FACULTÉ DES SCIENCES DHAR EL MAHRAZ
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH

Projet tutoré

Intitulé:

Collaborative filtering approach in recommender systems

Présenté par:

- El Yagouby Mohamed Amine
- Taghzouti Othmane

Encadré par:

- Pr. Nfaoui El Habib

Soutenu le 28 mai 2022

Le plan

I - Contexte général du projet

Les systèmes de recommandation → Problématique → Solution proposée → Objectifs → Organisation

II - Systèmes de recommandations basés sur le filtrage collaboratif

Data utilisée → Memory-based → Model-based → L'évaluation

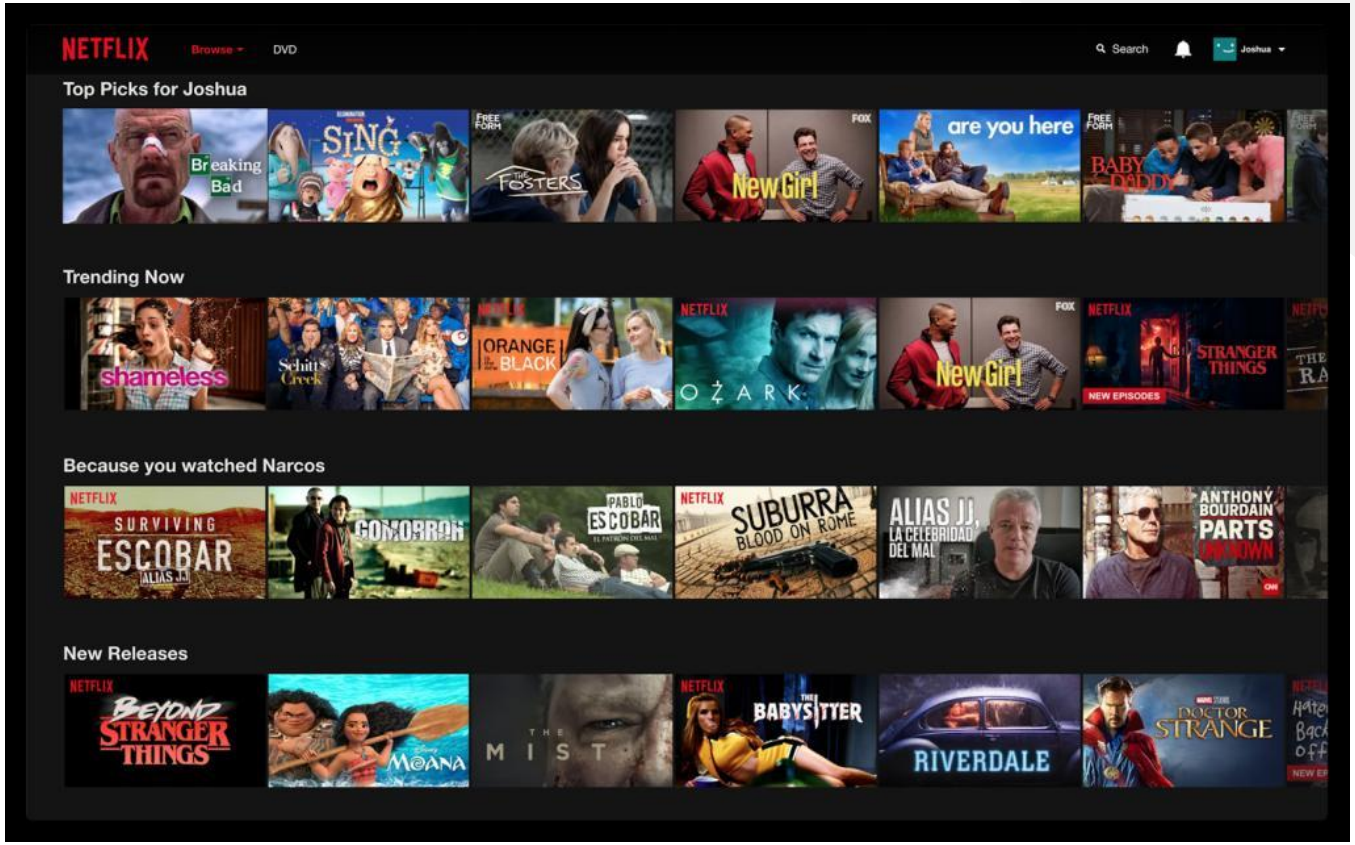
III - L'application web SmartRecommender

