
Factorisation en matrices non négatives pour le filtrage collaboratif

Jean-François Pessiot, Vinh Truong, Nicolas Usunier, Massih-Reza Amini et Patrick Gallinari

Laboratoire d'Informatique de Paris VI
8 rue du Capitaine Scott, 75015 Paris

{pessiot, truong, usunier, amini, gallinari}@poleia.lip6.fr

RÉSUMÉ. Les systèmes de filtrage collaboratif (FC) aident un utilisateur à trouver l'information qui l'intéresse à partir des jugements d'autres utilisateurs. Ces systèmes sont utilisés principalement par des sites Web pour faire des recommandations. En général, les systèmes de FC demandent aux utilisateurs de fournir des jugements sur des articles de leur choix, exprimés sous la forme d'une note, un réel positif borné. L'objectif d'un système est alors de comparer les goûts des utilisateurs afin de prédire les notes qui n'ont pas été fournies au système. Nous proposons un nouvel algorithme pour effectuer ces prédictions, basé sur la méthode de réduction dimensionnelle appelée factorisation en matrices non négatives, connue en analyse d'images et de textes. Nous testons l'approche sur une tâche de recommandation de films, et montrons qu'elle possède des performances comparables à l'état de l'art. De plus, la méthode présente les avantages d'être très simple à implanter, de faible complexité, et permet d'identifier des groupes de films dont les notes sont corrélées entre les utilisateurs.

ABSTRACT. Collaborative filtering (CF) systems help a user find the information satisfying his needs by using other users' judgements. These systems are mainly used in Web sites to provide recommendations. In general, CF systems ask users to provide a judgement to a number of items of their choice, expressed as a rating. The goal of the system is then to compare different user's tastes to predict the unknown ratings. We propose a new algorithm for CF based on a dimensionality reduction method called non-negative matrix factorization, well-known in the domains of image recognition or text processing. We test our approach on a movie recommendation task, and we show that it has a performance comparable with the state-of-the-art. Moreover, the method is simple to implement, has low complexity and enables to identify groups of films which have highly correlated ratings among the users.

MOTS-CLÉS : filtrage collaboratif, réduction dimensionnelle, apprentissage

KEYWORDS: collaborative filtering, dimensionality reduction, machine learning

1. Introduction

Les systèmes de filtrage collaboratif (FC) aident un utilisateur à trouver l'information qui l'intéresse à partir des jugements d'autres utilisateurs. Cette approche a pour but d'automatiser les recommandations que peuvent se faire des personnes partageant les mêmes centres d'intérêts. C'est une réponse efficace au problème de la sélection d'objets comme les produits culturels, qui possèdent des descripteurs naturels (auteur d'un livre, réalisateur d'un film) insuffisants pour prédire les goûts d'un individu. Le nombre grandissant de systèmes de recommandation sur Internet, par exemple pour les films (MovieLens¹, ymdb.com, ...), les pages Web (Del.icio.us), ou encore sur des sites commerciaux comme Amazon.com ou CDNow.com, montre que le développement de systèmes de FC performants présente des enjeux économiques importants.

La plupart des algorithmes de FC existants considèrent le cas où les jugements sont exprimés sous forme de notes [HER 04], et la taille trop importante des bases de jugements disponibles a amené les concepteurs à se focaliser sur des méthodes d'apprentissage pour prédire les notes inconnues [MAR 04a]. La diversité des applications du FC a alors amené à la création de plusieurs types d'approches, répondant à des spécifications différentes. Les premières approches, appelées à base d'instances (ou *memory-based*), prédisent les notes inconnues pour un utilisateur donné en combinant les notes des utilisateurs ayant les goûts les plus similaires. Elles sont très simples à implanter, mais le temps nécessaire à présenter une recommandation dépend du nombre d'utilisateurs de la base.

