

	Plan	Introduction
<div>Systèmes de recommandation</div> <div>Shengrui Wang</div> <div>11 novembre 2024</div> <div><div>UDS</div>Université de Sherbrooke</div>	<div>Introduction</div> <div>Motivations</div> <div>Comment anticiper sur le choix des utilisateurs?</div> <div>Filtrage basé sur le contenu</div> <div>Filtrage collaboratif</div> <div>Filtrage basé sur la connaissance</div>	<div><div>■ 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.</div><div><div><div><div><div></div></div><div>Un utilisateur</div></div><div><div><div></div></div><div>Un recommandateur</div></div><div><div><div></div></div><div>Sources d'information</div></div></div></div></div>
<div><div>UDS</div>Université de Sherbrooke</div> 11 novembre 2024Shengrui Wang2	<div><div>UDS</div>Université de Sherbrooke</div> 11 novembre 2024Shengrui Wang3	

Introduction (concepts de base)	Introduction (concepts de base)	Motivations
<div><div>■ Le système de recommandation a pour rôle de fouiller (<b>de manière anticipative</b>) dans une base de données un ensemble d'informations susceptibles d'intéresser l'utilisateur: <b>On parle de filtrage de l'information.</b></div><div>■ Par le biais d'un système de recommandation, nous avons donc une forme de plateforme interactive, où le client peut interagir avec le système.</div><div>■ <b>La navigation:</b> Lorsqu'un client demande un article en particulier, le système de recommandation le lui recommande selon les conditions proposées;</div><div>■ <b>Article semblable:</b> Le système de recommandation propose à la clientèle des articles susceptibles de l'intéresser en se référant à ceux de son panier d'achats;</div></div>	<div><div>■ <b>Courriels électroniques:</b> Le système de recommandation offre à la clientèle des informations relatives aux articles de son champ d'intérêt au moyen des courriels électroniques;</div><div>■ <b>Les commentaires:</b> Le système de recommandation fournit aux clients les commentaires des autres se rapportant sur les articles spécifiques;</div><div>■ <b>La moyenne pondérée:</b> Le système de recommandation affiche aux clients l'évaluation des autres portant sur les articles spécifiques;</div><div>■ <b>Top-N:</b> Le système de recommandation suggère au client suivant sa préférence N articles susceptibles de le séduire;</div><div>■ <b>Résultats de recherche ordonnés:</b> Le système de recommandation énumère tous les résultats de recherche et les classe suivant l'intérêt de sa clientèle.</div></div>	<div><div>■ 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.</div><div>■ Reproduire les routines de l'utilisateur;</div><div>■ Découverte de nouveaux produits;</div><div>■ Améliorer l'expérience de l'utilisateur et ainsi booster le niveau des ventes (pour un site marchand);</div></div>
<div><div>UdS</div>Université de Sherbrooke</div> 11 novembre 2024Shengrui Wang4	<div><div>UdS</div>Université de Sherbrooke</div> 11 novembre 2024Shengrui Wang5	<div><div>UdS</div>Université de Sherbrooke</div> 11 novembre 2024Shengrui Wang6

Motivations	Comment procéder?	Comment procéder?
<ul style="list-style-type: none"> <li>■ Convertir les navigateurs en des acheteurs;</li> <li>■ Augmenter la vente dans les sites électroniques;</li> <li>■ Renforcer la fidélité de la clientèle envers les sites électroniques.</li> </ul>	<ul style="list-style-type: none"> <li>■ Par interaction avec le système, l'utilisateur fournit un ensemble d'informations dont-il tirera profit pour effectuer sa recommandation.</li> <li>■ Les informations prises par le système peuvent être par navigation explicite ou implicite.</li> <li>■ La navigation explicite: désigne le fait que la navigation de la clientèle vise à communiquer ses préférences aux systèmes de recommandations.</li> </ul>	<ul style="list-style-type: none"> <li>■ La <b>navigation explicite</b>: désigne le fait que la navigation de la clientèle vise à communiquer ses préférences aux systèmes de recommandations. <ul style="list-style-type: none"> <li>▶ Demander à un utilisateur d'évaluer un article dans une échelle mobile;</li> <li>▶ Demander à un utilisateur de classer les articles selon sa préférence;</li> <li>▶ Présenter deux articles à un utilisateur et lui demander prédilection;</li> <li>▶ Demander à un utilisateur de faire une liste avec ses articles favoris.</li> </ul> </li> </ul>
<div> <div>UdS</div> <div> <div>Université de Sherbrooke</div> <div>11 novembre 2024</div> <div>Shengrui Wang</div> <div>7</div> </div> </div>	<div> <div>UdS</div> <div> <div>Université de Sherbrooke</div> <div>11 novembre 2024</div> <div>Shengrui Wang</div> <div>8</div> </div> </div>	<div> <div>UdS</div> <div> <div>Université de Sherbrooke</div> <div>11 novembre 2024</div> <div>Shengrui Wang</div> <div>9</div> </div> </div>
Comment procéder?	Comment procéder?	Comment procéder?
<ul style="list-style-type: none"> <li>■ La <b>navigation implicite</b>: 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. <ul style="list-style-type: none"> <li>▶ Observer les articles qu'un utilisateur consulte dans un magasin en ligne;</li> <li>▶ Prendre en note les articles qu'un utilisateur a acheté sur Internet;</li> <li>▶ Obtenir une liste des articles qu'un utilisateur a écouté ou regardé sur son poste de travail.</li> </ul> </li> </ul> <p>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.</p>	<ul style="list-style-type: none"> <li>■ Ils existent plusieurs techniques de filtrage d'information.</li> <li>■ Parmi les plus connus nous avons: <ol style="list-style-type: none"> <li>1. Le filtrage basé sur le contenu</li> <li>2. Le filtrage collaboratif</li> <li>3. Le filtrage basé sur la connaissance</li> </ol> </li> <li>■ Formellement parlant, étant donné un ensemble d'utilisateurs <math>U = \{U_1, U_2, \dots, U_n\}</math>, un ensemble d'items <math>I = \{I_1, I_2, \dots, I_m\}</math>, un ensemble de réactions sur les items correspondants <math>R = \{r_{ui}\}</math>, ici <math>1 \leq u \leq n, 1 \leq i \leq m</math>.  À noter qu'il ne faut pas interpréter comme quoi <math>r_{ui}</math> est définie pour toutes les valeurs admissibles de <math>u</math> et <math>i</math>.</li> </ul>	<ul style="list-style-type: none"> <li>■ Sur la base des réactions que l'utilisateur <math>U_u</math> aurait eu sur les items <math>I_u = \{I_i   r_{u,i} \in R\}</math>, on voudrait trouver un nouvel item <math>I_{i^*}</math> qui pourrait susciter son intérêt. <math display="block">I_{i^*} = \underset{I_j \in I \setminus I_u}{\operatorname{argmax}} f(U_u, I_j)</math> </li> <li>■ où <math>f</math> est une fonction d'estimation de la préférence des utilisateurs.</li> </ul>
<div> <div>UdS</div> <div> <div>Université de Sherbrooke</div> <div>11 novembre 2024</div> <div>Shengrui Wang</div> <div>10</div> </div> </div>	<div> <div>UdS</div> <div> <div>Université de Sherbrooke</div> <div>11 novembre 2024</div> <div>Shengrui Wang</div> <div>11</div> </div> </div>	<div> <div>UdS</div> <div> <div>Université de Sherbrooke</div> <div>11 novembre 2024</div> <div>Shengrui Wang</div> <div>12</div> </div> </div>