Description → Diagramme CU → Outils → Démonstration → Résultats et discussion

Contexte général du projet

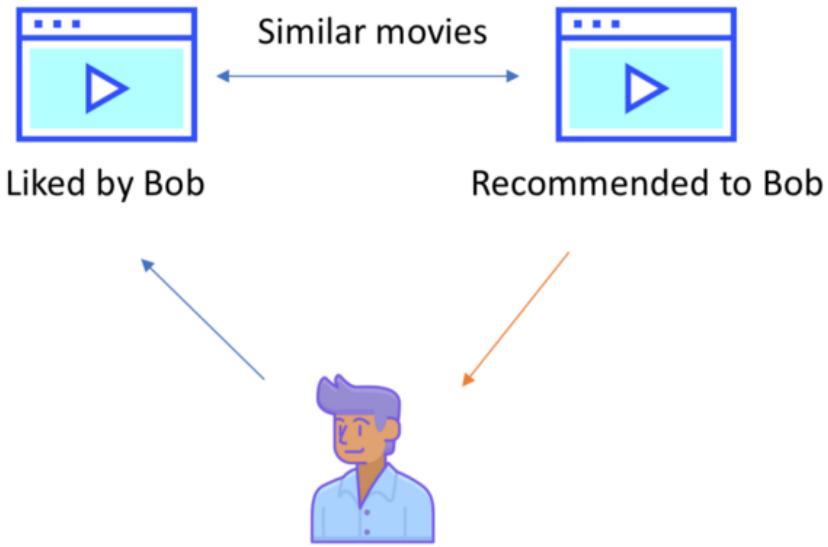
Aperçu sur les systèmes de recommandation

- Guider l'utilisateur dans ses choix
- Maximiser les profits (Amazon 5ème plus grande entreprise au monde, 70% des visionnages sur Youtube…)

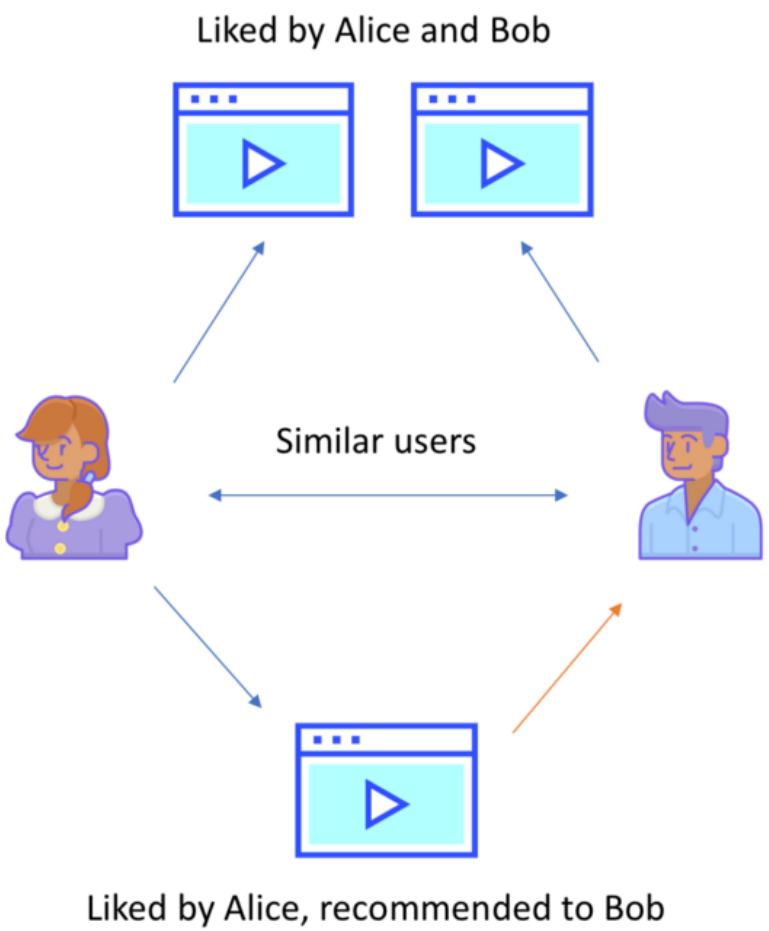


Principe de fonctionnement

Content-based filtering



Collaborative filtering



Problématique

- Comment peut-on construire un système de recommandation?
  - Laquelle des deux méthodes est mieux adaptée?
- Comment peut-on valider le résultat de notre système de recommandation?

