
Conception et formalisation d'un système de recommandation de prestations touristiques

Olivier Hotel¹, Frédéric Pourraz¹, Hervé Verjus¹

Université Savoie Mont-Blanc

LISTIC

Annecy, France

{olivier.hotel, frederic.pourraz, herve.verjus}@univ-smb.fr

RÉSUMÉ. Dans cet article, nous introduisons la notion d'avatar pour les systèmes de recommandation. Cette notion permet non seulement d'apporter une solution au problème du démarrage à froid mais aussi de manipuler des informations incomplètes concernant les utilisateurs du système de recommandation. Nous décrivons également un processus en trois étapes permettant d'affiner ces informations : identification des avatars, partitionnement des utilisateurs et application de la méthode des facteurs latents. Enfin, nous décrivons un exemple de notre méthodologie dans le cadre d'un système de recommandation de prestations touristiques.

ABSTRACT. In this article, we introduce the notion of avatar for recommender systems to solve the cold start problem and to manipulate incomplete information about the users of the recommender system. We also describe a three-step process for refining this information: avatars identification, clustering of the users, latent factors method. Finally, we describe a concrete example of our methodology as part of a recommender system of tourist services.

MOTS-CLÉS : Système de Recommandation - Informations Incomplètes - Système d'aide à la Décision - Avatar - Partitionnement - Facteurs Latents

KEYWORDS: Recommender System - Incomplete Information - Business Intelligence - Avatar - Clustering - Latent Factors

. -/

1. Introduction

De nos jours et avec le volume croissant de données disponibles, la conception de systèmes de recommandation est un axe de recherche important. Ces systèmes sont largement utilisés dans les plateformes de commerce en ligne et tendent à se développer dans de nombreux secteurs. Sans remplacer l'humain, un système de recommandation l'aide dans ses choix en lui faisant des suggestions. Il appartient à l'humain, de l'utilisateur novice à l'expert aguerri, de suivre les propositions qui lui sont faites par le système de recommandation, de s'inspirer de ces propositions tout en les accommodant ou, de ne pas en tenir compte.

Cet article propose un cadre générique et indépendant d'un domaine métier spécifique pour concevoir un système de recommandation couvrant les spécificités suivantes: au regard d'un ensemble d'articles, d'utilisateurs et de notes attribuées aux articles par les utilisateurs, le système de recommandation sera en capacité de faire des propositions pertinentes à l'utilisateur, qu'il ait ou non connaissance au préalable de ce dernier. Pour cela, et après avoir présenté un état de l'art des méthodes de recommandation et des problématiques qui s'y rapportent en section 2, nous introduisons une notion d'avatar au sein du système de recommandation en section 3; et nous décrivons des mécanismes d'adaptation en section 4.

2. État de l'art

Sans prétendre à une totale exhaustivité, cette section s'efforce de présenter succinctement un état de l'art des méthodes de recommandation proposées dans la littérature s'inscrivant dans le cadre précédemment décrit. Une description plus détaillée est disponible dans (Béchet, 2016; Benouaret, 2017). Ces méthodes peuvent être divisées en trois familles:

1. les méthodes basées sur le contenu (Pazzani, Billsus, 2007) consistent essentiellement à comparer les attributs des objets avec les attributs d'intérêt de l'utilisateur actif;
2. les méthodes basées sur des principes de filtrage collaboratif (Schafer *et al.*, 2007) reposent sur la conjecture que les opinions des autres utilisateurs peuvent être utilisées pour prédire les préférences des autres utilisateurs, et sur l'hypothèse que si des utilisateurs ont les mêmes préférences sur certains objets alors ils auront également les mêmes préférences sur d'autres objets; et
3. les méthodes hybrides qui combinent les deux précédentes approches de recommandation.

Ces dernières années, des propositions ont émergé (Chen, Chen, 2015; Hernández-Rubio *et al.*, 2018) afin de cerner le profil de l'utilisateur à partir d'informations recueillies sur Internet, des traces numériques que l'utilisateur aurait pu laisser (réseaux sociaux, forums, blogs,...); ces approches tentent d'éviter le problème du démarrage à froid en faisant l'hypothèse que des informations permettant de qualifier l'utilisateur existent et sont disponibles.

Système de recommandation avec avatars

Bien que les approches décrites dans cette section permettent, dans la plupart des cas, de recommander des objets pertinents aux utilisateurs, elles sont inopérantes lorsqu'aucune information concernant les objets ou les utilisateurs n'est disponible. Ce problème de démarrage à froid est encore largement ouvert même si des approches permettant d'ajouter un nouvel utilisateur ou un nouvel objet au système de recommandation ont été proposées dans la littérature (Negre *et al.*, 2013).

