

Une approche de recommandation hybride guidée par la sémantique

Roza Lémdani, Nacéra Bennacer, Géraldine Polailon, Yolaine Bourda

Supélec Sciences des Systèmes (E3S)

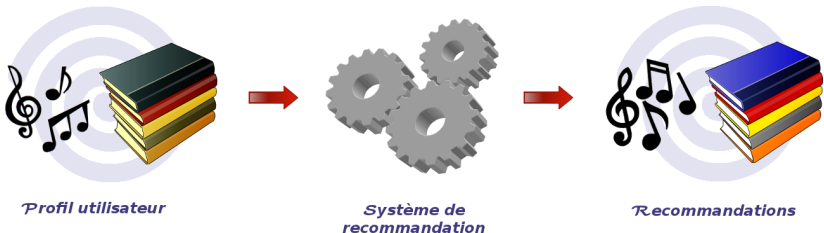
11 juin 2012



Contexte : la personnalisation

Definition

Les systèmes de recommandation sont des outils qui proposent des items aux utilisateurs.



Les systèmes de recommandation (SR)

Les approches utilisées

- Approches basées sur le contenu
- Approches collaboratives
- Approches hybrides

Les systèmes de recommandation (SR)

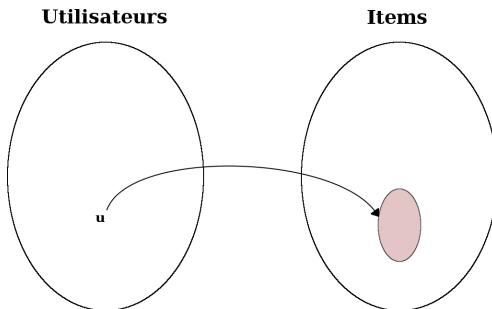
Les approches utilisées

- Approches basées sur le contenu
- Approches collaboratives
- Approches hybrides

Les systèmes de recommandation (SR)

Les approches utilisées

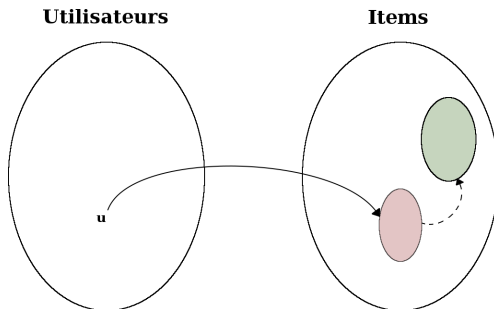
- Approches basées sur le contenu
- Approches collaboratives
- Approches hybrides



Les systèmes de recommandation (SR)

Les approches utilisées

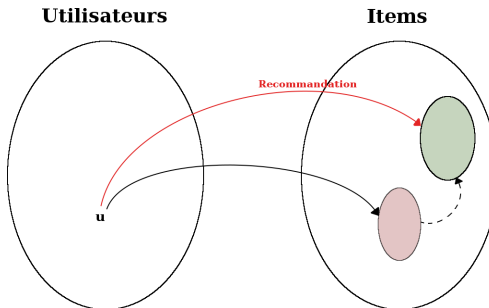
- Approches basées sur le contenu
- Approches collaboratives
- Approches hybrides



Les systèmes de recommandation (SR)

Les approches utilisées

- Approches basées sur le contenu
- Approches collaboratives
- Approches hybrides



Les systèmes de recommandation (SR)

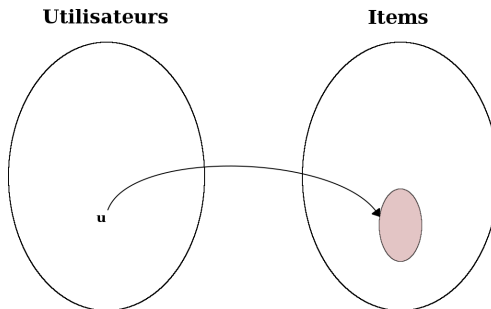
Les approches utilisées

- Approches basées sur le contenu
- **Approches collaboratives**
- Approches hybrides