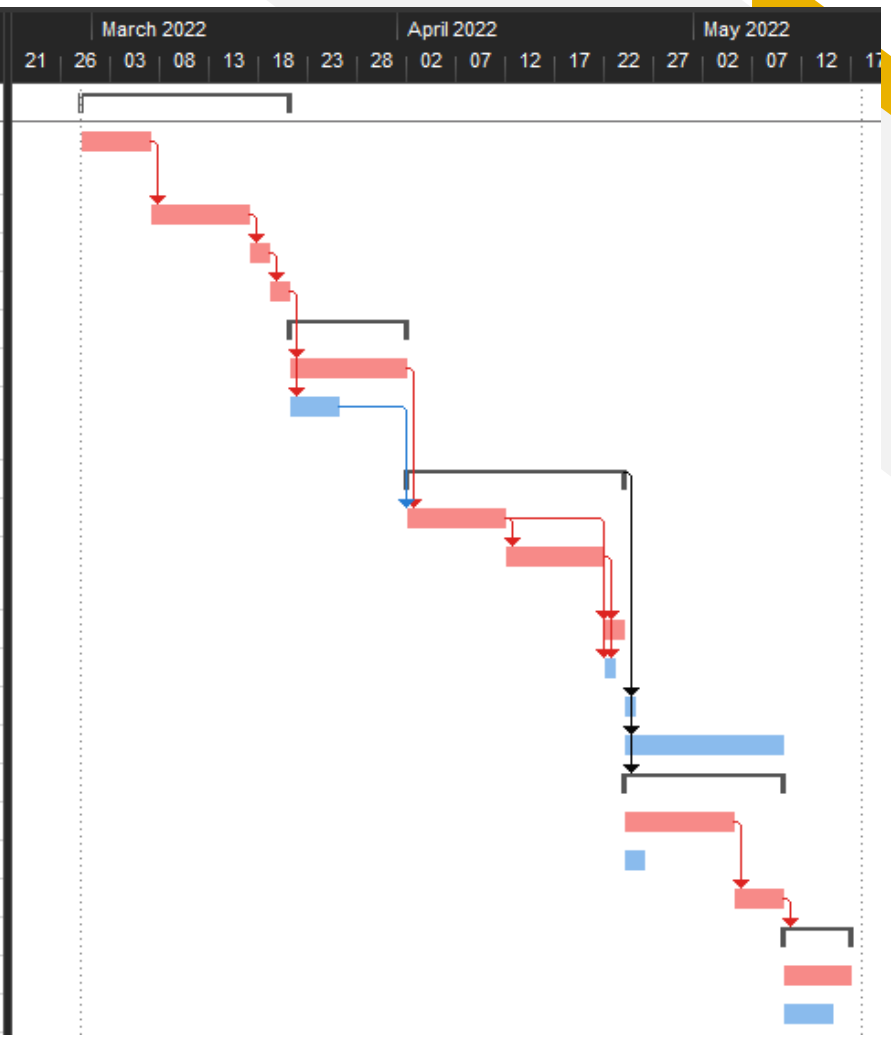
Filtrage collaboratif:

- Pas de connaissance approfondie requise
  - Data à utiliser moins volumineuse
    - L ’effet « surendipitous »

Objectifs du projet

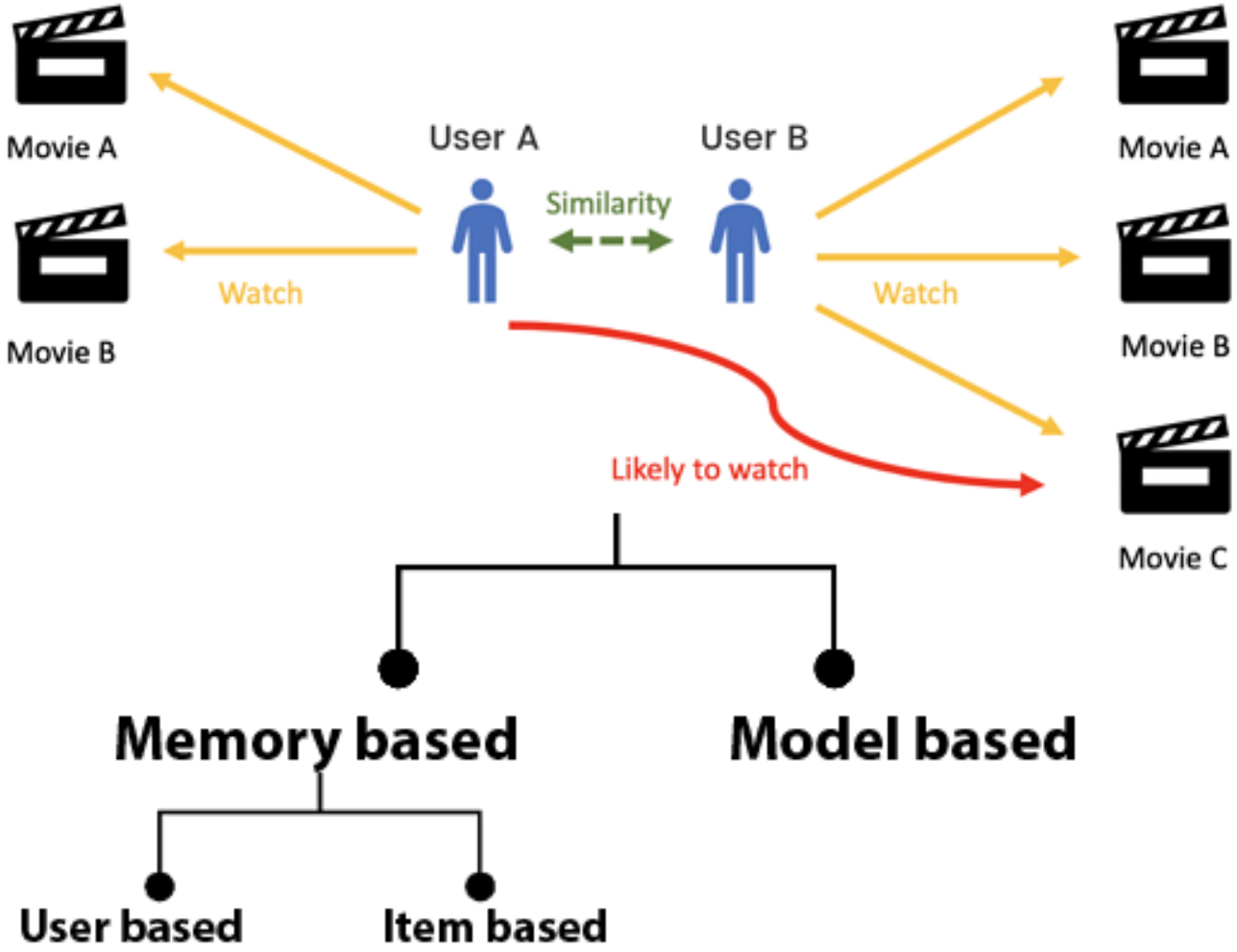
- Étude du filtrage collaboratif avec ses différents modèles
  - Comparaison et analyse de ces modèles
- Réalisation d'une application web de recommandation avec des datasets variées