3. Proposition d'avatars

3.1. Définition

Dans le cadre de notre approche, nous introduisons la notion d'avatars pour résoudre le problème du démarrage à froid. Un avatar est une représentation fictive choisie par l'utilisateur pour représenter ces intérêts.

Dans la phase initiale de fonctionnement du système, l'identification des avatars doit répondre à un triple objectif :

1. les utilisateurs doivent pouvoir s'identifier ne serait-ce que partiellement à un des avatars proposés par le système;
2. il doit être possible d'associer des notes fictives à chaque couple *objet – avatar*; et
3. le nombre d'avatars doit être limité pour que chaque utilisateur puisse rapidement choisir son avatar.

3.2. Étude de cas

Dans le cadre du projet Européen Interreg France-Suisse Transfrontour dont un des objectifs est la conception d'un système de recommandation de prestations touristiques, des études et des sondages réalisés auprès des acteurs du tourisme du département de la Haute-Savoie (74) ont permis d'identifier les avatars des usagers des différents services touristiques : *tourisme d'affaires*, *tourisme familial*, *tourisme culturel*, *tourisme gastronomique*, *tourisme sportif* et *tourisme d'hiver*.

Ces mêmes études ont également permis d'identifier les différentes prestations touristiques, de les classer par grands domaines (*hébergement*, *restauration*, *activités culturelles*, *activités sportives*, etc.) et par sous-domaines (restauration : *gastronomie - snack*; activités culturelles : *patrimoine - art*; activités sportives : *sport outdoor - sport indoor*, etc.).

Dans l'objectif de pouvoir associer une note à chaque couple *avatar – prestation*, nous proposons d'utiliser les sous-domaines identifiés comme des descripteurs permettant de représenter formellement les avatars des usagers ainsi que les différentes prestations touristiques et nous associons à chaque descripteur un poids variant par exemple de 0 à 6 (0 = "non pertinent" et 6 = "tout à fait pertinent"). Par exemple, les

. -/

avatars *tourisme sportif* et *tourisme culturel* sont respectivement représentés par les vecteurs :

$$\begin{aligned}
 \text{tourisme sportif} &= \{gastronomie = 0, snack = 0, patrimoine = 0, art = 0, \\
 &\quad sport\ outdoor = 6, sport\ indoor = 6\}; \text{ et} \\
 \text{tourisme culturel} &= \{gastronomie = 3, snack = 0, patrimoine = 6, art = 6, \\
 &\quad sport\ outdoor = 0, sport\ indoor = 0\}.
 \end{aligned}$$

et les prestations touristiques "kayak" et "musée d'Annecy" sont respectivement représentées par les vecteurs :

$$\begin{aligned}
 \text{kayak} &= \{gastronomie = 0, snack = 0, patrimoine = 0, art = 0, \\
 &\quad sport\ outdoor = 6, sport\ indoor = 0\}; \text{ et} \\
 \text{musée} &= \{gastronomie = 0, snack = 0, patrimoine = 6, art = 6, \\
 &\quad sport\ outdoor = 0, sport\ indoor = 0\}.
 \end{aligned}$$

La note associée à tout couple *utilisateur* – *prestation* est alors définie comme le résultat du produit scalaire entre le vecteur descripteur de l'avatar de l'utilisateur et le vecteur descripteur de la prestation. Par exemple, la note d'un utilisateur *sportif* sur la prestation *kayak* est 36 tandis que la note de ce même utilisateur sur la prestation *musée* est de 0 : cet utilisateur préfère le kayak au musée.

4. Formalisation du système de recommandation

Le système de recommandation est formalisé par une matrice **N** de notes contenant à la fois les notes explicitement fournies par les utilisateurs et celles générées à partir des avatars de ceux-ci. La figure 1 représente cette matrice avec en vert les notes des utilisateurs et en jaune celles des avatars.

	Objet 1	Objet 2		Objet O		
Utilisateur 1	1	2	3	1	1	5
Utilisateur 2	2	2	3	1	1	5
	3	2	2	1	1	5
	1	3	3	4	5	3
Utilisateur U	1	3	3	5	4	3

Figure 1. Représentation matricielle de l'état du système de recommandation avec en vert les notes explicites des utilisateurs et en jaune celles des avatars.

Système de recommandation avec avatars

Dans le reste de ce document, nous distinguons les notes explicites obtenues des utilisateurs de celles des avatars et nous notons N_{ult} l'ensemble des notes des utilisateurs et N_{avt} l'ensemble des notes des avatars.

Le principe de recommandation consiste à suggérer aux utilisateurs les objets ayant les meilleures notes. Cependant, comme les notes des avatars sont intrinsèquement approximatives, cette approche peut mener à des recommandations peu pertinentes. Nous proposons une méthode permettant d'adapter les notes des avatars en utilisant celles explicitement obtenues des utilisateurs. La méthode proposée consiste à grouper les utilisateurs ayant des intérêts communs et à appliquer pour chacun de ces groupes la méthode des facteurs latents pour permettre un rétrocontrôle des notes des utilisateurs sur les notes des avatars. L'opération de partitionnement des utilisateurs permet d'améliorer la précision de la méthode des facteurs latents en minimisant son erreur quadratique moyenne.

