HAL open science

Systèmes de recommandation et Recherche d'Information

Laurent Candillier, Etienne Chai, Estelle Delpech

To cite this version:

Laurent Candillier, Etienne Chai, Estelle Delpech. Systèmes de recommandation et Recherche d'Information. Journée d'étude " Systèmes de recommandation " (CNAM), Jun 2012, Paris, France. 6 p. hal-00912323

HAL Id: hal-00912323

https://hal.science/hal-00912323v1

Submitted on 2 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Systèmes de recommandation et Recherche d'Information

Laurent Candillier[*,] — Étienne Chai[*,**] — Estelle Delpech[*,**]**

[*] *Nomao, 1 avenue Jean Rieux, 13 500 Toulouse, France, www.nomao.com*
[**] *Ebuzzing, 98 rue du Cherche-Midi, 75 006 Paris, France, www.ebuzzing.com*

ABSTRACT. *This paper deals with the usage of recommender systems in information retrieval systems. We focus on the case of local businesses search through the presentation of Nomao's local and recommendation search engine.*

RÉSUMÉ. *Cet article traite de l'application des technologies de recommandation aux systèmes de Recherche d'Information. Nous y détaillons un cas d'application à la recherche de lieux à travers la présentation de l'annuaire Nomao.*

KEYWORDS: *collaborative filtering, descriptive profiling, e-reputation, social recommendation, Iocal search.*

MOTS-CLÉS : *filtrage collaboratif, profiling descriptif, e-réputation, recommandation sociale, recherche locale.*

1re soumission à *Actes de la journée d'étude sur les moteurs de recommandation (CNAM)*, le 26/06/2013

1. Introduction : les acteurs de la recommandation sur Internet

La masse des données échangées aujourd'hui sur Internet constitue un atout sans précédent pour l'accès de tous à l'information. En parallèle, elle constitue également une difficulté car elle oblige à d'importants traitements de filtrage pour que les données retournées soient effectivement pertinentes pour l'utilisateur. Dans ce cadre, la prise en compte du contexte dans lequel est soumise une requête fait l'objet d'une attention particulière de la part de la communauté en Recherche d'Information.

Les différents acteurs du Web qui développent ces outils de recommandations sont donc de plus en plus nombreux aujourd'hui. Le site web de vente en ligne Amazon a en effet accru significativement le nombre de ses visites et de ses ventes grâce à cet outil. Certains utilisateurs vont même maintenant sur ce site principalement dans le but de bénéficier de ces recommandations, même s'ils n'ont originellement pas l'intention d'y acheter de produit. Et il en va de même pour des sites comme Allociné pour les cinéphiles, ou de Deezer pour la musique. C'est aussi le cas pour OverBlog, la première plateforme de blogs en France, qui a intégré depuis peu un nouveau module de mise en relation entre articles similaires.

Dans cet article, nous ferons un bref survol des diverses méthodes de recommandation existantes (section 2) puis nous détaillerons un cas d'application à la recherche de commerces locaux (section 3).

2. Approches de la recommandation

Historiquement, les moteurs de recommandation utilisent deux techniques majeures :

– Le filtrage collaboratif cherche à croiser les navigations des utilisateurs pour faire profiter les uns des liens entre données récoltés sur les autres. La liaison entre un item et un utilisateur se fait sur la base de l'hypothèse suivante : "Les utilisateurs qui ont aimé les mêmes items que l'utilisateur λ, aiment aussi ces items là, que nous pouvons recommander à λ".

– Le profiling descriptif cherche plutôt à établir des profils descriptifs des utilisateurs puis à comparer ces profils aux données rencontrées pendant leur navigation. L'hypothèse de travail est : "L'utilisateur λ aiment les items présentant les caractéristiques $\{c_1, ...c_n\}$, nous pouvons lui recommander des items présentant ces mêmes caractéristiques".

Une évolution logique des systèmes de recommandation est d'exploiter les possibilités offertes par le Web 2.0 et en particulier sa dimension sociale. Dans ce cadre, on peut distinguer deux méthodes de recommandation :

– L'e-réputation se base sur les notes attribuées à un item par l'ensemble des internautes. L'hypothèse suivante : "Plus les internautes ont aimé un item, plus il est pertinent de le recommander à l'utilisateur".

– La recommandation sociale exploite les goûts des internautes appartenant au réseau social de l'utilisateur. La recommandation d'un pair est souvent jugée plus fiable que la recommandation d'un internaute inconnu. L'hypothèse est alors : "Les pairs d'un internaute ont aimé un item, nous pouvons le recommander à cet internaute".