Task Name	Duratio	Start	Finish	Predec	Resource Names
Building background knowledge	**21 days**	**28/02/22**	**21/03/22**		**Amine El Yagouby;Othmane Taghzouti**
Discovery of artificial intellingence and machine learning	7 days	28/02/22	07/03/22		
Studying of machine learning basic algorithms	10 days	07/03/22	17/03/22	2	
Learning basics of Python	2 days	17/03/22	19/03/22	3	
Familiarization with Google Colab and Jupyter	2 days	19/03/22	21/03/22	4	
Learning about recommender systems	**12 days**	**21/03/22**	**02/04/22**		**Amine El Yagouby;Othmane Taghzouti**
Recommender Systems approaches	12 days	21/03/22	02/04/22	5	
Singular Value Decomposition mathematical aspects	5 days	21/03/22	26/03/22	5	
Elaborating the algorithms	**22 days**	**02/04/22**	**24/04/22**		**Amine El Yagouby;Othmane Taghzouti**
Elaboration of some model based approaches	10 days	02/04/22	12/04/22	8;7	
Elaboration of some memory based approaches	10 days	12/04/22	22/04/22	10	
Testing multiple datasets	2 days	22/04/22	24/04/22	11;10	
Crossvalidating results	1 day	22/04/22	23/04/22	10;11	
Defining the use-case diagram	1 day	24/04/22	25/04/22	9	
Developping the web application	16 days	24/04/22	10/05/22	9	Amine El Yagouby
Preparing the report	**16 days**	**24/04/22**	**10/05/22**	**9**	**Othmane Taghzouti**
Writing the report	11 days	24/04/22	05/05/22		
Designing the images	2 days	24/04/22	26/04/22		
Checking and validating	5 days	05/05/22	10/05/22	17	
Final steps	**7 days**	**10/05/22**	**17/05/22**	**19**	**Amine El Yagouby;Othmane Taghzouti**
Reviewing the application	7 days	10/05/22	17/05/22		
Preparing the presentation	5 days	10/05/22	15/05/22		



Le filtrage collaboratif

Le filtrage collaboratif



Data utilisée

Matrice utilisateur-objet (User-item matrix)

