Cours de Data Mining – Engin de Recommendation

Andreea Dragut

Univ. Aix-Marseille, IUT d'Aix-en-Provence

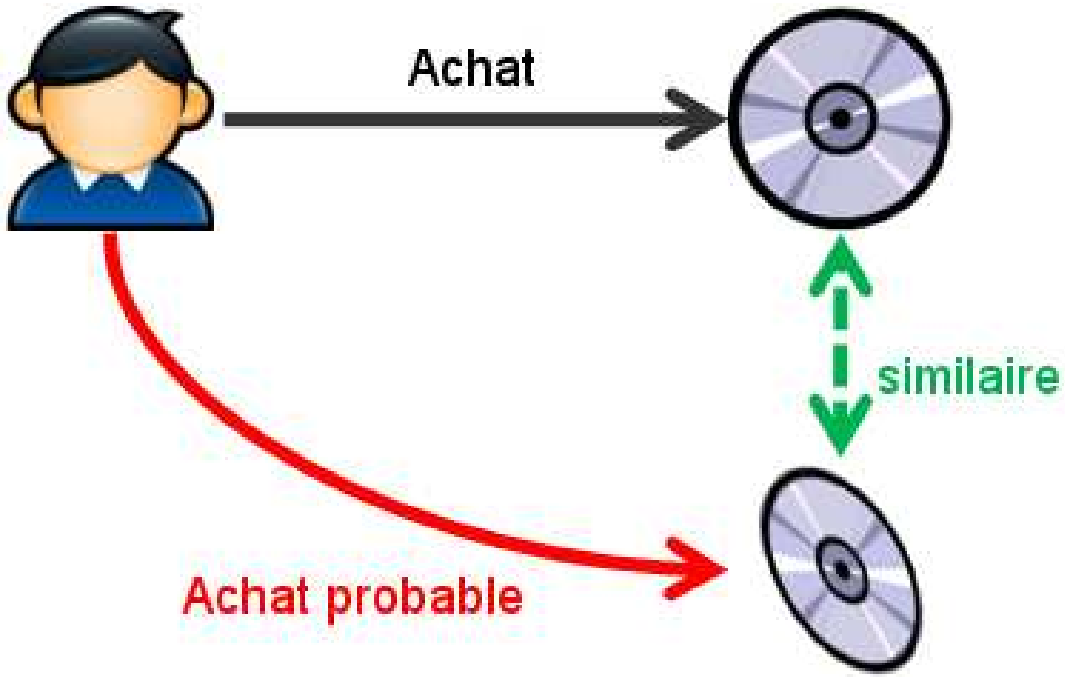
Plan du cours

- Présentation
- Paradigmes
- Techniques

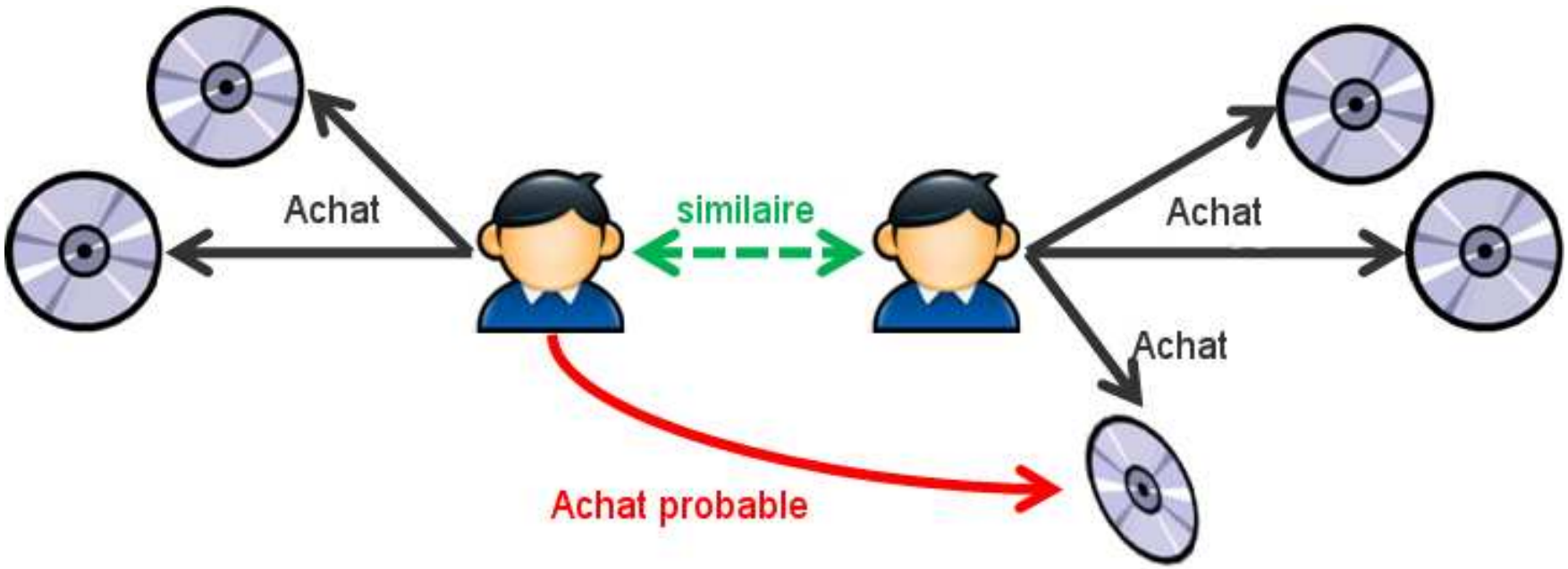
Engin de Recommendation

- Étant donné :
  - un modèle d'utilisateur (estimations, préférences, données démographiques, IP, données contextuelles)
  - des articles (avec ou sans description des caractéristiques des articles)
- Trouver
  - la pertinence d'un article pour un utilisateur. utilisée pour trouver le rang d'un article dans la liste des recommendations possibles.

Paradigmes des engins de recommendation



Paradigmes des engins de recommendation



Collaborative

- Pour
  - Presque aucun effort d'amorçage
  - Variété de résultats
- Contre
  - Besoin d'une forme de notation
  - Le feedback – problématique pour les utilisateurs nouveaux
  - apprentissage de segments de marché et de nouveaux objets

Basée sur le contenu

- Pour
  - Absence de besoin de communauté d'utilisateurs
  - Besoin de descriptions des objets
  - Possibilité de comparaison entre objets
- Contre
  - Démarrage à froid pour nouveaux utilisateurs
  - Résultats très prévisibles

Basée sur les connaissances

- Pour
  - Recommandations déterministes
  - Qualité assurée
  - Pas de démarrage à froid
- Contre
  - Effort de génie de la connaissance (*knowledge engineering*)
  - statique
  - ne réagit pas aux tendances à court terme

Filtrage Collaboratif

- L'approche la plus utilisée pour produire des recommandations
- Applicable dans beaucoup de domaines (livresm films, DVDs)
- Approche :
  - employer la « sagesse du groupe » pour recommander des articles
- Hypothèse
  - Les utilisateurs donnent des estimations (notations) aux articles de catalogue (implicitement ou explicitement) : clicks, page views, temps passé sur certaines pages, téléchargement de versions démo
  - un goût semblable dans le passé $\Longrightarrow$ un goût semblable dans l'avenir

Filtrage Collaboratif plus-proche-voisin, basé sur l'utilisateur

- Problème
  - Soit Alice, un « utilisateur actif », et un objet 5 pas encore vu par Alice. Estimer la notation d'Alice pour cet objet.
- Solution
  - trouver un ensemble d'utilisateurs (équivalents) qui ont apprécié les mêmes objets 1,2,3 et 4 qu'Alice dans le passé et qui ont noté cet objet N