Les systèmes de recommandation (SR)

Les approches utilisées

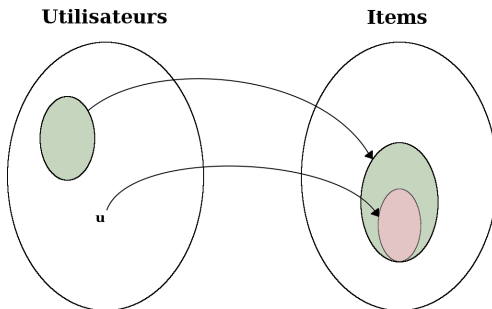
- Approches basées sur le contenu
- **Approches collaboratives**
- Approches hybrides



Les systèmes de recommandation (SR)

Les approches utilisées

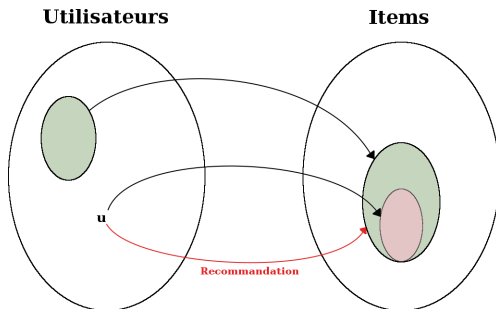
- Approches basées sur le contenu
- **Approches collaboratives**
- Approches hybrides



Les systèmes de recommandation (SR)

Les approches utilisées

- Approches basées sur le contenu
- **Approches collaboratives**
- Approches hybrides



Les systèmes de recommandation (SR)

Les approches utilisées

- Approches basées sur le contenu
- Approches collaboratives
- **Approches hybrides**

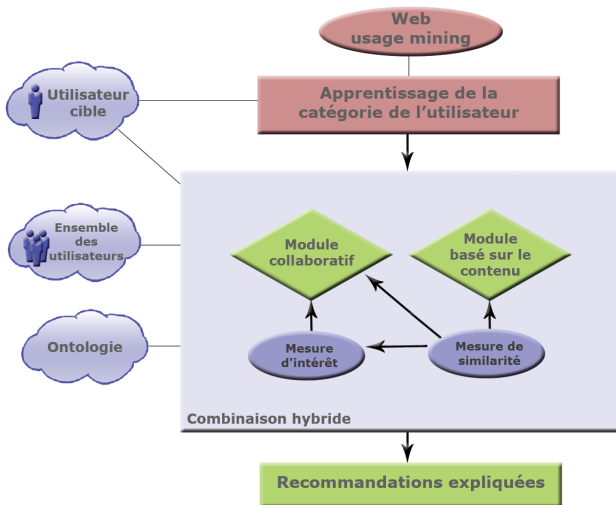
Objectifs

Objectifs

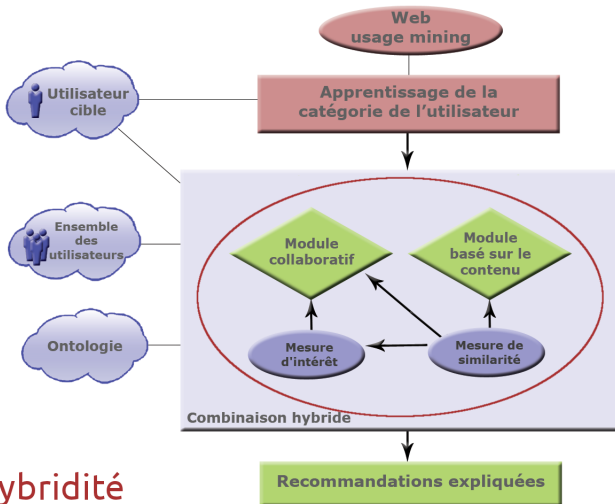
Définition d'un framework de recommandation :