userId	movieId	ratings
196	242	3
186	302	3
22	377	1
244	51	2
166	346	1

Objets

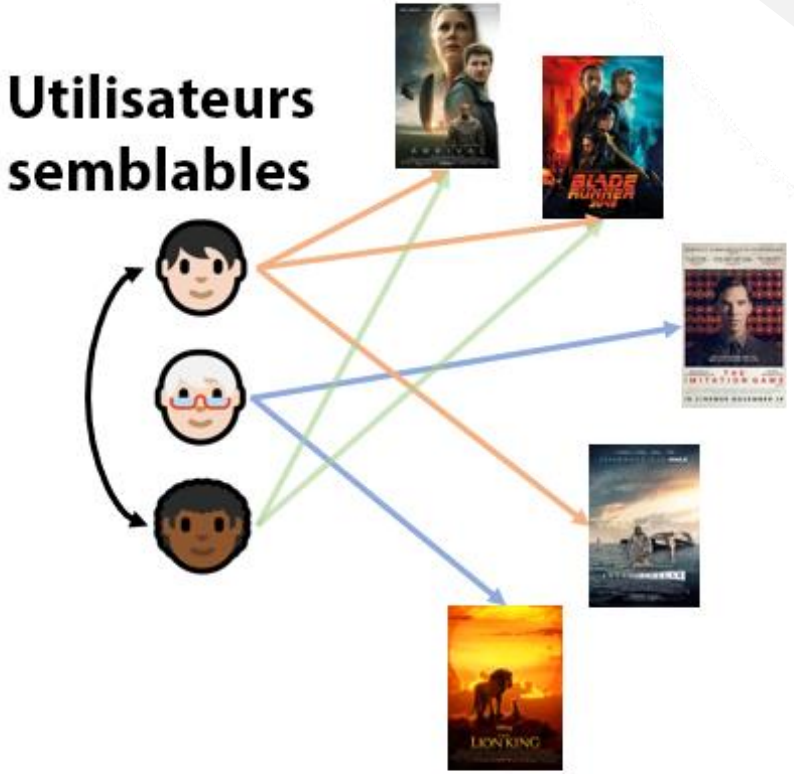
Utilisateurs

4				5	
4	5		5		
	5		4		3
5	4	?	?	?	?

Les types du filtrage collaboratif: memory based (User based)

$$\hat{r}_{u,i} = aggr_{v \in N_i(u)} r_{v,i}$$

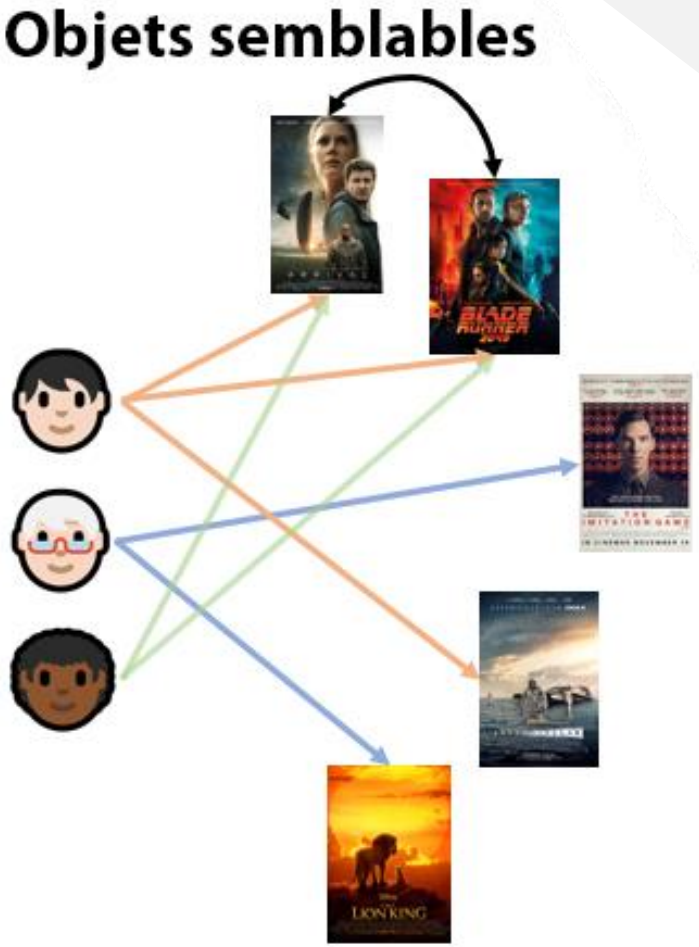
$$\hat{r}_{u,i} = \frac{\sum_{v \in N_i(u)} sim(u,v).r_{v,i}}{\sum_{v \in N_i(u)} sim(u,v)}$$



Les types du filtrage collaboratif: memory based (Item based)

$$\hat{r}_{u,i} = aggr_{j \in N_u(i)} r_{u,j}$$

$$\hat{r}_{u,i} = \frac{\sum_{j \in N_u(i)} sim(i,j).r_{u,j}}{\sum_{j \in N_u(i)} sim(i,j)}$$