  - utiliser la moyenne de leur notations pour prédire si Alice aimera l'objet N
  - itérer ce processus pour tous les objets pas encore vus par Alice et recommander celui le mieux noté.

Utilisateur	Objet 1	Objet 2	Objet 3	Objet 4	Objet 5
Alice	5	3	4	4	**?**
Utilisateur1	3	1	2	3	3
Utilisateur2	4	3	4	3	5
Utilisateur3	3	3	1	5	4
Utilisateur4	1	5	5	2	1

Filtrage Collaboratif plus-proche-voisin, basé sur l'utilisateur

- Questions
  - Comment mesure-t-on la similarité ?
  - Combien d'utilisateurs voisins devrions-nous considérer ?
  - Comment générer une prédiction à partie des notations des utilisateurs voisins ?

Améliorer les métriques et la fonction de prédiction

- Toutes les notations des utilisateurs voisins n'ont pas nécessairement la même valuer
  - l'accord sur les objets communément aimés est moins informatif que celui sur les objets communément détestés
  - Solution possible : donner plus de poids aux objets qui ont une variance plus grande
- Valeur du nombre d'objets notés par plusieurs utilisateurs
  - Utiliser une « pondération par significance », e.g. en diminuant linéairement le poids lorsque le nombre d'objets notés par plusieurs utilisateurs est petit
- Amplification de cas
  - Intuition : donner plus de poids aux voisins « très similaires », c'est-à-dire lorsque la valeur de la similarité est proche de 1.
- Selection de voisinage
  - Utiliser un seuil pour la similarité ou un nombre fixe de voisins

Approches basées sur la mémoire ou sur un modèle

- Le Filtrage Collaboratif plus-proche-utilisateur-voisin est « basées sur la mémoire »
  - la matrice des notations est utilisée directement pour trouver les voisins et faire les prédictions
  - les grands sites de e-commerce ont des 10^7 clients et des 10^6 de clients
- Approches basées sur un modèle
  - basées sur une étape de prétraitement offline ou d'apprentissage de modèle
  - pendant l'exécution, uniquement le modèle appris est utilisé pour faire les prédictions
  - les modèles sont mis à jour et réappris périodiquement

Filtrage Collaboratif basé sur les objets

- Idée de base
  - Utiliser la similarité entre objets (et non pas entre utilisateurs) pour faire des prédictions
- Exemple
  - Chercher les objets similaires à l'objet 5
  - Utiliser les notations d'Alice pour ces objets afin de prédire la notation d'Alice pour l'objet 5.