# Filtrage collaboratif et prise en compte de la polarité des jugements

**Luc-Aurélien Gauthier — Benjamin Piwowarski — Patrick Gallinari**

*Sorbonne Universités, UPMC Univ Paris 06, CNRS*
*Laboratoire d'Informatique de Paris 6*
*4, place Jussieu*
*75 005 Paris*

*{Luc-Aurelien.Gauthier,Benjamin.Piwowarski,Patrick.Gallinari}@lip6.fr*

RÉSUMÉ. *Nous nous intéressons à la recommandation par des systèmes de filtrage collaboratif. Nous proposons de combiner une approche globale basée sur une factorisation matricielle et une approche locale basée sur l'exploitation directe d'un voisinage de l'utilisateur. L'hypothèse explorée dans l'article est que les jugements utilisateurs ont une sémantique et donc une utilité différente suivant qu'ils sont positifs ou négatifs. Nous proposons un modèle qui exploite cette polarité et apprend à pondérer l'influence de ses voisins en se basant sur des caractéristiques de polarité. Des expériences effectuées sur deux corpus permettent d'évaluer la validité de ce modèle.*

ABSTRACT. *We consider the problem of item recommendation with collaborative filtering systems. We propose to combine a global approach based on matrix factorization and a local approach based on the direct use of a neighborhood of the user. The hypothesis explored in this paper is that the positive and negative user judgements do have different semantics and therefore different utilities for the recommandation task. We propose a model that exploits this polarity and learns to weight the influence of a user neighbors based on polarity characteristics. Experiments performed on two corpus allow us to assess the validity of this model.*

MOTS-CLÉS : *système de recommandation*

KEYWORDS: *recommandation system*

CORIA 2014, pp. 11–26, Nancy, 18-21 mars 2014

## 1. Introduction

Avec l'explosion des données sur le Web, les systèmes de recommandation sont de plus en plus présents. Ils permettent de conseiller des services ou des produits qui ont le potentiel d'intéresser des utilisateurs, qu'il s'agisse de musique ou de films (services) ou de matériel informatique (produits). Il existe deux grands types de systèmes de recommandation, qui peuvent être combinés, les premiers sont des systèmes de *filtrage par contenu* et reposent sur la construction de profils de produits et d'utilisateurs, tandis que les seconds sont des systèmes de *filtrage collaboratif*, où un ensemble de jugements utilisateur est pris en compte pour effectuer une recommandation.

Dans cet article, nous nous intéressons aux méthodes de filtrage collaboratif, et plus particulièrement, nous nous appuyons sur les méthodes de factorisation matricielle où utilisateurs et produits sont représentés dans un même espace latent. Ces méthodes ont souvent montré d'excellentes qualités (Koren *et al.*, 2009) et exploitent une information globale à l'ensemble des utilisateurs. Deux utilisateurs partageant une proportion importante de jugements communs auront des représentations similaires dans l'espace de représentation latent. D'autres familles de techniques de filtrage exploitent des informations locales : la préférence d'un utilisateur sera une moyenne pondérée de celles de ses voisins qui sont eux même soit connus explicitement comme dans le cas de graphes sociaux, soit inférés au sens d'une certaine métrique ou similarité sur l'espace des utilisateurs (Herlocker *et al.*, 1999). Enfin, des méthodes combinent ces deux idées comme par exemple (Jamali et Ester, 2010 ; Delporte *et al.*, 2013 ; Yang *et al.*, 2012).

Dans cet article, nous allons nous aussi utiliser une combinaison d'information globale et locale pour une tâche de filtrage collaboratif. L'hypothèse que nous explorons est que la polarité (positive ou négative) des préférences des utilisateurs a une sémantique bien précise et que les opinions positives apportent une information plus forte et plus facilement exploitable que les opinions négatives. Les méthodes de filtrage collaboratif ne font pas cette distinction et donnent la même importance à toutes les préférences utilisateurs. Nous allons montrer dans ce travail que cette différence est fondée et proposons un modèle permettant de l'exploiter.

Nos contributions sont les suivantes :

– Nous caractérisons l'importance du phénomène de polarité des notes pour la recommandation ;

– Nous proposons un modèle permettant d'intégrer cette caractéristique en apprenant à pondérer l'importance des voisins non seulement selon leur similarité, mais en fonction de caractéristiques de la polarité des accords entre l'utilisateur et les voisins.

– Nous évaluons ce modèle sur deux corpus représentatifs : Yahoo ! KDD Cup 2011 et Epinions.

L'article est organisé de la façon suivante. Nous commençons par un état de l'art sur les systèmes de recommandation et l'utilisation de la polarité dans les systèmes sociaux dans la section 2. Nous présentons le modèle utilisé en section 3, puis des

mesures caractérisant la polarité des accords entre utilisateurs en section 4. Les expériences conduites et leurs résultats qualitatifs et quantitatifs sont détaillés dans la section 5.

## 2. État de l'art

De nombreux systèmes de recommandation ont été proposés dans la littérature et (Bobadilla *et al.*, 2013) donne un bon résumé de ce domaine. Notre travail s'appuie sur les techniques dérivées de la factorisation matricielle (Zhou *et al.*, 2008) qui ont pour caractéristique principale de représenter utilisateurs et produits dans un espace latent commun. Cette famille de méthodes a été largement utilisée lors de la KDD Cup 2011, et a permis d'obtenir d'excellents résultats en prédiction de notes (Koenigstein *et al.*, 2011 ; Dror *et al.*, 2012). Ces méthodes exploitent de manière globale l'information apportée par l'ensemble des notes disponibles. Elles permettent par ailleurs de développer une grande variété de modèles en intégrant sous forme d'ensembles de contraintes des caractéristiques additionnelles des données. Par exemple, nous trouverons dans (Chen *et al.*, 2011) 81 variantes de la factorisation matricielle appliquées aux données Yahoo !, ou encore dans (Zheng *et al.*, 2011) des modèles qui prennent en compte l'horodatage ou une taxinomie.