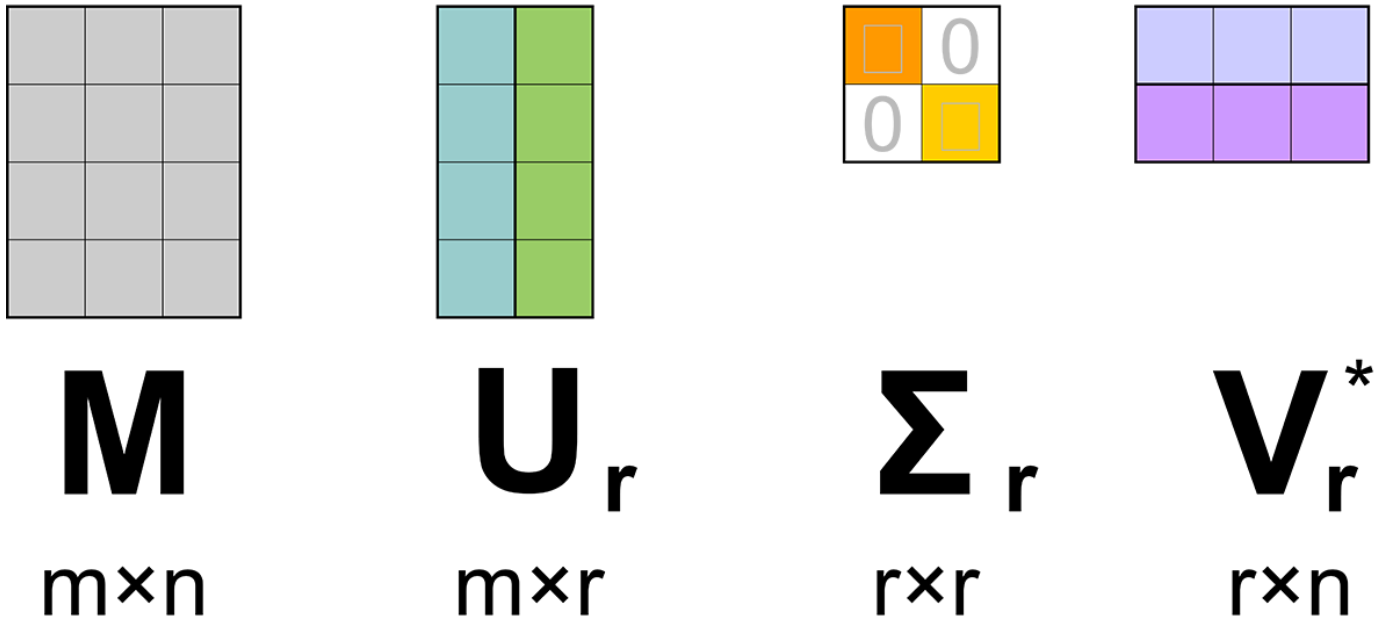
Mesure de similarité

$$\boldsymbol{cos\ similarity(u,v) = \frac{\sum_{i \in I_{uv}} r_{ui} r_{vi}}{\sqrt{\sum_{i \in I_u} r_{ui}^2 \sum_{i \in I_v} r_{vi}^2}}}$$

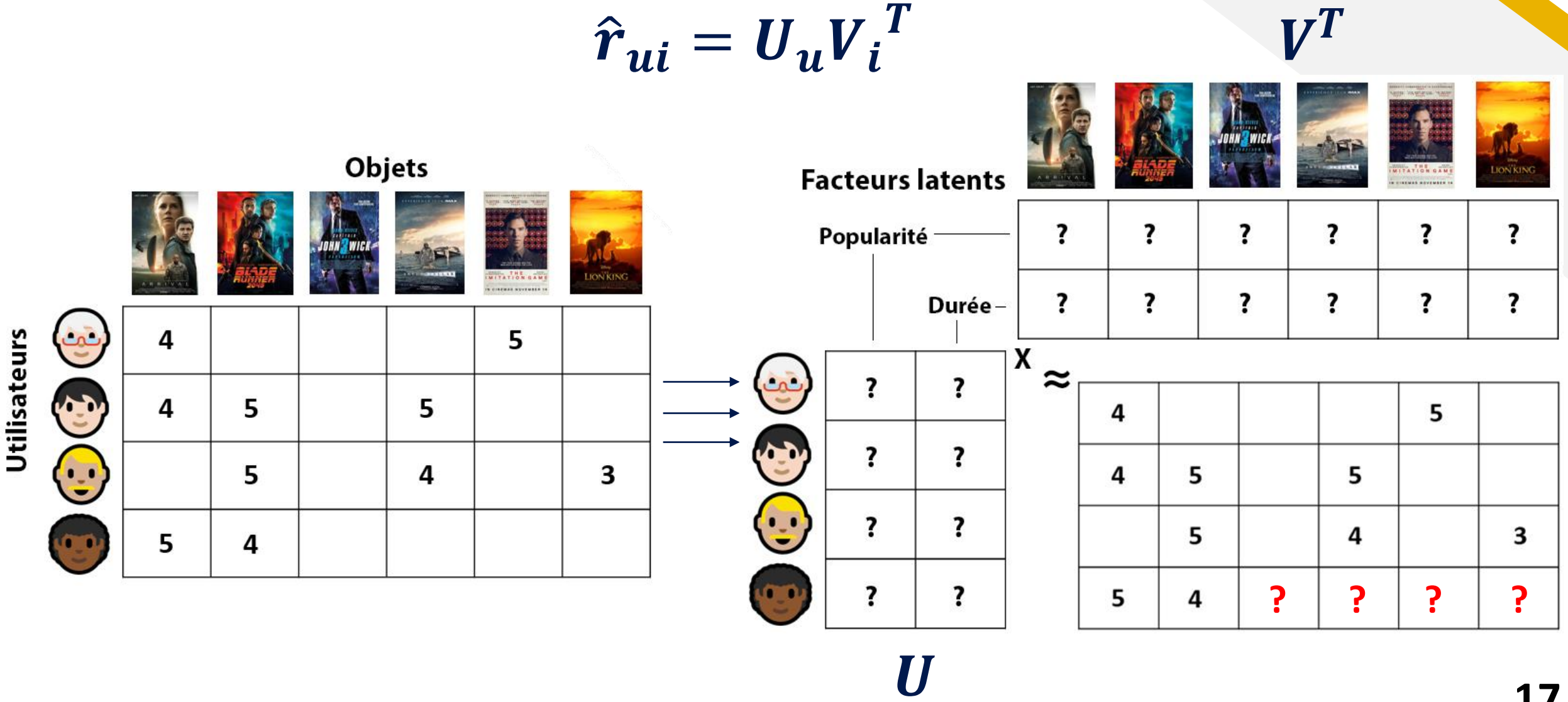
$$\boldsymbol{msd\ similarity(u,v) = \frac{1}{\frac{1}{|I_{uv}|} \sum_{i \in I_{uv}} (r_{ui} - r_{vi})^2 + 1}}$$

