

Systèmes de recommandation

Shengrui Wang

11 novembre 2024



Plan

Introduction

Motivations

Comment anticiper sur le choix des utilisateurs?

Filtrage basé sur le contenu

Filtrage collaboratif

Filtrage basé sur la connaissance

Introduction

- Un système de recommandation est un ensemble de programme(s) mis sur pied pour assister l'utilisateur à faire un/des choix parmi un certain nombre de propositions.



Introduction (concepts de base)

- Le système de recommandation a pour rôle de fouiller (**de manière anticipative**) dans une base de données un ensemble d'informations susceptibles d'intéresser l'utilisateur:
On parle de filtrage de l'information.
- Par le biais d'un système de recommandation, nous avons donc une forme de plateforme interactive, où le client peut interagir avec le système.
- **La navigation:** Lorsqu'un client demande un article en particulier, le système de recommandation le lui recommande selon les conditions proposées;
- **Article semblable:** Le système de recommandation propose à la clientèle des articles susceptibles de l'intéresser en se référant à ceux de son panier d'achats;

Introduction (concepts de base)

- **Courriels électroniques:** Le système de recommandation offre à la clientèle des informations relatives aux articles de son champ d'intérêt au moyen des courriels électroniques;
- **Les commentaires:** Le système de recommandation fournit aux clients les commentaires des autres se rapportant sur les articles spécifiques;
- **La moyenne pondérée:** Le système de recommandation affiche aux clients l'évaluation des autres portant sur les articles spécifiques;
- **Top-N:** Le système de recommandation suggère au client suivant sa préférence N articles susceptibles de le séduire;
- **Résultats de recherche ordonnés:** Le système de recommandation énumère tous les résultats de recherche et les classe suivant l'intérêt de sa clientèle.

Motivations

- Les données qui sont mises à la disposition de l'utilisateur sont de plus en plus volumineuses; Besoin de l'aide pour faire des choix.
- Reproduire les routines de l'utilisateur;
- Découverte de nouveaux produits;
- Améliorer l'expérience de l'utilisateur et ainsi booster le niveau des ventes (pour un site marchand);

Motivations

- Convertir les navigateurs en des acheteurs;
- Augmenter la vente dans les sites électroniques;
- Renforcer la fidélité de la clientèle envers les sites électroniques.

Comment procéder?

- Par interaction avec le système, l'utilisateur fournit un ensemble d'informations dont-il tirera profit pour effectuer sa recommandation.
- Les informations prises par le système peuvent être par navigation explicite ou implicite.
- La navigation explicite: désigne le fait que la navigation de la clientèle vise à communiquer ses préférences aux systèmes de recommandations.

Comment procéder?

- La **navigation explicite**: désigne le fait que la navigation de la clientèle vise à communiquer ses préférences aux systèmes de recommandations.
 - ▶ Demander à un utilisateur d'évaluer un article dans une échelle mobile;
 - ▶ Demander à un utilisateur de classer les articles selon sa préférence;
 - ▶ Présenter deux articles à un utilisateur et lui demander prédilection;
 - ▶ Demander à un utilisateur de faire une liste avec ses articles favoris.

Comment procéder?

- La **navigation implicite**: désigne le fait de considérer la navigation de la clientèle comme la saisie du système de recommandation, et ce, sans la connaissance de la personne concernée.
 - ▶ Observer les articles qu'un utilisateur consulte dans un magasin en ligne;
 - ▶ Prendre en note les articles qu'un utilisateur a acheté sur Internet;
 - ▶ Obtenir une liste des articles qu'un utilisateur a écouté ou regardé sur son poste de travail.

Sur la base des informations explicites ou implicites recueillies, le système de recommandation peut donc aller procéder à la fouille de données et ainsi filtrer celles qui pourraient être importantes pour l'utilisateur.

Comment procéder?

- Ils existent plusieurs techniques de filtrage d'information.
- Parmi les plus connus nous avons:
 1. Le filtrage basé sur le contenu
 2. Le filtrage collaboratif
 3. Le filtrage basé sur la connaissance
- Formellement parlant, étant donné un ensemble d'utilisateurs $U = \{U_1, U_2, \dots, U_n\}$, un ensemble d'items $I = \{I_1, I_2, \dots, I_m\}$, un ensemble de réactions sur les items correspondants $R = \{r_{ui}\}$, ici $1 \leq u \leq n, 1 \leq i \leq m$.