D'autres techniques de filtrage collaboratif reposent sur l'utilisation explicite du voisinage d'un utilisateur et/ou d'un item (Bobadilla *et al.*, 2013), ce qui permet d'exploiter directement une information locale à un utilisateur. La prise en compte d'information locale a également été introduite sous la forme de contraintes au niveau des techniques de factorisation matricielle. Nous discutons ci-dessous ces travaux avant de les présenter plus formellement dans la section suivante.

(Zheng *et al.*, 2011) utilisent un filtrage basé produit : la note prédite pour un produit dépendra des $k$ plus proches produits avec une distance qui dépend de la proportion d'utilisateurs les ayant jugés de la même manière. (Bell et Koren, 2007) utilisent une combinaison de filtrage basé produit et basé utilisateur qui considère les $k$ plus proches voisins d'un utilisateur. On parlera de voisinage *implicite* quand il est défini à partir de similarités. Dans ces approches, les similarités entre utilisateurs prennent en compte le nombre de jugements communs sans distinction de polarité.

Lorsqu'un graphe social existe, utiliser les voisinages *explicites* peut s'avérer être une excellente source d'information *indépendante* (Crandall *et al.*, 2008) pour améliorer les modèles de factorisation matricielle. Le voisinage social peut influencer les utilisateurs dans le choix des produits et dans la note qu'ils vont donner. Parmi les modèles utilisant ce voisinage explicite, nous pouvons citer (Jamali et Ester, 2010) et (Delporte *et al.*, 2013) qui utilisent la représentation latente des voisins, en plus de celle de l'utilisateur, lors de la prédiction.

La sémantique de la polarité des notes n'a pas à notre connaissance été étudiée ou exploitée explicitement dans les systèmes de recommandation. Par contre, quelques travaux se sont penchés sur l'étude de graphes sociaux signés pour différentes tâches

comme le scoring de nœuds, le clustering ou la prédiction de liens signés. Par exemple (Yang *et al.*, 2011) et (Yang *et al.*, 2012) proposent des modèles pour prédire la polarité des liens sociaux (confiance et méfiance) et analysent l'impact du signe du lien social sur la similarité des notes entre voisins. Les auteurs montrent alors que les liens sociaux négatifs (méfiance) ne sont pas corrélés aux jugements des utilisateurs alors que les liens positifs le sont. Ils proposent pour la prédiction de liens signés un modèle de factorisation matricielle où la relation de voisinage (notamment le poids de chaque relation) est prise en compte (Leskovec *et al.*, 2010 ; Kunegis *et al.*, 2009).

Ces travaux ont donc permis d'apporter une information supplémentaire aux techniques de factorisation matricielle et, dans certains cas (Yang *et al.*, 2012), ont pris en compte la polarité des liens sociaux pour affiner la prédiction de liens. En revanche, cette polarité n'est pas étudiée dans le cas des jugements, mais uniquement dans le cas de liens sociaux.

## 3. Représentation latente et voisinage

Dans cette section, nous décrivons tout d'abord le modèle standard de factorisation matricielle pour le filtrage collaboratif, avant de présenter ceux qui intègrent le voisinage.

Les principaux modèles utilisés dans (Chen *et al.*, 2011), (Koren et Bell, 2011) ou encore (Yang *et al.*, 2012) s'appuient sur l'utilisation d'un espace latent $\mathbb{R}^n$ dans lequel sont représentés l'ensemble des utilisateurs et des produits. La note $\hat{r}_{ui}$ de l'utilisateur $u \in U$ pour le produit $i \in I$ est alors estimée à partir des deux vecteurs représentatifs de l'espace latent pour l'utilisateur ($p_u$) et le produit ($q_i$) :

$$\hat{r}_{ui} = \mu + \mu_i + \mu_u + {q_i}^T \cdot p_u \quad [1]$$

où $\mu$ est la moyenne des notes du corpus, $\mu_i$ la moyenne des notes du produit $i$ et $\mu_u$ des notes de l'utilisateur $u$. Ces biais permettent de prendre en compte les différences systématiques des notes pour certains utilisateurs et produits.

L'apprentissage de ce type de modèle se fait en général en minimisant l'erreur $e_{ui} = r_{ui} - \hat{r}_{ui}$ pour les notes $r_{ui}$ connues. Enfin, pour éviter tout phénomène de sur-apprentissage, on ajoute à la fonction objectif un terme de régularisation en norme L2, ce qui donne :

$$E = min \sum_{(u,i)} (r_{ui} - \hat{r}_{ui})^2 + \lambda(\|p_u\|^2 + \|q_i\|^2) \quad [2]$$

La somme étant prise sur toutes les notes connues.

Dans cette représentation, il est facile de voir que plus deux utilisateurs donnent les mêmes notes, plus ils auront des représentations proches et que cette proximité dépend uniquement de la similarité des jugements entre utilisateurs.

La prise en compte d'informations de voisinage peut être exprimée en introduisant les voisins dans la formule (1) :

$$\hat{r}_{ui} = \mu + \mu_i + \mu_u + \sum_{v \in V_u \cup \{u\}} w_{uv} p_v^T q_i \qquad [3]$$

où $V_u$ est l'ensemble des voisins de $u$ et $w_{uv}$ une similarité entre $u$ et $v$.

C'est ce type de formulation qu'exploitent par exemple (Yang *et al.*, 2012 ; Bell et Koren, 2007). Les approches diffèrent dans l'expression de $w_{uv}$.

