$29.67 ✓Prime

Les clients ayant acheté cet article ont également acheté = 30% des ventes

Netflix Prize



grand prize of US\$1,000,000

Training set: 100 480 507 notes attribuées par 480 189 utilisateurs à 17 770 films

Test set: 1 408 342 évaluations

à partir d'un set de données de plus de 100 millions de ratings de films pourrait offrir des recommandations 10% plus précises que celles générées par le système original de la compagnie Netflix. Le 21 septembre 2009, le prix a été attribué à une équipe appelée BellKor's Pragmatic Chaos. Les gagnants avaient misé sur un mélange de centaines d'approches algorithmiques et méthodes prédictives qui réunies ensemble permettent d'être plus performant dans la prédiction.

La solution la plus performante semble donc bien s'appuyer sur un ensemble de méthodes algorithmiques plutôt que sur le raffinement et l'optimisation d'une seule technique spécifique.

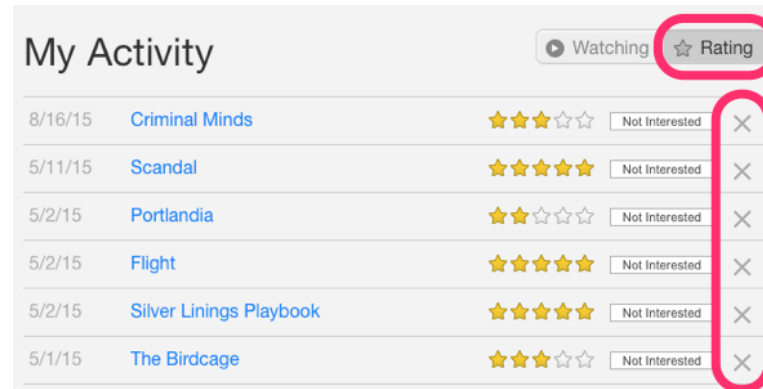
Netflix Prize



Pour l'anecdote, le Netflix prize n'a pas été renouvelé en 2010, car malgré le fait que le set de million de données fournies aux participants du concours aient été rendues anonymes pour préserver l'identité des clients, deux chercheurs de l'Université du Texas ont été capables d'identifier des utilisateurs présents dans le set de données utilisé lors du concours en associant et croisant ces données avec des ratings de films soumis sur l'Internet Movie Database (IMDb). Il s'en est suivi qu'en décembre 2009, un des utilisateurs anonymes du set de données a poursuivi Netflix pour violation de la législation américaine en terme de confidentialité.

Collecte de données explicite VS implicite

Collecte de données explicite – Filtrage actif: repose sur le fait que l'utilisateur indique explicitement au système ses intérêts.



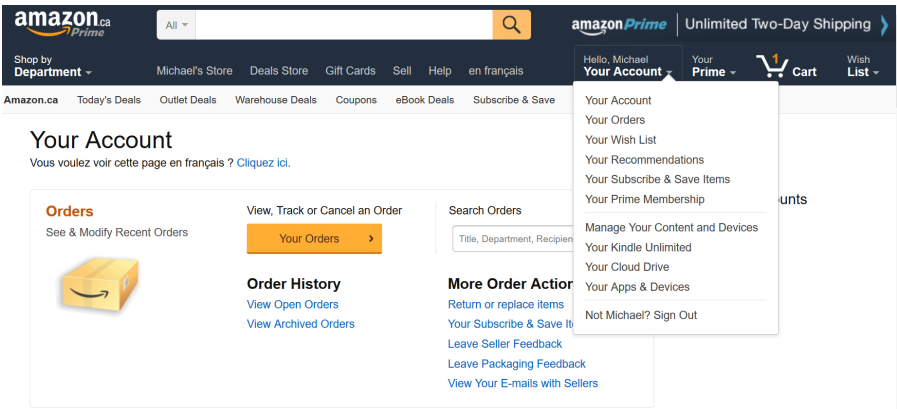
Exemple: demander à un utilisateur de commenter, taguer/étiqueter, noter, liker ou encore ajouter comme favoris des contenus (objets, articles...) qui l'intéressent. On utilise souvent une échelle de ratings allant de 1 étoile (je n'aime pas du tout) à 5 étoiles (j'aime beaucoup) qui sont ensuite transformées en valeurs numériques afin de pouvoir être utilisées par les algorithmes de recommandation.