- hybride
- indépendant du domaine
- adaptatif
- justifiant la recommandation

Le framework de recommandation

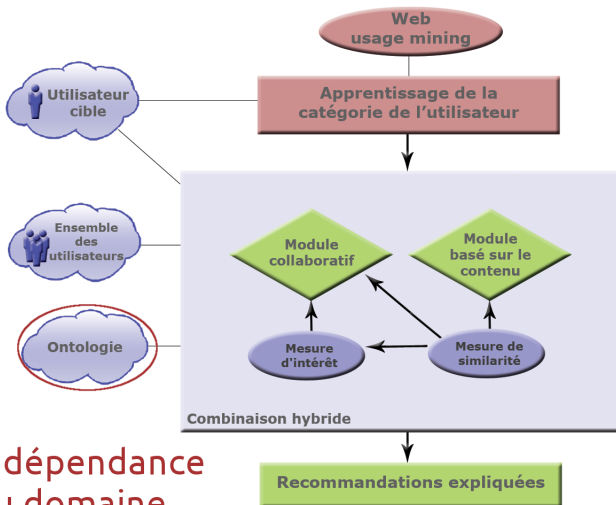


Le framework de recommandation



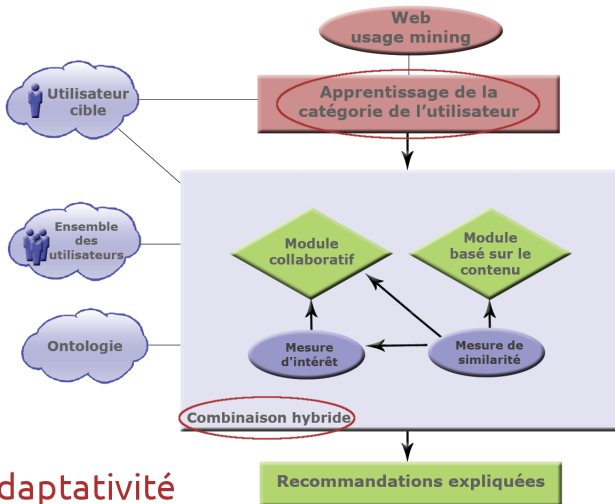
Hybridité

Le framework de recommandation

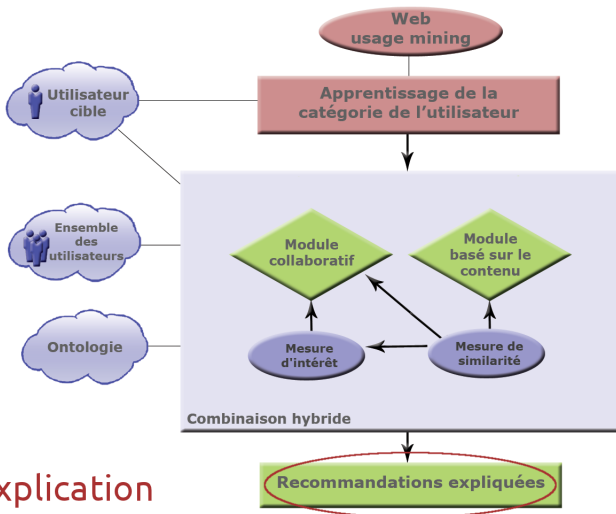


Indépendance
du domaine

Le framework de recommandation



Le framework de recommandation



Explication

Modèle utilisateur et items

Modèle utilisateur

Données explicites plus fiables (notes des items attribuées par l'utilisateur, formulaires)

Données implicites plus disponibles (clicks, recherches)

Modèle des items

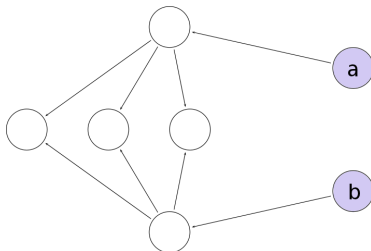
Les items sont décrits dans une ontologie qui contient :