Yang et al. (Yang *et al.*, 2012) supposent qu'un graphe social décrivant les relations entre les utilisateurs est disponible et supposent que le poids $w_{uv}$ est directement proportionnel à la similarité entre les deux utilisateurs dans l'espace latent ($p$ est le poids donné à la représentation de l'utilisateur cible) :

$$w_{uv} = \begin{cases} p & \text{si } v = u \\ (1-p) p_u \cdot p_v & \text{sinon} \end{cases}$$

Bell et al. (Bell et Koren, 2007) ne font pas l'hypothèse d'un graphe social, mais utilisent un voisinage implicite. Les voisins $V$ d'un utilisateur sont ses $k$ plus proches voisins calculés dans l'espace des notes obtenus grâce à une fonction de similarité prenant en compte l'historique des notes ; $w_{uv}$ sera proportionnel à la similarité entre utilisateurs.

Dans cet article, nous utiliserons également la formulation générale (3). La nouveauté est que le poids $w_{uv}$ n'est pas donné a priori par une fonction de similarité basée sur la représentation des nœuds, mais est appris et que pour cela, nous exploitons des caractéristiques qui utilisent la polarité des jugements communs à deux utilisateurs.

## 4. Modèle

Nous allons maintenant décrire plus précisément les hypothèses qui sous-tendent notre travail, avant de présenter la forme fonctionnelle qui nous permet de calculer le poids $w_{uv}$ entre deux utilisateurs ainsi que les caractéristiques dont nous nous servons pour estimer ce lien.

### 4.1. *Relations entre utilisateurs*

Nous nous plaçons dans le contexte du filtrage collaboratif et supposons qu'il existe un ensemble de jugements utilisateurs. Nous supposons, de plus, que ceux-ci sont binaires (positifs/négatifs) ou que l'on peut se ramener à ce cas quand on dispose de jugements valués. C'est ce que nous avons fait dans les expériences.

Notre hypothèse de départ concernant la sémantique des liens est que les jugements positifs n'ont pas le même sens que les jugements négatifs. Nous supposons

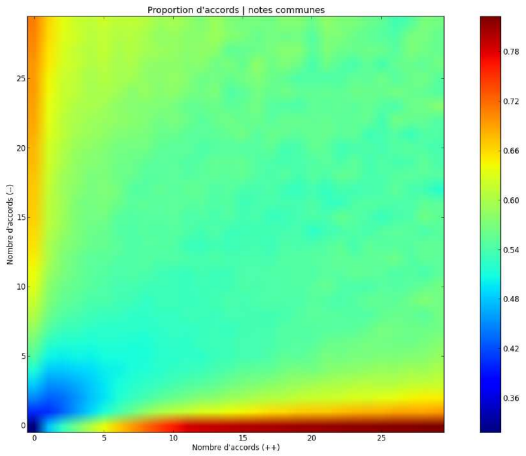


**Figure 1.** *Le pourcentage d'accords pour un couple d'utilisateurs donné, en fonction du nombre d'avis positifs et négatifs en communs.*

que l'utilisateur émet des jugements sur les produits susceptibles de l'intéresser. Cette hypothèse communément admise est d'ailleurs source de biais dans l'évaluation des systèmes de recommandation (Pradel et Paris, n.d.). Deux utilisateurs partageant des notes auront donc des communautés d'intérêts. Deux utilisateurs qui ne partageraient que des avis positifs auraient des goûts communs et on peut penser que la prédiction serait aisée à condition de disposer de suffisamment de notes. Par contre, s'ils ne partagent que des avis négatifs, l'information disponible pour recommander des produits est bien plus faible. C'est ce qu'indique la figure 1 qui donne la proportion d'accords utilisateurs (nombre de notes identiques / nombre de notes communes) en fonction du nombre d'accords positifs et du nombre d'accords négatifs. Ces statistiques ont été calculées sur le corpus Yahoo ! KDD utilisé pour les expériences. Sur cette figure on distingue 3 zones, une zone (bleue) de faible accord entre utilisateurs quand il y a peu d'avis communs, une grande zone verte quand ils partagent de nombreux avis communs et deux zones jaune/ rouge quand ils partagent de façon exclusive soit des avis négatifs soit des avis positifs. Ce sont ces dernières zones qui sont les plus facilement exploitables pour des systèmes de recommandation utilisant l'information locale. À noter que la proportion d'accords est bien plus forte quand ils partagent des accords positifs. Notons toutefois que cette figure ne donne que des informations partielles, car le nombre de désaccords croit avec le nombre de notes données.

### 4.2. *Modèles*

Le modèle que nous proposons s'appuie sur l'équation (3) dont les coefficients $w_{uv}$ seront appris de façon à prendre en compte la polarité des jugements. Les poids

sont exprimés sous la forme d'une fonction logistique qui dépend d'un ensemble de variables $\Phi$ qui caractérisent cette polarité :

$$w_{uv} = \sigma(\tau \cdot \Phi) \quad [4]$$

où $\sigma(x) = (1+e^{-x})^{-1}$ désigne la fonction logistique, $\tau$ est un vecteur de pondérations qui sera appris et $\Phi = (1, \phi_1, \ldots, \phi_n)^T$ est l'ensemble des caractéristiques, la valeur «1» correspond à un terme de biais.

Nous noterons $J_u^+$ l'ensemble des items ayant reçu une note positive de l'utilisateur $u$, $J_u^-$ l'ensemble des items avec des notes négatives de $u$ et enfin $J_u$ l'ensemble de tous les produits notés par $u$. Pour deux utilisateurs $u$ et $v$, nous notons $D_{uv}$ l'ensemble des items sur lesquels ils sont en désaccord et $A_{uv}$ l'ensemble des items sur lesquels ils sont en accord. Enfin, nous distinguons l'accord positif (ils aiment le même produit) de l'accord négatif (ils n'aiment pas le même produit). Ainsi, nous avons $A_{uv} = A_{uv}^+ \cup A_{uv}^-$.