Model based

Décomposition en valeurs singulières (SVD)



$$\hat{r}_{ui} = U_u V_i^T$$



Matrice incomplète (95% des valeurs sont inconnues)

Objets

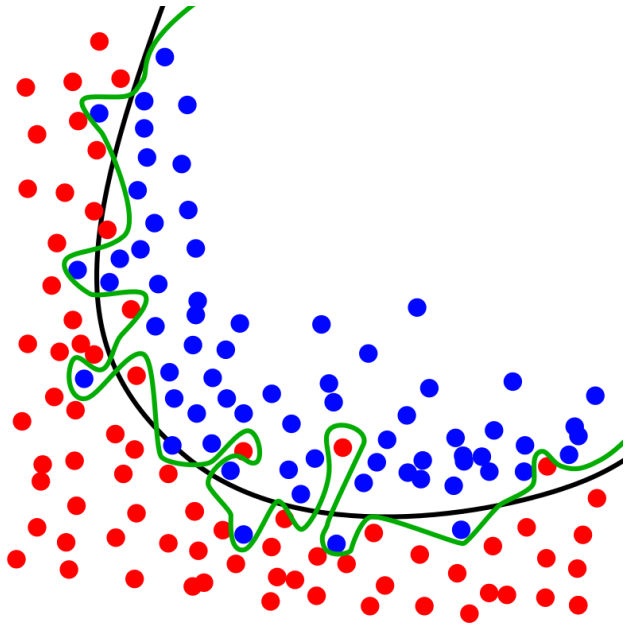
Utilisateurs

4				5	
4	5		5		
	5		4		3
5	4				

Model based

Minimisation de la fonction objective

$$min_{p,q} \sum_{(u,i)\in K} \left(r_{ui} - U_u V_i^T\right)^2 + \lambda\left(\|U_u\|^2 + \|V_i\|^2\right)$$



Validation des modèles de prédiction

- **Mesure de l'écart entre les scores estimés et les scores réels (non utilisés dans la construction des modèles) grâce aux relations suivantes**

L'erreur absolue moyenne (MAE)
Mean Absolute Error

$$MAE = \frac{1}{|\widehat{R}|} \sum_{\hat{r}_{ui} \in \widehat{R}} |r_{ui} - \hat{r}_{ui}|$$

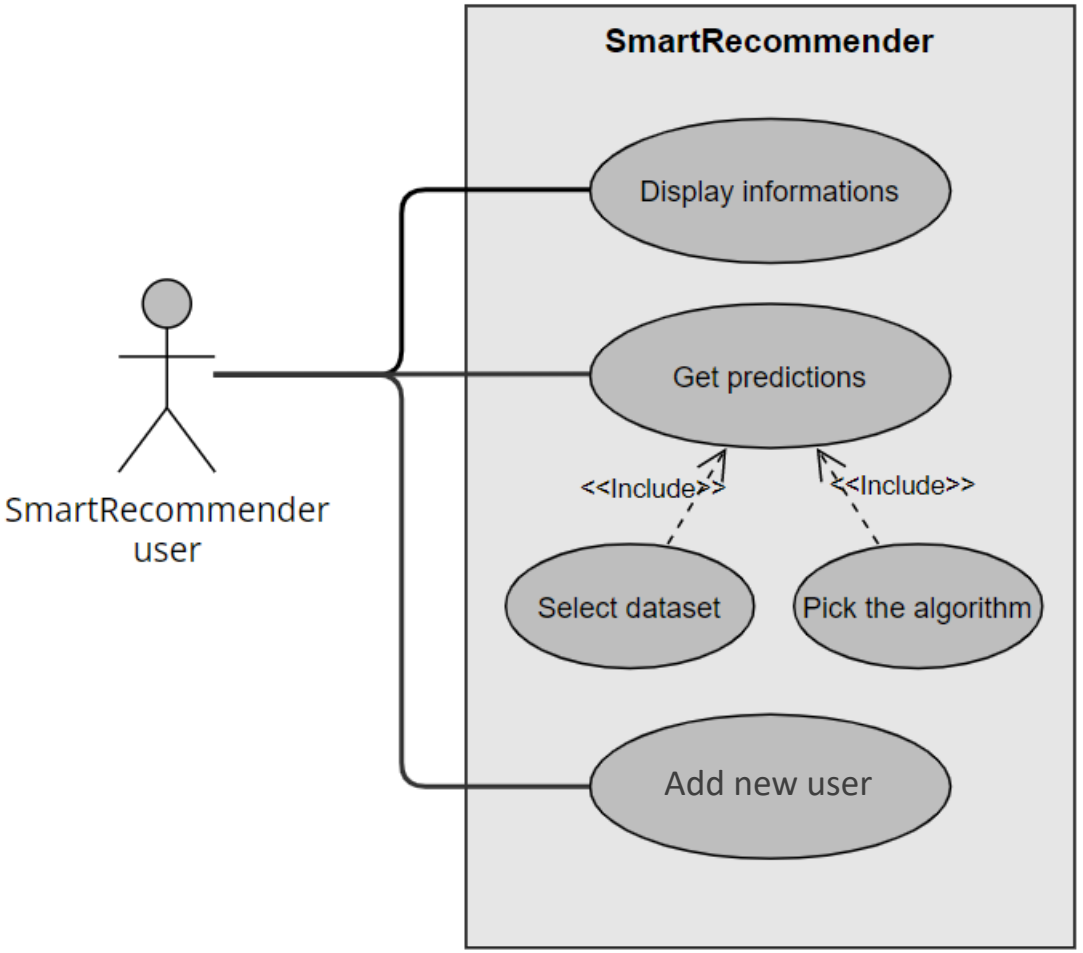
L'erreur quadratique moyenne (RMSE)
Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{|\widehat{R}|} \sum_{\hat{r}_{ui} \in \widehat{R}} (r_{ui} - \hat{r}_{ui})^2}$$