- un ensemble $\mathcal{C}$ de concepts (Ex : le concept Acteur)
- un ensemble $\mathcal{R}$ de relations (Ex : la relation aJoueDans entre les concepts Acteur et Film)
- un ensemble $\mathcal{I}$ d'instances (Ex : Tom Hanks est une instance du concept Acteur)

Mesure de similarité

Principe

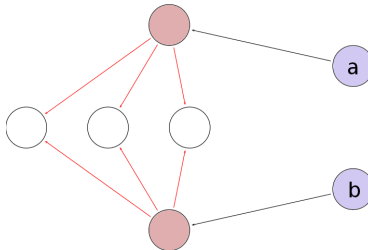
Deux instances de concepts de l'ontologie, a et b , sont similaires si elles sont reliées aux mêmes instances par une propriété quelconque, ou si elles sont reliées à des instances elles-mêmes similaires.



Mesure de similarité

Principe

Deux instances de concepts de l'ontologie, a et b , sont similaires si elles sont reliées aux mêmes instances par une propriété quelconque, ou si elles sont reliées à des instances elles-mêmes similaires.



Mesure d'intérêt

Principe

Évaluer pour chaque item, l'intérêt de l'utilisateur pour chaque concept de cet item.

Exemple

$$\begin{aligned}
 interet^{inst}(u, "S. Kubrick") &= mediane(\\
 &\quad note(u, "Dr. Folamour"), \\
 &\quad note(u, "Orange Mecanique")) \\
 &= mediane(2, 1) \\
 &= 1.5 \\
 interet^{Realisateur}(u, "Shining") &= interet^{inst}(u, "S. Kubrick") \\
 &= 1.5
 \end{aligned}$$

Module collaboratif : extraction des règles d'association

	Film ₁ ^l	Film ₁ ^d	Film ₂ ^l	Film ₂ ^d	Film ₃ ^l	Film ₃ ^d	Film ₄ ^l	Film ₄ ^d	Film ₅ ^l	Film ₅ ^d
u_1	×				×		×		×	
u_2		×	×		×				×	
u_3	×			×	×			×	×	
u_4		×	×		×				×	
u_5	×						×		×	
u_6	×			×				×		

Table: Matrice d'extraction de règles

- Utilisation de l'algorithme Apriori pour la génération [Agrawal et al. 1994]
- Exemple de règles extraites :

$$r_1 : \{ \text{Film}_1^l, \text{Film}_3^l \} \Rightarrow \{ \text{Film}_5^l \}$$

$$r_2 : \{ \text{Film}_1^l, \text{Film}_2^d \} \Rightarrow \{ \text{Film}_4^d \}$$

$$r_3 : \{ \text{Film}_1^l, \text{Film}_4^l \} \Rightarrow \{ \text{Film}_5^l \}$$

$$r_4 : \{ \text{Film}_2^l, \text{Film}_3^l \} \Rightarrow \{ \text{Film}_1^l, \text{Film}_5^l \}$$

Module collaboratif : extension et sélection des RA

- Ajouter des règles d'association telles que leurs prémisses soient composées d'items similaires aux items des règles générées par l'algorithme Apriori.
- Extraire les éléments en conclusion de règle non vus par l'utilisateur cible tel quel les éléments en prémisse de règle appartiennent tous à l'historique dudit utilisateur.

Exemple

$$\{\text{Sacré Graal}', \text{Jurassic Park}'\} \Rightarrow \{\text{Indiana Jones}'\}$$

'Sacré Graal' et 'La vie de Brian' étant similaires, la règle suivante est ajoutée :

$$\{\text{La vie de Brian}', \text{Jurassic Park}'\} \Rightarrow \{\text{Indiana Jones}'\}$$

Module basé sur le contenu

Principe

Inférer les items intéressants pour l'utilisateur qui sont similaires à ceux de son historique.