Nous pouvons alors définir les quatre caractéristiques suivantes :

**Taux d'accord positif sur l'ensemble des accords** Cette première caractéristique permet de caractériser parmi les utilisateurs similaires (avec des jugements en commun similaires) ceux qui ont une proportion importante d'accords positifs. Elle favorisera les utilisateurs qui ont principalement aimé les mêmes produits. Formellement,

$$\phi_1(u,v) = \frac{|A_{uv}^+|}{|A_{uv}|} = \frac{|J_u^+ \cap J_v^+|}{|(J_u^+ \cap J_v^+) \cup (J_u^- \cap J_v^-)|} \quad [5]$$

Notons que cette mesure ne tient compte que des accords entre utilisateurs.

**Taux d'accord positif sur l'ensemble des notes communes** Cette caractéristique est similaire à la première à la différence qu'elle prend également en compte les désaccords. Ainsi, elle rapprochera les personnes qui sont peu souvent en désaccord et aiment les mêmes produits.

$$\phi_2(u,v) = \frac{|A_{uv}^+|}{|A_{uv} \cup D_{uv}|} = \frac{|J_u^+ \cap J_v^+|}{|J_u \cap J_v|} \quad [6]$$

**Taux d'accord négatif sur l'ensemble des notes communes** Cette caractéristique est équivalente à la précédente, sauf que nous considérons le cas des accords négatifs.

$$\phi_3(u,v) = \frac{|A_{uv}^-|}{|A_{uv} \cup D_{uv}|} = \frac{|J_u^- \cap J_v^-|}{|J_u \cap J_v|} \quad [7]$$

Notons que ces trois caractéristiques ne prennent pas en compte le nombre total de produits que les utilisateurs ont notés, comparativement au nombre d'items notés en commun. Par exemple deux utilisateurs partageant dix accords positifs (et rien d'autre) seront considérés similaires par la mesure $\phi_2(u,v)$, même si ces dix accords ne représentent qu'une infime partie des items qu'ils ont notés. Pour cela nous introduisons la caractéristique suivante.

**Indice de Jaccard «positif».** Cette caractéristique, combinée aux trois premières, permet de prendre en compte le nombre total de produits notés par les deux utilisateurs et ainsi d'évaluer l'importance des notes partagées. Il permet ainsi de nuancer les cas où deux utilisateurs seraient fortement en accord positif, mais que l'ensemble des notes communes aux deux utilisateurs ne serait pas significatif. L'indice de Jaccard correspond au rapport entre les notes communes et l'ensemble des notes données par les utilisateurs du couple. Nous utiliserons ici une version «positive» de cet indice calculée comme le rapport entre les accords positifs communs et l'ensemble des notes positives des deux utilisateurs.

$$\phi_4(u,v) = \frac{|A^+_{uv}|}{|J^+_u \cup J^+_v|} = \frac{|J^+_u \cap J^+_v|}{|J^+_u \cup J^+_v|} \quad [8]$$

Ces quatre caractéristiques ont été sélectionnées parmi un ensemble de caractéristiques candidates que nous avons définies et testées.

## 5. Expériences

Dans cette section, nous présentons les données que nous utilisons pour l'ensemble des expériences. Nous définissons les différents protocoles utilisés pour nos modèles et discutons de l'étude empirique effectuée pour mesurer le potentiel de notre approche.

### 5.1. *Les données*

Les expériences ont été réalisées sur deux jeux de données.

#### 5.1.1. *Yahoo ! KDD Cup 2011*

Ce corpus dispose d'une grande quantité de notes. Nous disposons d'une collecte de 10 ans sur le site Yahoo ! Music, ce qui représente 262.810.175 notes et 1.000.090 utilisateurs. Construit pour la KDD Cup 2011, le corpus propose trois ensembles de notes pour l'apprentissage, la validation et le test des modèles. Les notes se situent entre 0 et 100. La moyenne de l'ensemble des notes est aux alentours de 48 sur 100. Nous les avons binarisées de la façon suivante : celles inférieures ou égales à 30 sur 100 sont considérées comme négatives, celles supérieures ou égales à 70 sont considérées positives. Les autres notes sont ignorées. Deux utilisateurs seront en accord si leur note est du même signe.

#### 5.1.2. *Epinions*

Il est composé de 132.000 utilisateurs et contient 1.560.144 articles et 13.668.319 notes. Les notes se situent entre 0 et 5 et la moyenne est aux alentours de 4. Une note inférieure (strictement) à la moyenne est considérée comme négative, une note supérieure (strictement) à la moyenne sera positive.

Ce corpus n'est pas décomposé en ensembles d'apprentissage et de test. Pour les évaluations, nous avons utilisé une cross-validation (70% apprentissage, 30% test).

### 5.2. *Algorithme d'apprentissage*

Les modèles présentés en section 4.2 sont appris grâce à une descente de gradient stochastique. Cette approche est très utilisée dans la littérature, notamment dans (Chen *et al.*, 2011 ; Koenigstein *et al.*, 2011 ; Wu *et al.*, 2011), car elle allie une facilité d'implémentation, une rapidité d'exécution et de bons résultats. Nous avons utilisé une descente alternée, c'est-à-dire en modifiant alternativement un ensemble de paramètres, les autres étant fixés. Nous parcourons donc chaque exemple d'apprentissage et mettons à jour successivement les vecteurs latents représentant les utilisateurs et les produits, puis, les paramètres de notre fonction logistique.