Ces nouvelles approches n'éludent pas pour autant l'intérêt du système de recommandations basé sur ses goûts personnels, car nos amis ne partagent pas forcément nos goûts, mais elle a l'intérêt d'ajouter une dose de confiance supplémentaire dans la suggestion qui est réalisée.

3. Nomao : recherche et recommandation de lieux

Nomao est une application Web et mobile permettant à ses utilisateurs de trouver des lieux qui correspondent à leurs goûts. Nomao présente les mêmes fonctionnalités qu'un annuaire (un champ de recherche *quoi* et un champ de recherche *où*). Les requêtes sont du type [TYPE DE LIEU] [EMPLACEMENT] (ex: "restaurant Paris") où [NOM DE LIEU] [EMPLACEMENT] (ex: "Chartier Paris"). Le moteur de recherche intégré à Nomao propose alors une liste de résultats ordonnés en fonction de leur degré d'appariement avec la requête de l'utilisateur, d'un score de recommandation correspondant à l'une des approches présentées dans la section 2 et, dans le cas de l'application mobile, en fonction de la proximité du lieu.

La méthode de recommandation est laissée au choix de l'utilisateur :

– Par popularité : ce score prend en compte l'e-réputation des lieux ;

– Par affinité : ce score exploite filtrage collaboratif et profiling descriptif ;

– Par amis : ce score prend en compte les recommandations des amis Facebook de l'utilisateur.

La figure 1 en annexe donne un exemple de résultat du moteur de recherche qui intègre ces trois recommandations.

3.1. *Score de popularité*

Le score de popularité est basé sur les notes attribuées au lieu par les internautes. Cinq notes sont possibles : {1, 2, 3, 4, 5}. Le score de popularité correspond à la moyenne des notes, ramenée à une échelle de 0 à 1 (équation 1).

$$\text{popularité}(i) = \frac{\sum_{n \in N(i)} 0.2n}{||N(i)||} \qquad [1]$$

où $N(i)$ est l'ensemble des notes attribuées au lieu i.

3.2. *Score d'affinité*

Le score d'affinité est calculé prioritairement à partir du filtrage collaboratif.

3.2.1. *Filtrage collaboratif*

Le filtrage collaboratif est basé sur une matrice $A = (a_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$ où $I = \{1, ..., m\}$ est l'ensemble des utilisateurs ; $J = \{1, ..., n\}$ est l' ensemble des lieux et $a_{i,j}$ est un booléen indiquant si l'utilisateur i aime le lieu j.

De cette matrice peuvent être déduites des similarités entre lieux. Nous considérons deux lieux comme similaires s'ils tendent à être aimés par les mêmes utilisateurs. Le score de similarité entre deux lieux l et l' est donné par le coefficient Jaccard (équation 2).

$$\text{similarité}(l, l') = \frac{||U(l) \cap U(l')||}{||U(l) \cup U(l')||} \qquad [2]$$

où $U(x)$ représente l'ensemble des utilisateurs ayant aimé le lieu x: $U(x) = (u|a_{xu} \text{ is TRUE})$.

Le score d'affinité FC entre un utilisateur u et un lieu l à recommander correspond à la moyenne des similarités entre les lieux aimés par l'utilisateur et les lieux similaires au lieu à recommander (équation 3).

$$\text{affinité FC}(l, u) = \frac{\sum_{(l', l'') \in L(u) \times knn(l)} sim(l', l'')}{||L(u) \times knn(l)||} \qquad [3]$$

où $L(u) = (l|a_{lu} \text{ is TRUE})$ (l'ensemble des lieux aimés par l'utilisateur u) et $knn(l)$ sont les k lieux les plus similaires à l (au sens de la mesure de similarité donnée dans l'équation 2).

La principale limite du filtrage collaboratif est qu'il ne peut pas recommander un lieu qui n'a été aimé par aucun utilisateur. Ceci peut arriver dans le cas de lieux nouvellement intégrés dans la base. Dans ce cas, Nomao utilise le profiling descriptif.