Exemple

Soit l'utilisateur u qui a noté le film « La vie de Brian ». Soit le film « Sacré Graal », similaire à « La vie de Brian », et un seuil $\beta = 3$.

$$\begin{aligned} eval(u, "Sacré Graal") &= sim_{10}("Sacré Graal", "La vie de Brian") \\ &\quad \times note(u, "La vie de Brian") \\ &= 0.79 \times 5 \\ &= 3.95 \end{aligned}$$

Module basé sur le contenu

Principe

Inférer les items intéressants pour l'utilisateur qui sont similaires à ceux de son historique.

Exemple

Soit l'utilisateur u qui a noté le film « La vie de Brian ». Soit le film « Sacré Graal », similaire à « La vie de Brian », et un seuil $\beta = 3$.

$$\begin{aligned} eval(u, "Sacré Graal") &= sim_{10}("Sacré Graal", "La vie de Brian") \\ &\quad \times note(u, "La vie de Brian") \\ &= 0.79 \times 5 \\ &= 3.95 \end{aligned}$$

Module basé sur le contenu

Principe

Inférer les items intéressants pour l'utilisateur qui sont similaires à ceux de son historique.

Exemple

Soit l'utilisateur u qui a noté le film « La vie de Brian ». Soit le film « Sacré Graal », similaire à « La vie de Brian », et un seuil $\beta = 3$.

$$\begin{aligned} eval(u, "Sacré Graal") &= sim_{10}("Sacré Graal", "La vie de Brian") \\ &\quad \times \textcolor{red}{note}(u, \textcolor{red}{La vie de Brian}) \\ &= 0.79 \times 5 \\ &= 3.95 \end{aligned}$$

Module basé sur le contenu

Principe

Inférer les items intéressants pour l'utilisateur qui sont similaires à ceux de son historique.

Exemple

Soit l'utilisateur u qui a noté le film « La vie de Brian ». Soit le film « Sacré Graal », similaire à « La vie de Brian », et un seuil $\beta = 3$.

$$\begin{aligned} eval(u, "Sacré Graal") &= sim_{10}("Sacré Graal", "La vie de Brian") \\ &\quad \times note(u, "La vie de Brian") \\ &= 0.79 \times 5 \\ &= 3.95 \end{aligned}$$

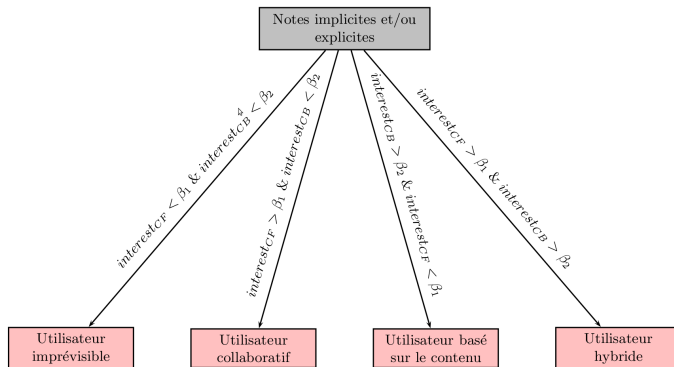
Apprentissage de la catégorie de l'utilisateur

Principe

Filtrage des données Détecter les données à utiliser pour l'utilisateur (explicites et/ou implicites).

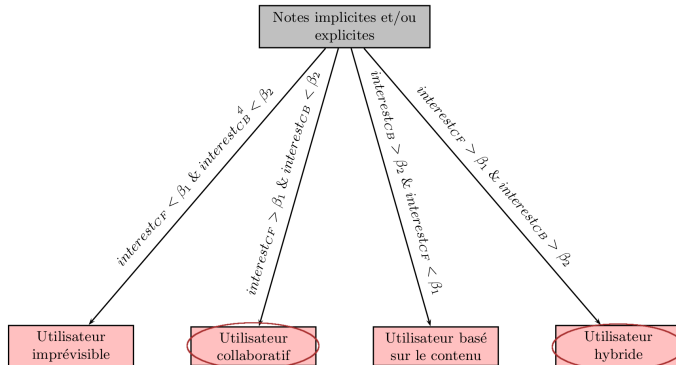
Classification de l'utilisateur Analyser les données sélectionnées afin de découvrir le type de recommandations auxquelles l'utilisateur est le plus sensible.