La fonction de coût 2 est utilisée avec le modèle de recommandation suivant :

$$\hat{r}_{ui} = \mu + \mu_i + \mu_u + \left( (1 - \frac{\sum_{v \in V_u} w_{uv}}{2k}) p_u + \frac{\sum_{v \in V_u} w_{uv}}{2k} p_v \right)^T \cdot q_i \qquad [9]$$

En notant $e_{ui} = r_{ui} - \hat{r}_{ui}$, cela nous donne pour chaque mise à jour :

$$p_u \leftarrow p_u + 2\gamma \left( e_{ui} \left( 1 - \frac{\sum_{v \in V} w_{uv}}{2k} \right) q_i - \lambda p_u \right) \qquad [10]$$

$$q_i \leftarrow q_i + 2\gamma \left( e_{ui} \left( (1 - \frac{\sum_{v \in V} w_{uv}}{2k}) p_u + \frac{\sum_{v \in V} w_{uv}}{2k} p_v \right) - \lambda q_i \right) \qquad [11]$$

$$\tau \leftarrow \tau + \gamma_\tau \left( e_{ui} \frac{\sum_{v \in V} \phi_{uv} \sigma(\tau \cdot \phi_{uv})(1 - \sigma(\tau \cdot \phi_{uv}))}{2k} (p_v - p_u)^T q_i \right) \qquad [12]$$

où $\gamma$ et $\gamma_\tau$ sont les pas de gradient[1], et $\lambda$ et $\lambda_\tau$ sont les coefficients de régularisation correspondants. Notons que pour les modèles sans prise en compte du voisinage, $w_{uv}$ est égal à zéro sauf lorsque $u = v$, et la mise à jour (12) n'est pas utilisée.

Lors de cette descente de gradient, les paramètres utilisés pour les produits et les items sur l'ensemble d'apprentissage ont été : $\gamma = 3 \times 10^{-4}$ et $\lambda = 1 \times 10^{-2}$ . Pour les poids, les paramètres utilisés sur l'ensemble de validation ont été : $\gamma_\tau = 1 \times 10^{-7}$ et $\lambda_\tau = 1 \times 10^{-2}$.

1. nous avons un pas différent pour les poids des représentations latentes et pour les paramètres de la fonction logistique

### 5.3. *Calcul du graphe de similarité*

Compte tenu de la quantité de données et de la taille de la matrice utilisateur×utilisateur utilisée pour calculer les voisinages, nous avons utilisé des heuristiques pour limiter le nombre de voisins considérés. Pour un utilisateur donné, nous gardons (au maximum) les 30 utilisateurs les plus en accord avec lui (en proportion des notes communes). Ensuite, nous ajoutons à notre présélection la même quantité d'utilisateurs choisis aléatoirement (dans l'ensemble des personnes ayant un produit jugé en commun) pour lesquels nous nous attendons à ce que le $w$ soit proche de zéro.

Cette sélection est utilisée pour effectuer une initialisation du modèle, elle nous permet d'effectuer une première phase d'apprentissage dans le but de calculer les coefficients de notre régression logistique. Une fois cette première phase effectuée, nous pouvons relancer une phase de sélection des voisins au cours de laquelle nous sélectionnons les 50 utilisateurs dont le $w$ associé est le plus élevé.

## 6. Résultats

### 6.1. *Évaluation qualitative*

Dans un premier temps, nous voulons étudier la pertinence des caractéristiques proposées. Pour cela, nous introduisons une mesure de similarité entre utilisateurs qui est basée sur leur voisinage commun. Idéalement, cette mesure devrait nous indiquer si les deux utilisateurs font partie d'un groupe cohérent d'utilisateurs qui ont une opinion positive des mêmes produits.

Dans la pratique, nous procédons de la façon suivante. Pour un couple d'utilisateurs, nous considérons comme voisinage commun l'ensemble des utilisateurs qui ont noté au moins un produit en commun avec chacun des deux utilisateurs. La mesure de similarité basée sur le voisinage du couple est alors la proportion de ces voisins qui ont aimé au moins un même produit (pas nécessairement le même) avec l'un comme avec l'autre. Lorsque cette mesure est nulle, cela signifie qu'il n'y a aucun voisin qui a une opinion positive commune avec chaque utilisateur du couple considéré. Inversement, une mesure à 1 voudra dire que n'importe quel voisin du couple partage un avis positif — ce qui correspond d'une certaine façon à cette notion de groupe cohérent.

Nous avons sélectionné 25 000 000 couples d'utilisateurs sur un échantillon du corpus Yahoo ! pour lesquels nous avons calculé l'ensemble des caractéristiques décrites en section 4.2. Nous avons calculé la similarité de voisinage en fonction des caractéristiques proposées, nous décrivons ici les observations correspondantes pour deux d'entre elles : le taux d'accord positif sur l'ensemble des notes communes (équation 6) et l'indice de Jaccard. La figure 2 donne en abscisse les quantiles d'une caractéristique (on considère 25 quantiles) et en ordonnée la similarité de voisinage. Elle montre que la première caractéristique est plus fortement corrélée à la similarité de voisinage que l'indice de Jaccard qui ne prend pas en compte la polarité des jugements. Dans le cas de la première caractéristique, les couples au $\phi$ faible partagent beaucoup moins

de voisins avec qui ils partagent des jugements positifs que ceux au $\phi$ élevé (plus de 80%). Avec l'indice de Jaccard, la moyenne pour l'ensemble des quantiles ne varie pas énormément (entre 80% et 90%), ce qui est beaucoup moins informatif.

Nous pouvons observer une corrélation similaire entre les autres caractéristiques introduites en section 4.2 et la qualité du voisinage. Pour la caractéristique prenant en compte la quantité de jugements négatifs, la progression est inverse. Pour la caractéristique normalisée sur l'ensemble des accords, la corrélation est moins nette et moins marquée que celle observable sur 2.