Pour pouvoir effectuer des recommandations en temps réel, un premier type d'approches est de calculer toutes les notes inconnues durant l'apprentissage. Ces systèmes utilisent des adaptations de méthodes de réduction dimensionnelle linéaire à la matrice (utilisateurs, articles) [SAR 00]. Ils sont faciles à implanter, mais ont une complexité d'apprentissage ne leur permettant pas d'être appliqués à des bases contenant des millions de jugements. Des approches de plus faible complexité, utilisant des modèles probabilistes des notes des utilisateurs, ont alors été proposés [HOF 04, MAR 04b]. Elles sont beaucoup plus sophistiquées, et certaines [MAR 04b] nécessitent une connaissance approfondie du domaine de l'apprentissage pour être mises en oeuvre. Elles sont néanmoins reconnues comme les plus performantes [MAR 04a], et présentent l'intérêt majeur que les modèles peuvent être interprétés comme expliquant le comportement des utilisateurs [HOF 04, POL 04]. La compréhension des comportements permet alors d'améliorer le développement du système, de l'interface utilisateur ou d'identifier des communautés d'utilisateurs ou des ensembles d'articles similaires [POL 04]. Chaque méthode réalise donc un compromis différent entre complexité, performances, facilité d'implantation et explication des prédictions. Le concepteur d'un système doit donc choisir un algorithme selon ses spécifications.

Dans cet article, nous présentons un nouvel algorithme de prédictions de notes pour le FC utilisant la méthode de réduction dimensionnelle appelée factorisation en

1. <http://movielens.umn.edu/login>

matrices non négatives (FMN) [LEE 99]. Initialement, la FMN a été introduite en traitement d'images pour la reconnaissance de visages, car elle permet d'identifier les parties des objets (yeux, bouches des visages). Appliquée au FC, elle permet d'identifier les articles pour lesquels les notes sont fortement corrélées selon les groupes d'utilisateurs, fournissant ainsi une explication des notes qu'elle prédit. De plus la méthode est de faible complexité, simple à implanter, et a des performances comparables aux méthodes de l'état de l'art. Elle présente donc à la fois les avantages des méthodes de réduction dimensionnelle existantes et des modèles probabilistes.

La suite de l'article est organisée de la façon suivante. La section 2 présente plus précisément le cadre de FC dans lequel nous nous plaçons, ainsi que l'utilisation de la réduction dimensionnelle dans un tel cadre. La section 3 présente l'algorithme principal de FMN, ainsi que les adaptations nécessaires pour la prédiction de notes dans le cas du FC. La section 4 présente les expériences qui ont été réalisées sur la base MovieLens 100K², et la section 5 décrit et discute les résultats obtenus.

2. État de l'art et motivation

2.1. Caractéristiques et spécifications des systèmes de FC

Les différents cadres développés pour le FC comprennent des approches pures ou hybrides, et elles peuvent être séquentielles ou non-séquentielles. Dans cet article, nous nous plaçons dans un cadre de FC pur, non-séquentiel, et où les jugements sont des notes.

Le terme *pur* indique que l'on n'utilise que les jugements des autres utilisateurs pour effectuer les recommandations, sans prendre en compte de descriptions des articles ; cela revient à prédire les jugements dans des cas où la seule information disponible sur chaque article est son identifiant. Ce cadre a été largement étudié (voir [MAR 04a] pour un état de l'art complet), parce qu'il montre la capacité d'un algorithme à combiner uniquement les jugements utilisateurs. Les approches utilisant à la fois les jugements utilisateurs et des descripteurs des articles sont dites *hybrides* car elles mélangent les filtrages d'information collaboratif et basé sur le contenu. Elles ont aussi été étudiées et peuvent améliorer les prédictions dans des applications [BAS 04].

L'aspect non-séquentiel des systèmes de FC correspond à l'hypothèse que l'ordre dans lequel les jugements sont collectés n'influe pas sur la recommandation. C'est le cadre le plus largement étudié, mais des applications particulières peuvent introduire l'ordre chronologique des jugements, par exemple si l'on souhaite utiliser l'historique des pages visitées durant la navigation d'un utilisateur sur un site pour guider la recommandation [SHA 05].