Classification de l'utilisateur



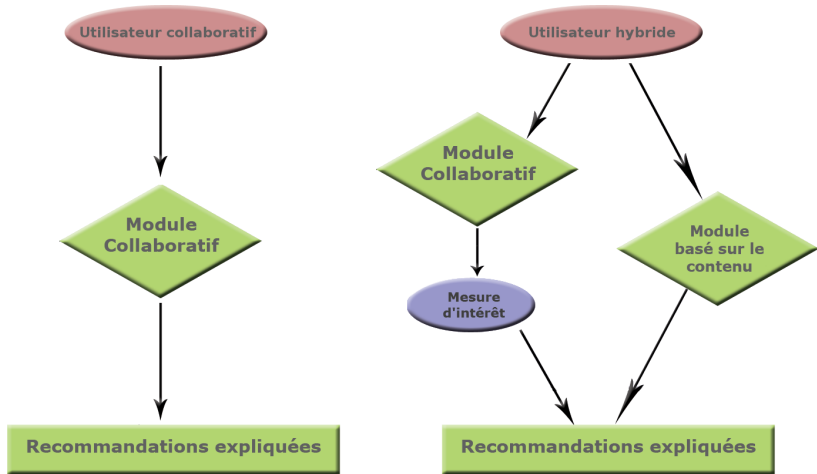
- $interest_{CB}$ est le degré d'intérêt de l'utilisateur pour les recommandations basées sur le contenu
- $interest_{CF}$ est le degré d'intérêt de l'utilisateur pour les recommandations collaboratives
- β_1 et β_2 sont des seuils fixés par l'expert

Classification de l'utilisateur



- $interest_{CB}$ est le degré d'intérêt de l'utilisateur pour les recommandations basées sur le contenu
- $interest_{CF}$ est le degré d'intérêt de l'utilisateur pour les recommandations collaboratives
- β_1 et β_2 sont des seuils fixés par l'expert

Classification de l'utilisateur



Expérimentation

Data

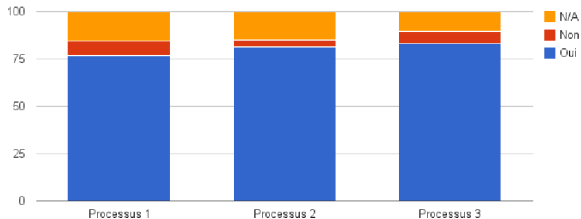
- Sous-ensemble du dataset de MovieLens
- 523 films, 934 utilisateurs et 67 217 notes
- 1813 acteurs principaux, 413 réalisateurs, 876 auteurs, 54 langues, 33 pays et 24 genres
- Données décrites dans une ontologie OWL

Users

- 12 volontaires âgés de 16 à 36 ans ayant noté au moins 30 films chacun
- Comparaison de 3 processus de recommandation
- Évaluation des recommandations (oui/non)

Evaluation

- Processus 1** Module collaboratif sans mesure de similarité sémantique (règles d'association classiques).
- Processus 2** Module collaboratif + mesure d'intérêt, sans mesure de similarité sémantique.
- Processus 3** Module collaboratif + mesure de similarité sémantique (règles d'association étendues) + mesure d'intérêt + module basé sur le contenu.



Conclusion

- Système de recommandation hybride
- Processus de recommandation personnalisé
- Plusieurs modules indépendants permettant d'apporter une explication de la recommandation
- Approche générique, utilisable quel que soit le domaine
- Évaluation du système sur de vrais utilisateurs. Premiers résultats encourageants