Avantage: capacité à reconstruire l'historique d'un individu et capacité à éviter d'agréger une information qui ne correspond pas à cet unique utilisateur (plusieurs personnes sur un même poste).

Inconvénient: les informations recueillies peuvent contenir un biais dit de déclaration.

Collecte de données explicite VS implicite

Collecte de données implicite – Filtrage passif: repose sur une observation et une analyse des comportements de l'utilisateur effectuées de façon implicite dans l'application qui embarque le système de recommandation, le tout se fait en “arrière-plan” (en gros sans rien demander à l'utilisateur).



- Exemples:
- Obtenir la liste des éléments que l'utilisateur a écoutés, regardés ou achetés en ligne.
 - Analyser la fréquence de consultation d'un contenu par un utilisateur, le temps passé sur une page.
 - Moniter le comportement en ligne de l'utilisateur.
 - Analyser son réseau social.

Avantage: Aucune information n'est demandée aux utilisateurs, toutes les informations sont collectées automatiquement. Les données récupérées sont a priori justes et ne contiennent pas de biais de déclaration.

Inconvénient: Les données récupérées sont plus difficilement attribuables à un utilisateur et peuvent donc contenir des biais d'attribution (utilisation commune d'un même compte par plusieurs utilisateurs). Un utilisateur peut ne pas aimer certains livres qu'il a acheté, ou il peut l'avoir acheté pour quelqu'un d'autre.

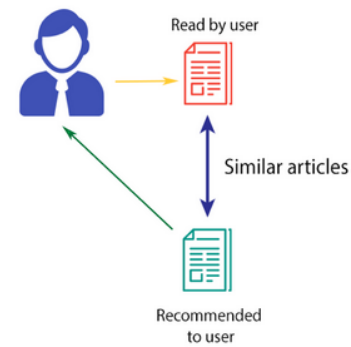


Les Different modèles

Recommandation Personnalisée

Il s'agit de recommander des objets sur la base du comportement passé de l'utilisateur.

CONTENT-BASED FILTERING



Exemples:

Des produits achetés ou sélectionnés sur un site de e-commerce, ainsi qu'un certain nombre d'actions ou décisions effectuées par l'utilisateur qui permettent de prédire de nouveaux produits susceptibles de l'intéresser.

Les annonces publicitaires (par ex. Adsense de Google) sont considérées comme des systèmes de recommandation personnalisée qui se basent sur le comportement passé de l'utilisateur (navigation, clics, historique de recherche...)

Recommandation Personnalisée

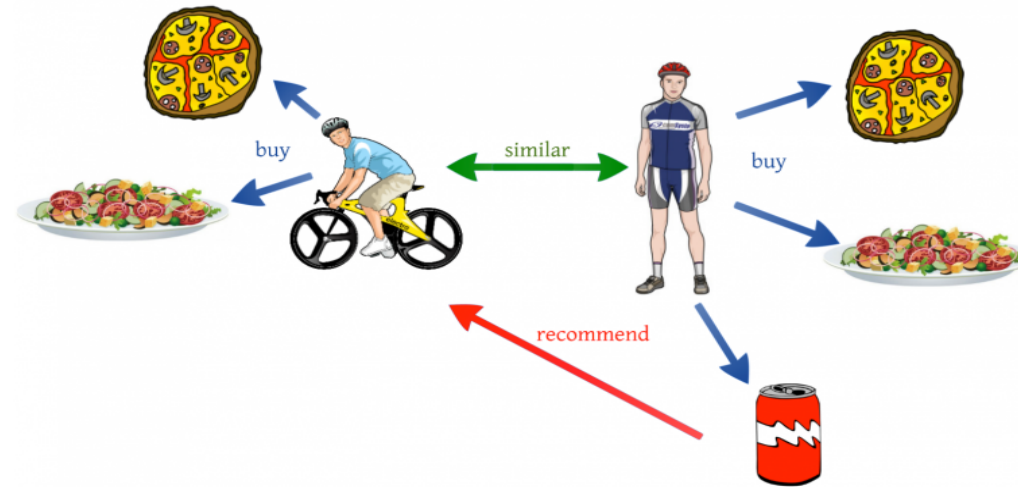
$$\text{sim}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

$$\text{Euclidean Distance} = \sqrt{(x_1 - y_1)^2 + \dots + (x_N - y_N)^2}$$

$$\text{sim}(u, v) = \frac{\sum (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum (r_{ui} - \bar{r}_u)^2} \sqrt{\sum (r_{vi} - \bar{r}_v)^2}}$$

formule qui permettent de calculer la proximité des données

Recommandation Sociale Personne-Personne



Recommander des choses sur la base du comportement passé des utilisateurs similaires, en effectuant une corrélation entre des utilisateurs ayant des préférences et intérêts similaires.