Finalement, nous considérons le cas où les jugements sont exprimés par des notes. Cela correspond par exemple au cas où le système de FC demande à chaque utilisateur

2. <http://www.cs.umn.edu/research/GroupLens/index.html>

de fournir des jugements sur des articles de son choix, exprimés sous la forme d'une note, un réel positif borné. C'est le cas de la majorité des systèmes, par exemple pour la recommandation de films, où les utilisateurs attribuent des notes de 1 à 5. Stocké en machine sous la forme d'une matrice de notes, ce type de jugements présente le double avantage d'être facile à collecter et à traiter du point de vue du système.

L'évaluation des systèmes de FC pur, non-séquentiel avec des jugements sous forme de notes est un problème complexe. En effet, il serait naturel d'évaluer la qualité des recommandations en simulant une situation de recommandation où seuls quelques articles seraient sélectionnés, le jugement de l'utilisateur étant ensuite comparé à ces articles. Cependant, ce type d'évaluation n'est pas possible hors-ligne, car la majorité des jugements des articles ne sont pas disponibles. Il est donc très probable que les jugements pour les quelques articles sélectionnés soient inconnus. Bien que certaines méthodes de simulation aient été proposées pour répondre à ce problème, elles sont réputées peu fiables [HER 04]. Les auteurs ont alors proposé d'évaluer simplement la qualité des notes prédites, c'est-à-dire, étant donné des articles dont les jugements sont connus mais qui n'ont pas été considérés durant l'apprentissage, comparer les notes prédites avec les notes réelles [HER 04, MAR 04a]. Bien que cette évaluation en terme de *prédiction de notes* ne corresponde pas avec l'utilisation réelle d'un système de FC, elle permet néanmoins d'avoir une mesure stable et fiable pour comparer les systèmes. C'est donc en terme de prédiction de notes que nous allons évaluer notre système, avec la mesure décrite section 4.

2.1.1. *Caractéristiques des algorithmes*

La qualité de prédiction des notes est un critère important des systèmes de FC, mais ce n'est pas le seul ; [MAR 04a] propose d'étudier les caractéristiques suivantes :

Performances en généralisation : afin de mesurer la qualité de prédiction d'un système de FC, [MAR 04a] propose deux définitions de la notion de généralisation : la généralisation *faible* mesure la qualité de prédiction des notes pour des utilisateurs qui étaient présents dans la base d'apprentissage ; les systèmes de FC ayant pour but de considérer des communautés d'utilisateurs relativement restreintes, ce critère est le plus important et le plus étudié [MAR 04a]. Cependant, les sites Web utilisant des systèmes de FC doivent aussi pouvoir faire des recommandations à des utilisateurs immédiatement après leur inscription (i.e. sans attendre la mise à jour du site), c'est à dire à des utilisateurs non présents dans la base d'apprentissage. C'est cette capacité que mesure la généralisation *forte*.

Complexité de l'apprentissage : cette complexité se traduit en terme du temps de réponse des systèmes de FC. Bien que la plupart des algorithmes considèrent un entraînement hors-ligne (par exemple une fois par jour ou par semaine), les mises à jour des notes prédites par rapport aux mises à jour des jugements utilisateurs doivent pouvoir se faire en temps raisonnable, sous peine d'être inapplicables en pratique.

Complexité de la prédiction : cette complexité se traduit en terme du temps de calcul nécessaire pour effectuer une recommandation. D'une façon générale, les systèmes de FC ont pour but d'être implantés sur des sites Internet, et utilisés pour faire des

recommandations spécifiques à un utilisateur et à une visite. Il est donc nécessaire que cette complexité soit faible pour effectuer les recommandations en temps réel.

À ces caractéristiques, [HOF 04] rajoute la possibilité pour un algorithme de FC d'être utilisé comme outil de fouille de données. De façon équivalente, il s'agit de savoir si un algorithme permet de fournir des explications de ces prédictions. [POL 04] montrent que certains modèles de FC possèdent des interprétations et des visualisations naturelles permettant d'identifier des communautés d'utilisateurs ou d'améliorer les systèmes ou les interfaces utilisateurs. La notion d'interprétabilité des modèles est donc une caractéristique importante d'un algorithme de FC.