## 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 :

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
$X$	$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$
$U_1$	$r_{1,1}$		$r_{1,3}$	$r_{1,4}$	$r_{1,5}$	
$U_2$		$r_{2,2}$		$r_{2,4}$	$r_{2,5}$	
$U_3$	$r_{3,1}$	$r_{3,2}$	$r_{3,3}$			
$U_4$			$r_{4,3}$	$r_{4,4}$	$r_{4,5}$	

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

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
$X$	$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$
$U_1$	$r_{1,1}$		$r_{1,3}$	$r_{1,4}$	$r_{1,5}$	
$U_2$		$r_{2,2}$		$r_{2,4}$	$r_{2,5}$	
$U_3$	$r_{3,1}$	$r_{3,2}$	$r_{3,3}$			
$U_4$			$r_{4,3}$	$r_{4,4}$	$r_{4,5}$	

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

## Filtrage basé sur le contenu

- Les informations du contenu sont données sous forme matricielle  $X \in \mathbb{R}^{m \times d}$ ,  $X = (x_{i,j})_{1 \leq i \leq m, 1 \leq j \leq d}$
- À partir desquelles on peut créer, pour chaque utilisateur  $U_u$  des profils  $P_u$  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  $I_j$  dont le vecteur  $x_j$  serait similaire.

## Filtrage basé sur le contenu

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

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

- Approche Bayésienne:

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

## Filtrage basé sur le contenu

- Avantages:
  - Indépendance de l'utilisateur: 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.

## Filtrage basé sur le contenu

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  $I_j$  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 réactions chacun des utilisateurs  $U_u$  aurait eu à l'endroit des différents items.

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

2. Construction d'une matrice d'association;

UTILISATEUR	ITEM 1	ITEM 2	ITEM 3	ITEM 4
U <sup>1</sup>	★★★★		★★★★	★★★
U <sup>2</sup>		★★★★		★★★★
U <sup>3</sup>	★★★★	★★★★		
U <sup>4</sup>			★★★★	★★★★
U <sup>5</sup>	★★★★		★★★★	

3. Recommandation suivant les similarités dans les associations.

## Filtrage collaboratif

- La recommandation peut se faire :

- suivant les similarités entre utilisateurs : filtrage collaboratif utilisateur;



- suivant les similarité entre items: filtrage collaboratif item.



## Filtrage collaboratif



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

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

## Filtrage collaboratif

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  $U_{v \sim u} = \{U_v | U_v \neq U_u / I_u \cap I_v \neq \emptyset\}$ .
- Un vote auquel on s'attendrait que  $U_u$  ait sur l'item  $I_i$ , pourrait être

$$r_{u,i} = \left\lceil \left( \frac{1}{|U_{v \sim u}|} \sum_{U_v \in U_{v \sim u}} r_{v,i} \right) \right\rceil$$

## Filtrage collaboratif

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.



	ITEM 1	ITEM 2	ITEM 3	ITEM 4	ITEM 5
U1	★★★★	★★★★	★★★★	★★★★	★★★★
U2	★★★★	★★★★	★★★★	★★★★	★★★★
U3	★★★★	★★★★	★★★★	★★★★	★★★★
U4	★★★★	★★★★	★★★★	★★★★	★★★★
U5	★★★★	★★★★	★★★★	★★★★	★★★★

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

$$Recommend(I_i | U_u) = \begin{cases} I_j & \text{if } I_j \in \{I_k / \\ & I_k = \underset{I_k \in I, I_k}{argmax} sim(I_i, I_k) \} \\ \emptyset & \text{else} \end{cases}$$

## 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  $I_{j \sim i} = \{I_j / I_j = \underset{I_k \in I, I_k}{argmax} sim(I_i, I_k)\}$  l'ensemble des items similaires à  $I_i$
- Un vote qu'aurait l'utilisateur  $U_u$  sur l'item  $I_j$  serait,

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

## Filtrage collaboratif

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