L’application web SmartRecommender

Description de l'application

- Exploiter le travail achevé précédemment au sein de l'étude des différents modèles du filtrage collaboratif
- Construction et test des systèmes de recommandation en utilisant 2 bases de données:
  - Films (MovieLens 1M ratings, 3883 films, 6040 utilisateurs)
  - Livres (GoodBooks 981757 ratings, 10000 livres 53000 utilisateurs)


SmartRecommender
Display informations
Get predictions
<<Include>>
<<Include>>
Select dataset
Pick the algorithm
Add new user
SmartRecommender user

Les outils principalement utilisés

Python
Langage de programmation qui fournit des outils adaptés à l'apprentissage automatique

Pandas
Libraire de manipulation des données

Django
Framework qui suit l'architecture MVT

Scikit-Suprise

- Librairie orientée spécifiquement à la construction et l'analyse des systèmes de recommandation
- Convient parfaitement à notre projet grâce aux algorithmes de CF qui y sont définis: SVD pour le model-based, KNN inspired pour le memory based

Démonstration en vidéo

SmartRecommender

Fichier | C:/Users/intel/PycharmProjects/pandasprojectv3/app-home/index.html

Home Datasets About

SmartRecommender

Select the dataset :

Books dataset

Select the algorithm :

MSD

Select the choice for the KNN algorithm:

Item based

Enter the user Id :

Enter the item Id :

Résultats et discussions

SVD

accuracy of the algorithm :

RMSE	0.8809531994362201
MAE	0.6911881047892232

KNN User-based

accuracy of the algorithm :

RMSE	0.9283307004098478
MAE	0.7321873820573952

KNN Item-based

accuracy of the algorithm :

RMSE	0.9178920596660614
MAE	0.7230052919717014

Résultats et discussions

Exemple:

L'utilisateur 5 qui est intéressé par les films du genre comédie et drama

SVD

Top 5 recommandations for the user 5 :

Annie Hall (1977)	Comedy|Romance
Do the Right Thing (1989)	Comedy|Drama
Manhattan (1979)	Comedy|Drama|Romance
Magnolia (1999)	Drama
Seventh Seal, The (Sjunde inseglet, Det) (1957)	Drama

KNN User-based

Top 5 recommandations for the user 5 :

Gate of Heavenly Peace, The (1995)	Documentary
Schlafes Bruder (Brother of Sleep) (1995)	Drama
Paralyzing Fear: The Story of Polio in America, A (1998)	Documentary
Ulysses (Ulisse) (1954)	Adventure
Smashing Time (1967)	Comedy

KNN Item-based

Top 5 recommandations for the user 5 :

Open Season (1996)	Comedy
Harlem (1993)	Drama
Condition Red (1995)	Action|Drama|Thriller
Angela (1995)	Action|Drama|Thriller
Nueba Yol (1995)	Comedy|Drama

Résultats et discussions

- On ne peut pas décider directement le meilleur système de recommandation
- Chaque base de données s'adapte différemment avec un certain algorithme

  - Les entreprises divisent les données collectées
  - Testent chaque partie avec un modèle diffèrent
  - Prennent finalement le modèle qui génère un maximum de revenu

Conclusion

Fin

Merci pour votre attention