Sur la figure 3, nous avons reproduit le même type de graphique, mais avec en abscisse la valeur des poids $w$ appris. Les valeurs élevées de $w$ correspondant bien à une similarité de voisinage élevée, avec le même effet de seuil déjà observé pour la caractéristique $\phi_2(u, v)$ sur la figure 2. Notre modèle semble donc avoir appris des coefficients pertinents et la nouvelle mesure mise en place semble donc discriminante.

La table (1) donne les coefficients obtenus pour le vecteur de la régression logistique (colonne 2). Les coefficients mesurant les accords positifs sont supérieurs à zéro et tendent à faire augmenter la valeur de $w$, le coefficient correspondant à $\phi_2$ étant de loin le plus élevé. Le coefficient correspondant aux accords négatifs tend à abaisser la valeur de $w$. Ainsi deux utilisateurs partageant uniquement des accords négatifs auraient un $w$ proche de 0, alors que deux utilisateurs partageant principalement des accords positifs auraient un $w$ proche de 1. Cette observation tend à valider l'hypothèse de différentiation des jugements positifs et négatifs pour la recommandation. La principale information apportée par ces résultats est que le taux d'accord négatif sur l'ensemble des accords a un impact négatif sur la mesure de similarité. Autrement dit, posséder des accords négatifs ne joue pas en faveur d'une plus grande similarité. Deux personnes ayant un taux élevé d'accords négatifs seront donc moins rapprochées que deux utilisateurs ayant uniquement des accords positifs.

Les colonnes suivantes de la table (1) donnent des exemples de vecteurs de caractéristiques calculés sur des couples d'utilisateurs. Les utilisateurs $u$ et $v_1$ sont principalement en désaccord et $w_{uv_1}$ est nul. Pour les deux autres couples $(u, v_2)$ et $(u, v_3)$, les utilisateurs sont principalement en accord sur les jugements qu'ils ont donnés. La différence se situe alors sur la proportion de jugements communs négatifs. Les utilisateurs $u$ et $v_3$ partageant plus de produits qu'ils ont aimés, ils seront davantage similaires que $u$ et $v_2$. Ainsi, $w_{uv_2}$ est inférieur à $w_{uv_3}$.

### 6.2. *Évaluation quantitative*

Nous avons également évalué notre modèle dans le cadre de la prédiction de note et de l'ordonnancement afin de mesurer l'amélioration que peut apporter la prise en compte de la polarité.

Pour évaluer nos modèles en prédiction de notes, nous avons utilisé une mesure standard du domaine, la moyenne des erreurs quadratiques (RMSE) entre la note pré-

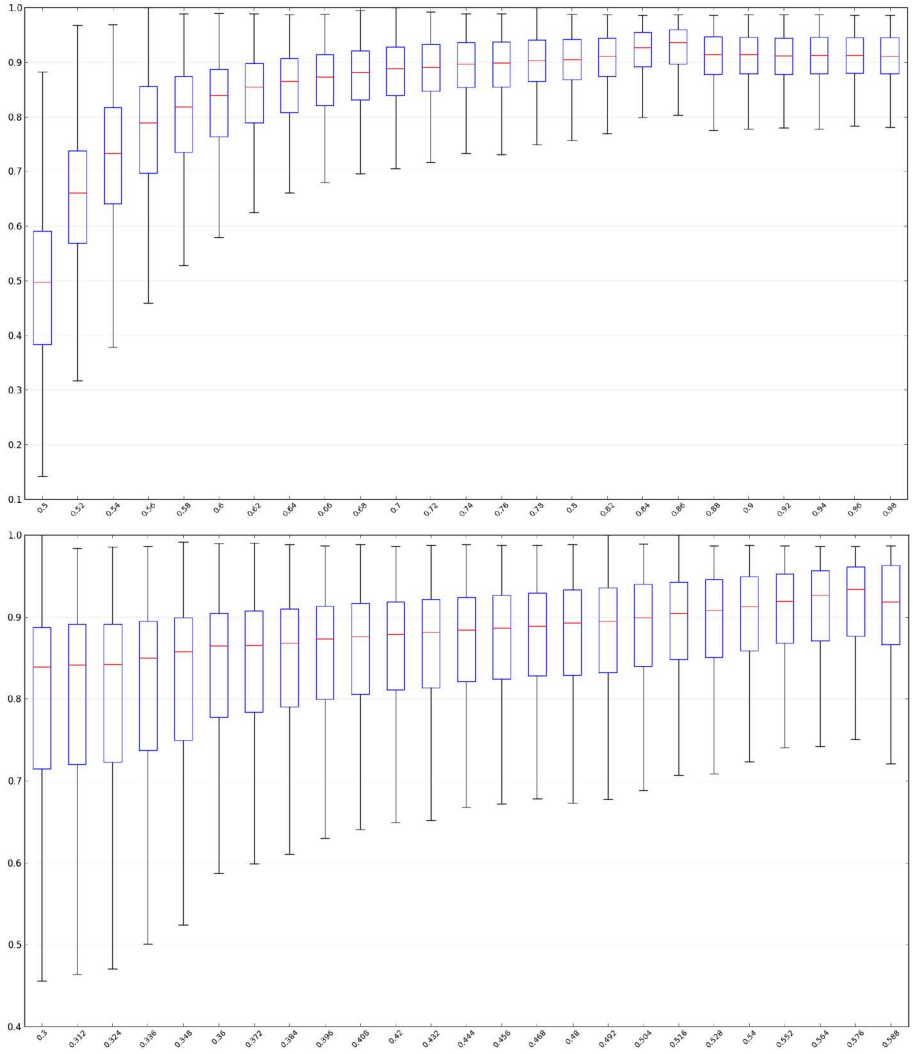

**Figure 2.** *La qualité du voisinage en fonction du taux d'accord positif sur l'ensemble des notes communes (en haut) et de l'indice de Jaccard (en bas). Chaque quantile représente 1 000 000 de couples d'utilisateurs.*