Enfin, nous pensons que la facilité d'implantation est un critère déterminant dans la conception d'un algorithme de FC. En effet, il apparaît de plus en plus de sites non commerciaux de mise en partage d'information. L'implantation d'interfaces Web permettant de demander des jugements utilisateurs, ainsi que des bases de données les enregistrant est aujourd'hui à la portée de tous les développeurs Web. La création d'algorithmes de FC accessibles aux non spécialistes pourrait donc permettre la généralisation de cette technologie sur des sites maintenus par des particuliers comme des blogs, des forums, et plus généralement des sites de partage d'information.

Dans la suite de cette section, nous présentons l'utilisation de la réduction dimensionnelle pour le FC. Nous présentons notre méthode dans la section 3, et montrons que son avantage par rapport aux méthodes existantes est de posséder à la fois de bonnes performances en terme de prédiction de notes, une faible complexité d'apprentissage et une complexité de prédiction indépendante du nombre d'utilisateurs du site, tout en permettant de fouiller les bases de jugements et d'être extrêmement simple à implanter.

2.2. Méthodes de réduction dimensionnelle

Le but de la réduction dimensionnelle est de représenter les données dans un espace de dimension inférieure à celle de départ. Une approche pour y parvenir est la factorisation matricielle, qui offre un ensemble de méthodes parmi lesquelles la décomposition en valeurs singulières (DVS). Cette dernière permet de factoriser une matrice V en un produit UST^t , où U et T sont des matrices orthogonales contenant les vecteurs singuliers respectivement gauches et droits de V , et S est une matrice diagonale contenant les valeurs singulières de V . La meilleure approximation de rang k de V pour la norme de Frobenius $\|V - \tilde{V}\| = \sqrt{\sum_{ij} (V_{ij} - \tilde{V}_{ij})^2}$ est obtenue en ne gardant que les k plus grandes valeurs singulières.

Pour le FC, on peut faire l'hypothèse que les notes des utilisateurs sont des combinaisons d'un petit ensemble de facteurs communs. Sous cette hypothèse, une DVS appliquée à la matrice (utilisateurs, articles) permettrait d'identifier ces facteurs (les vecteurs ligne de T), ainsi que les combinaisons correspondantes pour les utilisateurs grâce à la matrice U . Bien que la plupart des valeurs sont inconnues, la décompo-

sition est possible grâce à la DVS pondérée ([SRE 03]), qui permet ensuite de prédire les notes manquantes. Le critère minimisé est la norme de Frobenius pondérée $\mathcal{L}(V, \tilde{V}) = \sum_{i,j} a_{ij}(V_{ij} - \tilde{V}_{ij})^2$, où a_{ij} est l'indicateur valant 0 si v_{ij} est une donnée manquante, 1 sinon (on garde cette notation dans la suite). L'algorithme employé, de type EM, et au moyen duquel les données manquantes sont ré-estimées à chaque itération, est simple à mettre en oeuvre grâce aux nombreuses implantations de la DVS. La méthode obtient des bonnes performances que ce soit en faible ou forte généralisation [MAR 04a]. Cependant, la complexité de l'algorithme de DVS est de l'ordre de $O(nm^2 + m^3)$, où n est le nombre d'utilisateurs et m le nombre d'articles, ce qui rend l'apprentissage coûteux, donc difficile à appliquer sur de grandes bases.

3. FMN pour le filtrage collaboratif

La factorisation en matrices non négatives (FMN) est une méthode générale de décomposition matricielle, introduite par [LEE 99]. Elle permet d'approximer toute matrice V de taille $(n \times m)$ et dont les éléments sont tous positifs, grâce à une décomposition de la forme $V \approx WH$, où W et H sont des matrices $(n \times k)$ et $(k \times m)$. La matrice V contient les vecteurs réels de dimension m , la matrice W contient les vecteurs correspondants dans un espace de dimension $k < m$, et la matrice de passage H contient les vecteurs de base. L'originalité de la FMN réside dans les contraintes de non-négativité qu'elle impose à W et H ; c'est à dire que leurs éléments doivent être tous positifs. Ces contraintes font que les vecteurs de base comportent beaucoup de 0 et que leurs parties non nulles se chevauchent rarement. La représentation d'un objet (décrit par un vecteur de réels positifs) comme une somme de ces vecteurs de base, correspond alors à l'intuition d'une décomposition par parties. En comparaison, la DVS ne permet d'interpréter ni les vecteurs de base ni les nouvelles coordonnées réduites, à cause de la présence de coefficients négatifs.