À noter qu'il ne faut pas interpréter comme quoi r_{ui} est définie pour toutes les valeurs admissibles de u et i .

Comment procéder?

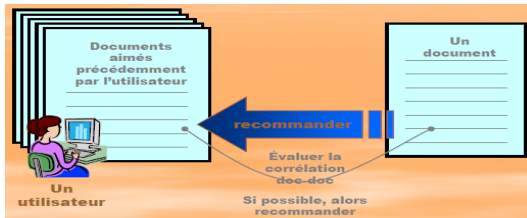
- Sur la base des réactions que l'utilisateur U_u aurait eu sur les items $I_u = \{I_i | r_{u,i} \in \mathbf{R}\}$, on voudrait trouver un nouvel item I_{i^*} qui pourrait susciter son intérêt.

$$I_{i^*} = \underset{I_j \in I \setminus I_u}{\operatorname{argmax}} f(U_u, I_j)$$

- où f est une fonction d'estimation de la préférence des utilisateurs.

Filtrage basé sur le contenu

- Les recommandations basées sur le contenu recueillent d'abord, au moyen du système, les données de l'utilisateur.
- Par la suite, elles analysent et créent son profil de préférence pour enfin suggérer les articles qui lui conviennent.



- Les contenus actuels de l'item aux contenus des items aimés auparavant par l'utilisateur sont comparés et ensuite les items les plus similaires sont recommandés.

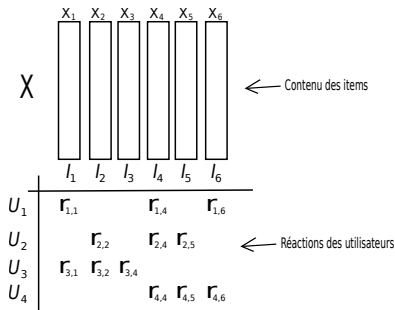
Filtrage basé sur le contenu

Étant le jeu de données suivant :

X	X_1	X_2	X_3	X_4	X_5	X_6	← Contenu des items
	I_1	I_2	I_3	I_4	I_5	I_6	
	<hr/>						
	U_1	$r_{1,1}$		$r_{1,4}$		$r_{1,6}$	
	U_2		$r_{2,2}$		$r_{2,4}$	$r_{2,5}$	
	U_3	$r_{3,1}$	$r_{3,2}$	$r_{3,4}$			
U_4				$r_{4,4}$	$r_{4,5}$	$r_{4,6}$	← Réactions des utilisateurs

De manière générale, un système de recommandation aura pour but de prédire les espaces blancs dans la matrice d'utilisateurs-items.

Filtrage basé sur le contenu



Pour ce faire, le filtrage basé sur le contenu tiendra particulièrement compte des propriétés des items X .

- Les informations du contenu sont données sous forme matricielle $\mathbf{X} \in \mathbb{R}^{m \times d}$, $\mathbf{X} = (x_{i,\tau})_{1 \leq i \leq m, 1 \leq \tau \leq d}$
- À partir desquelles on peut créer, pour chaque utilisateur U_u des profiles P_u comme suit:

$$P_u = \sum_{I_i \in I_u} r_{u,i} \cdot x_i$$

Filtrage basé sur le contenu

- Chaque usager est supposé traité de façon indépendante.
- Les items sont représentés par des caractéristiques (features)
 - ▶ Ex: chaussure: (couleur, catégorie, taille, marque, disponibilité, date, etc. ...)
- Le profil usager est créé et mis à jours automatiquement en réponse aux feedbacks de l'utilisateur sur la désirabilité des articles qui lui ont été présentés.
- Vu que le profil utilisateur P_u est une combinaison des items sur lesquels il a réagi, la recommandation du prochain item pourrait être un item I_i dont le vecteur x_i serait similaire.