dite et la note réelle. Pour l'évaluation en ordonnancement, nous utilisons la mesure de précision moyenne MAP. Elle correspond à la moyenne sur l'ensemble des utilisateurs de la précision moyenne de chaque produit de l'ensemble de test. Le modèle de référence est une factorisation matricielle correspondant aux équations (1) et (2). Les résultats sont donnés dans le tableau 2. Comme nous pouvons l'observer, le modèle avec prise en compte du voisinage local et apprentissage des coefficients permet d'améliorer les résultats en ordonnancement et en prédiction de notes pour le corpus

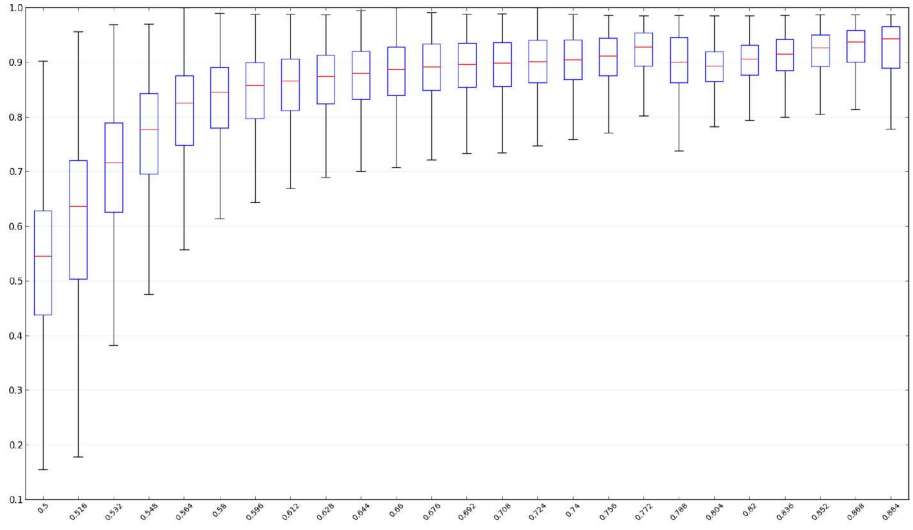

**Figure 3.** *La similarité de voisinage selon la valeur de notre régression logistique calculée lors de l'apprentissage*

| | $\tau$ | $\Phi_{uv_1}$ | $\Phi_{uv_2}$ | $\Phi_{uv_3}$ |
|---|---|---|---|---|
| $\phi_1 : \lvert A^+_{uv}\rvert / \lvert A_{uv}\rvert$ | 0.11 | 0.4 | 0.25 | 1.0 |
| $\phi_2 : \lvert A^+_{uv}\rvert / \lvert A_{uv} \cup D_{uv}\rvert$ | 6.74 | 0.11 | 0.75 | 0.75 |
| $\phi_3 : \lvert A^-_{uv}\rvert / \lvert A_{uv} \cup D_{uv}\rvert$ | -1.65 | 0.17 | 0.24 | 0.00 |
| $\phi_4 : \lvert A^+_{uv}\rvert / \lvert J^+_u \cup J^+_v\rvert$ | 0.55 | 0.12 | 0.08 | 0.19 |
| 1 : Biais | -5.20 | 1 | 1 | 1 |
| $w_{uv_i}$ | | 0.0 | 0.38 | 0.52 |

**Tableau 1.** *Le vecteur $\tau$ des coefficients de la régression logistique. Exemples de poids $w$ obtenus pour trois couples $(uv_1, uv_2, uv_3$ de plus en plus similaires)*

Yahoo ! et ceci, quels que soient le nombre de notes connues et la dimension $D$ de l'espace latent. En effet, nous observons pour la MAP un gain de l'ordre de 3 à 8% et pour le RMSE, une amélioration de l'ordre de 1 à 2%. Notre mesure de similarité apporte donc une information pertinente pour notre apprentissage.

Pour le corpus Epinions, nous obtenons un gain en ordonnancement, alors que nous perdons en performance dans le cadre de la prédiction de notes. Le voisinage construit ne semble donc pas suffisamment informatif dans le cadre de la prédiction de notes, mais permet d'améliorer l'ordonnancement qui profite de l'ordre établi par les produits les plus notés par les voisins. La différence entre les deux corpus est que le nombre de notes est bien plus faible dans Epinions, et que l'utilisation de voisinages locaux réduit encore le nombre de données utilisées pour apprendre le modèle.

| | Epinions | | | |
|---|---|---|---|---|
| | D = 25 | | D = 50 | |
| | RMSE | MAP | RMSE | MAP |
| Modèle standard | 0.553 | 0.128 | 0.675 | 0.125 |
| Modèle avec Voisinage | 0.608 | 0.129 | 0.798 | 0.132 |
| | Yahoo ! — 10% des notes | | | |
| | D = 25 | | D = 50 | |
| | RMSE | MAP | RMSE | MAP |
| Modèle standard | 28.27 | 0.565 | 28.10 | 0.491 |
| Modèle avec Voisinage | 27.60 | 0.589 | 27.55 | 0.535 |
| | Yahoo ! — 100% des notes | | | |
| | D = 25 | | D = 50 | |
| | RMSE | MAP | RMSE | MAP |
| Modèle standard | 26.02 | 0.556 | 26.18 | 0.505 |
| Modèle avec Voisinage | 25.78 | 0.583 | 25.88 | 0.521 |

**Tableau 2.** *Résultats RMSE et MAP sur un sous-corpus de Yahoo ! comptant 150 000 utilisateurs*

## 7. Conclusion

Dans cet article, nous nous sommes intéressés à l'utilisation de la polarité des jugements dans une tâche de filtrage collaboratif. Partant de l'hypothèse que les jugements positifs et négatifs avaient des sémantiques différentes, nous avons proposé un modèle utilisant des informations locales de voisinage utilisateur, qui exploite cette dissymétrie entre les deux polarités. En se basant sur des caractéristiques mesurant cette polarité, le modèle apprend à partir d'exemples à pondérer l'influence de ses voisins.