3.2.2. *Profiling descriptif*

Les lieux indexés par Nomao sont associés à un ensemble de descripteurs caractéristiques de leur activité. Par exemple un restaurant pourra être associé avec les descripteurs *manger, sushis, pas cher*. Le score d'affinité PD prend en compte les descripteurs associés aux lieux aimés par l'utilisateur $D(L(u))$ et les descripteurs associés au lieu à recommander $D(l)$:

$$\text{affinité PD}(l, u) = \frac{||D(l) \cap D(L(u))||}{||D(l) \cup D(L(u))||} \qquad [4]$$

3.3. *Recommandation sociale*

Le score de recommandation sociale prend en compte le nombre d'amis Facebook de l'utilisateur qui ont aimé le lieu à recommander (équation 5). Il s'appuie sur la matrice A décrite en section et sur une deuxième matrice F indiquant les liens d'amitié entre utilisateurs (f_{ij} vaut TRUE si l'utilisateur i et l'utilisateur j sont amis).

$$\textit{recommandation sociale}(l, u) = ||\{u' | f_{uu'} \text{ is TRUE and } a_{lu'} \text{ is TRUE}\}|| \qquad [5]$$

4. Perspectives : vers des systèmes de recommandation interactifs

Une prochaine évolution des systèmes de recommandation portera sur l'interaction entre les utilisateurs et le système.

Cet interaction sera utile pour contourner le problème du démarrage à froid : dans le cas de Nomao, lorsqu'un nouvel utilisateur arrive, nous ne connaissons pas ces goûts, il est donc difficile de lui recommander des lieux sur la base du filtrage collaboratif ou du profiling descriptif. Face à un nouvel utilisateur, le système de recommandation devra interagir avec lui de façon à acquérir un maximum de connaissances sur ses goûts en un minimum de temps.

Le système de recommandation doit également être capable d'expliquer son comportement à l'utilisateur : ceci peut se faire au niveau local ("nous vous recommandons ce lieu car vous aimez tel lieu / tel type de lieu") ou au niveau global (ex.: barres de progression indiquant le potentiel de recommandation pour un type de lieu). Enfin, nous envisageons de permettre à l'utilisateur de corriger la recommandation en indiquant les lieux qu'ils ne souhaitent pas ou plus que le système lui recommande.

Annexe

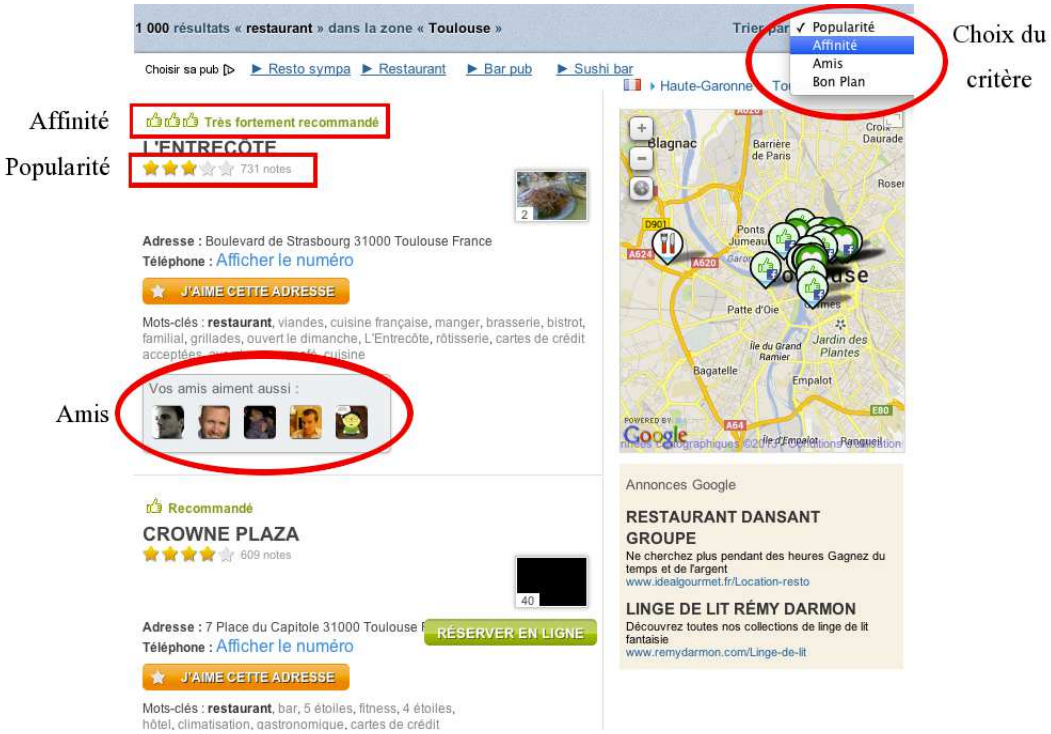

Figure 1. *Résultats du moteur de recherche de Nomao incluant les recommandations par popularité, affinité et amis*