Recommandation Sociale

$$P_{u,i} = \frac{\sum_v (r_{v,i} * s_{u,v})}{\sum_v s_{u,v}}$$

The prediction $P_{u,i}$ is given by:

$P_{u,i}$ is the prediction of an item

$R_{v,i}$ is the rating given by a user v to a movie i

$S_{u,v}$ is the similarity between users

Recommandation Sociale objet-objet



Recommandation Sociale objet-objet

$$P_{u,i} = \frac{\sum_N (s_{i,N} * R_{u,N})}{\sum_N (|s_{i,N}|)}$$

$$sim(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2}$$

Fonctionne comme le user-user mais avec en plus la somme pondérée des notes des « utilisateurs-voisins »

Recommandation Objet

Profil des livres

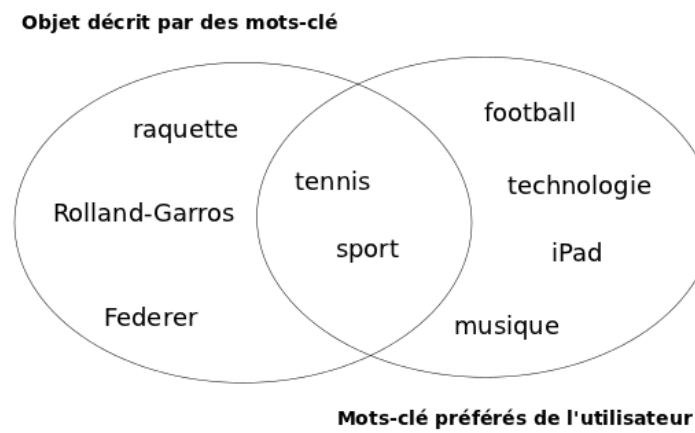
Title	Genre	Author	Type	Price	Keywords
<i>The Night of the Gun</i>	Memoir	David Carr	Paperback	29.90	press and journalism, drug addiction, personal memoirs, New York
<i>The Lace Reader</i>	Fiction, Mystery	Brunonia Barry	Hardcover	49.90	American contemporary fiction, detective, historical
<i>Into the Fire</i>	Romance, Suspense	Suzanne Brockmann	Hardcover	45.90	American fiction, murder, neo-Nazism
...					

Profil/Préférences Utilisateur

Title	Genre	Author	Type	Price	Keywords
...	Fiction, Suspense	Brunonia Barry, Ken Follett	Paperback	25.65	detective, murder, New York

Il s'agit de recommander des objets (ou contenus) en se basant sur les qualités et propriétés intrinsèques de l'objet lui-même et en les corrélant avec les préférences et intérêts de l'utilisateur.

Recommandation Objet



Les algorithmes de recommandation objet permettent de développer des modèles afin de trouver des patterns ou motifs semblables entre différentes données. Ils évaluent à quel point un contenu pas encore vu par l'utilisateur est similaire aux contenus que celui-ci a évalués positivement dans le passé.

Recommandation Hybride



Une combinaison des trois approches ci-dessus. Les méthodes hybrides sont de plus en plus utilisées, car elle permettent de résoudre des problèmes comme le cold start et la sparsity (rareté) qu'on retrouve dans une approche de recommandation uniquement sociale.

Recommandation Hybride

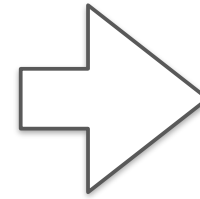
Collaborative filtering

Movie	Rank
A	1
B	0.8
C	0.6
D	0.4
E	0.2



Content Based Filtering:

Movie	Rank
B	1
D	0.8
A	0.6
C	0.4
E	0.2



Movie	New Rank
A	$1+0.6 = 1.6$
B	$0.8+1 = 1.8$
C	$0.6+0.4 = 1$
D	$0.4+0.8 = 1.2$
E	$0.2+0.2 = 0.4$

Efficacité d'un algorithme de recommandation

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

Pour déterminer l'efficacité d'un système, des indicateurs comme la précision et le recall sont utilisés.

The end



APPLAUSE !!

Merci d'applaudir vigoureusement sinon !!!!