Des expériences réalisées sur deux jeux de données ont montré l'intérêt de l'apprentissage de la prise en compte de la polarité des opinions. Sur les deux jeux de données utilisées, cette information permet d'augmenter la mesure MAP qui est représentative du but de la recommandation. Sur la mesure RMSE qui montre la qualité de l'approximation, les performances sont améliorées sur l'un des jeux de données et détériorées sur l'autre qui est caractérisé par un nombre de jugements beaucoup plus faible.

Finalement, notre travail ouvre des pistes quant à la prise en compte du voisinage dans les modèles de filtrage collaboratif basés sur un espace latent, en apprenant dans quelle mesure deux voisins sont similaires. Des extensions sont possibles, en particulier en prenant en compte le graphe social (quand celui-ci existe) et la polarité des relations entre utilisateurs.

## 8. Bibliographie

Bell R. M., Koren Y., « Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights », *IEEE International Conference on Data Mining (ICDM*, 2007.

Bobadilla J., Ortega F., Hernando A., Gutiérrez A., « Recommender systems survey », *Knowledge-Based Systems*, 2013.

Chen P.-l., Tsai C.-t., Chen Y.-n., Chou K.-c., Li C.-l., Tsai C.-h., Wu K.-w., Chou Y.-c., Li C.-y., Lin W.-s., Yu S.-h., Chiu R.-b., Lin C.-y., Wang C.-c., Wang P.-w., Su W.-l., Wu C.-h., Kuo T.-t., Mckenzie T. G., Chang Y.-h., Ferng C.-s., Ni C.-m., Lin H.-t., Lin C.-j., Lin S.-d., « A Linear Ensemble of Individual and Blended Models for Music Rating Prediction », 2011.

Crandall D., Cosley D., Huttenlocher D., Kleinberg J., Suri S., « Feedback Effects Between Similarity and Social Influence in Online Communities », *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, ACM, New York, NY, USA, p. 160-168, 2008.

Delporte J., Karatzoglou A., Matuszczyk T., Canu S., « Socially Enabled Preference Learning from Implicit Feedback Data », *in* H. Blockeel, K. Kersting, S. Nijssen, F. ÅœeleznÃœ (eds), *Machine Learning and Knowledge Discovery in Databases*, vol. 8189 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, p. 145-160, 2013.

Dror G., Koenigstein N., Koren Y., Weimer M., « The Yahoo ! Music Dataset and KDD-Cup'11 », *JMLR Workshop and Conference Proceedings*, vol. 18, p. 3-18, 2012.

Herlocker J. L., Konstan J. A., Borchers A., Riedl J., « An Algorithmic Framework for Performing Collaborative Filtering », *Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, ACM, New York, NY, USA, p. 230-237, 1999.

Jamali M., Ester M., « A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks », *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, ACM, New York, NY, USA, p. 135-142, 2010.

Koenigstein N., Dror G., Koren Y., « Yahoo ! Music Recommendations : Modeling Music Ratings with Temporal Dynamics and Item Taxonomy », *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, ACM, New York, NY, USA, p. 165-172, 2011.

Koren Y., Bell R., « Advances in Collaborative Filtering », *in* F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (eds), *Recommender Systems Handbook*, Springer US, p. 145-186, 2011.

Koren Y., Bell R., Volinsky C., « Matrix Factorization Techniques for Recommender Systems », *Computer*, vol. 42, n° 8, p. 30-37, August, 2009.

Kunegis J., Lommatzsch A., Bauckhage C., « The Slashdot Zoo : Mining a Social Network with Negative Edges », *Proceedings of the 18th international conference on World wide web*, 2009.

Leskovec J., Huttenlocher D., Kleinberg J., « Signed networks in social media », ACM Press, New York, New York, USA, p. 1361, 2010.

Pradel B., Paris L. I. P. U., « Ranking With Non-Random Missing Ratings : Influence Of Popularity And Positivity on Evaluation Metrics », p. 147-154, n.d.

Wu Y., Yan Q., Bickson D., Low Y., Yang Q., « Efficient Multicore Collaborative Filtering », *ACM KDD CUP workshop*, 2011.

Yang S. H., Long B., Smola A., Sadagopan N., Zheng Z., Zha H., « Like like alike Joint Friendship and Interest Propagation in Social Networks », 2011.

Yang S.-H., Smola A. J., Long B., Zha H., Chang Y., « Friend or Frenemy ? : Predicting Signed Ties in Social Networks », *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, ACM, New York, NY, USA, p. 555-564, 2012.

Zheng Z., Chen T., Liu N., Yang Q., Yu Y., « Rating Prediction with Informative Ensemble of Multi-Resolution Dynamic Models », p. 75-97, 2011.

Zhou Y., Wilkinson D., Schreiber R., Pan R., « Large-Scale Parallel Collaborative Filtering for the Netflix Prize », *Proc. 4th Int'l Conf. Algorithmic Aspects in Information and Management, LNCS 5034*, Springer, p. 337-348, 2008.