5. Adaptation des notes des avatars

Dans cette section, nous décrivons la méthode employée pour adapter les notes des avatars dans l'objectif de personnaliser les avatars de chaque utilisateur et de mieux définir les intérêts des utilisateurs envers les objets.

La méthode proposée, permettant d'adapter les notes des avatars en utilisant celles des utilisateurs, consiste à :

1. partitionner les utilisateurs, les lignes de la matrice \mathbf{N} , en groupes homogènes de manière à diviser cette dernière en k sous-matrices \mathbf{N}_k en utilisant la méthode des *k-moyennes* (Arthur, Vassilvitskii, 2007 ; Kodinariya, Makwana, 2013);
2. effectuer une factorisation $\mathbf{P}_k \mathbf{Q}_k^T$ de ces sous-matrices (Koren *et al.*, 2009); et à
3. mettre à jour les notes des avatars selon l'équation :

$$\forall k, \forall (u, o), u \in \{1, \dots, \text{card}(U)\} o \in \{1, \dots, \text{card}(O)\} : r_{u,o} \in N_{avt}$$

$$r_{u,o} \leftarrow \alpha \text{Avt}_{u,o} + (1 - \alpha) p_{k,u} q_{k,o}^T.$$

$p_{k,u}$ et $q_{k,o}$ représentent respectivement les vecteurs des matrices \mathbf{P}_k et \mathbf{Q}_k si $r_{u,o}$ appartient à la partition k ; et $\text{Avt}_{u,o}$ représente la note générée par l'avatar de l'utilisateur u pour l'objet o en calculant le produit scalaire de leurs vecteurs descripteurs respectifs tels que décrits dans la section 3. Le coefficient $\alpha \in [0; 1]$ permet de contrôler l'influence de l'estimation dans le processus d'adaptation et, au cours des mises à jour successives des notes, d'oublier la note initiale de l'avatar.

6. Conclusion

L'approche que nous proposons est assez générique pour être expérimentée dans des domaines d'application différents. L'introduction de la notion d'avatar est intéressante pour permettre d'initialiser le système et apporte une réponse à la question du

. -/

démarrage à froid. Notre proposition permet également d'adapter dynamiquement les notes initialement générées par les avatars pour les rendre de plus en plus pertinentes pour les utilisateurs.

Une des limites du cas d'étude telle que réalisée actuellement est que l'identification des descripteurs et l'attribution des poids à ces derniers sont réalisées manuellement. Cette question peut faire l'objet de perspectives à ce travail.

Remerciements

Ces travaux ont été en partie financés par le projet Interreg Franco-Suisse Transfrontour (2016-2019) sur des fonds FEDER.

Bibliographie

- Arthur D., Vassilvitskii S. (2007). k-means++: The advantages of careful seeding. In *Proc. of the 18th annual acm-siam symposium on discrete algorithms*, p. 1027 - 1035.
- Béchet N. (2016). *état de l'art sur les systèmes de recommandation*. Rapport technique. INRIA.
- Benouaret I. (2017). *Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels*. Thèse de doctorat non publiée, Université de Technologie de Compiègne.
- Chen G., Chen L. (2015, août). Augmenting service recommender systems by incorporating contextual opinions from user reviews. *User Modeling and User-Adapted Interaction*, vol. 25, n° 3, p. 295–329. Consulté sur <https://doi.org/10.1007/s11257-015-9157-3>
- Hernández-Rubio M., Cantador I., Bellogín A. (2018, novembre). A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews. *User Modeling and User-Adapted Interaction*. Consulté sur <https://doi.org/10.1007/s11257-018-9214-9>
- Kodinariya T. M., Makwana P. R. (2013). Review on determining number of cluster in k-means clustering. *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, n° 6, p. 90 - 95.
- Koren Y., Bell R., Volinsky C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer Society*, vol. 9, n° 1, p. 42 - 49.
- Negre E., Ravat F., Teste O., Tournier R. (2013). Problème du démarrage à froid pour les système de recommandation. In *Proc. of inforsid 2013*, p. 439 - 454.
- Pazzani M. J., Billsus. (2007). The adaptive web: Methods and strategies of web personalization. In P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), vol. 4321, p. 325-341. Springer-Verlag, Berlin Heidelberg New York.
- Schafer J. B., Frankowski D., Herlocker J., Sen S. (2007). The adaptive web: Methods and strategies of web personalization. In P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), vol. 4321, p. 291-324. Springer-Verlag, Berlin Heidelberg New York.