- Parmi les différentes techniques de recherche basée sur le contenu nous avons entre autres:

- ▶ Similarité cosinus:

$$\text{recommend}(I_i|P_u) = \begin{cases} 1 & \text{if } 1 - \cos(x_i, P_u) \leq \delta \\ 0 & \text{else} \end{cases}$$

- ▶ Approche Bayesienne:

$$Pr(r_{u,i}|P_u) = Pr(r_{u,i}) \prod_{j \in x_i} Pr(x_{i,j}|r_{u,i})$$

■ Avantages:

- ▶ Indépendance de l'usager: Le système exploite les réactions fournis par l'utilisateur en question pour construire son profil.
- ▶ Pas besoin de données sur les autres usagers.

■ Transparence: On peut facilement expliquer la décision prise par le système de recommandation en consultant les features correspondant à l'item recommandé.

■ Indépendance d'évaluation:

- ▶ Le système de recommandation est capable de recommander des nouveaux items sans que ces derniers n'aient déjà été évalués.

Comme limites nous avons:

- En cas de manque d'information sur les différents items, il devient difficile pour le système de faire une recommandation appropriée.
- Le contenu doit être encodé comme des caractéristiques (features) significatives
 - ▶ L'attribution des caractéristiques aux items peut ne pas être suffisante pour définir des aspects distinctifs des items.
 - ▶ Les mots clés ne sont pas appropriés pour représenter le contenu des items (synonymes, polysèmes, etc.).

Filtrage collaboratif

- Dans le filtrage collaboratif, on tient compte des réactions des utilisateurs sur les différents items,
- En d'autres termes, on recommanderait un item I_i à un utilisateur U_u s'il existe un autre utilisateur U_v ayant déjà réagi sur cet item, et qui aurait eu des réactions antérieures similaires à U_u ,
- Même en cas de manque d'information sur le contenu de l'item, on pourrait toujours recommander cet item à l'utilisateur
- Vu que le système se base sur les réactions des autres, un item non-relatif à l'attente de l'utilisateur pourrait-être proposé.

Filtrage collaboratif

Le filtrage collaboratif se déroule en trois étapes :

1. Collecte d'information sur les utilisateurs: Il est question de collecter l'ensemble des reactions chacun des utilisateurs U_u aurait eu à l'endroit des différents items.

$$\forall U_u \in \mathbf{U}, I_u = \{I_i / r_{u,i} \in \mathbf{R}\}$$


2. Construction d'une matrice d'association;

	GLADIATEUR	ROCKY IV	BEN-HUR	SKYFALL	IDIOCRACY
1	★★★★★			★★★★★	★★★★★
2		★★★★★		★★★★★	
3	★★★★★	★★★★★		★★★★★	
4			★★★★★	★★★★★	
5	★★★★★		★★★★★		

3. Recommandation suivant les similarités dans les associations.


Filtrage collaboratif

- La recommandation peut se faire :
 - ▶ suivant les similarités entre utilisateurs : filtrage collaboratif utilisateur;



	GLADIATEUR	ROCKY IV	BEN-HUR	SKYFALL	IDIOCRACY
①	★☆☆☆☆			★★★★★	★★★★★
②		★★★★★		★★★★★	
③	★★★★★	★★★★★		★★★★★	
④			★★★★★	★★★★★	
⑤	★★★★★		★★★★★		

- ▶ suivant les similarité entre items: filtrage collaboratif item.



	GLADIATEUR	ROCKY IV	BEN-HUR	SKYFALL	IDIOCRACY
①	★☆☆☆☆			★★★★★	★★★★★
②		★★★★★		★★★★★	
③	★★★★★	★★★★★		★★★★★	
④			★★★★★	★★★★★	
⑤	★★★★★		★★★★★		

Filtrage collaboratif



	GLADIATEUR	ROCKY IV	BEN-HUR	SKYFALL	IDIOCRACY
1	★☆☆☆☆			★★★★★	★★★☆☆
2		★★★★★		★★★★★	
3	★★★★★	★★★★★		★☆☆☆☆	
4			★★★★★	★★★★★	
5	★★★★★		★★★★★		

- Filtrage collaboratif utilisateur : trouver des utilisateurs qui ont des préférences similaires à l'utilisateur cible.
- Formellement,

$$Recommend(I_i | U_u) = \begin{cases} 1 & \text{if } \{U_v, U_v \neq U_u / I_u \cap I_v \neq \emptyset, I_i \in I_v\} \neq \emptyset \\ & \text{with } I_u = \{I_j, j \neq i / r_{u,j} \in \mathbf{R}, r_{u,j} \geq \delta\} \\ & I_v = \{I_k / r_{v,k} \in \mathbf{R}, r_{v,k} \geq \delta\} \\ 0 & \text{else} \end{cases}$$

Prédiction des réactions $r_{u,i}$ des utilisateurs en se basant sur la collaboration utilisateur:

- Trouver le vote qu'aurait l'utilisateur U_u sur l'item I_i en se référant sur les utilisateurs les plus semblables à U_u .
- Supposons $\mathbf{U}_{v \sim u} = \{U_v | U_v \neq U_u / I_u \cap I_v \neq \emptyset\}$.
- Un vote auquel on s'attendrait que U_u ait sur l'item I_i , pourrait être

$$r_{u,i} = \left(\frac{1}{|\mathbf{U}_{v \sim u}|} \sum_{U_v \in \mathbf{U}_{v \sim u}} r_{v,i} \right)$$

Le véritable problème avec le filtrage collaboratif par utilisateur est que,

- Les préférences des utilisateurs changent avec le temps; ce qui pourrait générer de nombreuses suggestions inappropriées.
- Plus le nombre d'utilisateurs augmente, plus il sera long de générer des recommandations.
- Démarrage à froid (cold start)
- Vulnérable au *Shilling Attack* : tentative de promouvoir ou rétrograder certains articles par un profil d'utilisateur malveillant inséré dans un ensemble de données de filtrage collaboratif.

Filtrage collaboratif



	GLADIATEUR	ROCKY IV	BEN-HUR	SKYFALL	IDIOCRACY
1	★★★★★			★★★★★	★★★★★
2		★★★★★		★★★★★	
3	★★★★★	★★★★★		★★★★★	
4			★★★★★	★★★★★	
5	★★★★★		★★★★★		

- Calculer la similarité entre items suivant les réactions des utilisateurs.
- Si un utilisateur U_u apprécie un item I_i semblable à l'item I_j , alors on pourrait lui recommander l'item I_j .
- Formellement

$$Recommend(I_j|U_u) = \begin{cases} 1 & \text{if } I_j \in \left\{ I_{i^*} / \right. \\ & \left. I_{i^*} = \underset{I_k \in I \setminus I_u}{argmax} \sim(I_i, I_k) \right\} \\ 0 & \text{else} \end{cases}$$

Filtrage collaboratif

Prédiction des réactions $r_{u,i}$ des utilisateurs en se basant sur la collaboration item:

- Trouver le vote qu'aurait l'utilisateur U_u sur l'item I_j en se référant sur les items semblables à l'item cible dont-il a déjà consulté/voté.
- Supposons $I_{j \sim i} = \left\{ I_{i^*} / I_{i^*} = \underset{I_k \in I \setminus I_u}{argmax} \ sim(I_i, I_k) \right\}$ l'ensemble des items similaires à I_i
- Un vote qu'aurait l'utilisateur U_u sur l'item I_j serait,

$$r_{u,j} = \lceil \left(\frac{1}{|I_{j \sim i}|} \sum_{I_i \in I_{j \sim i}} r_{u,i} \right)$$

l'avantage du filtrage collaboratif par item,

- Les items du système ne changent pas, donc les suggestions seront plus pertinentes.
- Du fait qu'on ait souvent moins d'éléments que d'utilisateurs, le temps de traitement est largement réduit.
- Plus difficile de tricher avec ce système.

filtrage basé sur la connaissance

- Utiliser la connaissance qu'on a de l'utilisateur et des items pour faire correspondre les deux
- Il ne tient pas compte des évaluations des utilisateurs,
- Tient compte des similitudes entre les exigences des utilisateurs et des descriptions d'items.
- S'applique bien dans le cas des items qui sont rarement achetés (ex. maison, voiture, objets de luxe etc. ...)

Inconvénients:

- Nécessite la compétence d'experts dans le domaine
- Savoir quelles caractéristiques sont importantes
- Les caractéristiques doivent être accessibles
- En cas de manque de contenu on a le même problème comme dans le filtrage basé sur le contenu