La FMN a été appliquée avec succès, notamment en reconnaissance des visages ([LEE 99]) et en classification de documents textuels ([XU 03], [BER 04]). Ici nous l'appliquons au FC, motivés par l'idée de pouvoir représenter les utilisateurs comme des sommes de comportements types, ces derniers étant alors directement interprétables (ils correspondent à des utilisateurs imaginaires ayant noté des sous-ensembles d'articles). Cette interprétabilité est développée dans la section 5.2.

Déterminer les matrices W et H revient à minimiser la distance entre la matrice initiale et le produit WH ; plus précisément, il faut minimiser la norme de Frobenius $\|V - WH\|^2 = \sum_{i,j} (V_{ij} - (WH)_{ij})^2$ sous les contraintes de non-négativité. C'est un problème d'optimisation non trivial, que [LEE 00] proposent de résoudre en initialisant W et H aléatoirement, puis en alternant les mises à jour suivantes :

$$H_{ij} \leftarrow H_{ij} \frac{(W^T V)_{ij}}{(W^T W H)_{ij}} \quad W_{ij} \leftarrow W_{ij} \frac{(V H^T)_{ij}}{(W H H^T)_{ij}}$$

Les auteurs montrent que leur algorithme converge vers un minimum local. Sa complexité est en $O(nmkt)$, où t est le nombre maximal d'itérations.

3.1. Algorithme

La version originale de la FMN ne permet pas de décomposer une matrice avec des données manquantes, et n'est donc pas directement applicable dans le cadre du FC. En nous inspirant des travaux de [SRE 03], nous en proposons ici une version pondérée, qui minimise la norme de Frobenius pondérée entre V et WH :

- Répéter jusqu'au critère d'arrêt
- (a) Etape E : $V \leftarrow A \odot V + (1 - A) \odot \tilde{V}$
- (b) Etape M : $[W, H] \leftarrow FMN_L(V)$
- (c) si $L > k$ alors réduire(L)

où \odot est le produit de Schur (élément par élément), A la matrice d'indicateurs a_{ij} , et \tilde{V} la matrice de données manquantes, initialement nulles. A l'étape E, ces dernières sont complétées par la reconstruction de la FMN au rang L . Comme le suggèrent [SRE 03], le rang de l'algorithme est initialisé avec un rang $L > k$, et, à chaque itération, la valeur de L est diminuée jusqu'à la valeur désirée k . Une fois atteint, le rang reste constant.

3.2. Prédiction de notes

En généralisation faible, la prédiction des notes manquantes avec les approches DVS pondérée et FMN pondérée est immédiate : il suffit de prendre les matrices obtenues par factorisation, et de les multiplier pour reconstruire la matrice finale \hat{V} de rang k . La note prédite pour l'utilisateur i et l'article j vaut simplement \hat{V}_{ij} .

En généralisation forte, le système doit être capable de prédire les notes d'utilisateurs qu'il n'a pas vus en apprentissage, ce que ne permet pas la version initiale de la DVS pondérée. [MAR 04a] propose une procédure pour y remédier, que nous avons adaptée dans le cadre de la FMN de la façon suivante :

- Répéter jusqu'au critère d'arrêt
- (a) $x \leftarrow a^u \odot v^u + (1 - a^u) \odot v^u$
- (b) $l^u \leftarrow \min_l \|x - H^t l\|^2$
- (c) $\hat{v}^u \leftarrow H^t l^u$

où v^u est le vecteur des notes disponibles pour l'utilisateur u , a^u est le vecteur d'indicateurs associés aux notes, H est obtenue par la factorisation FMN sur la base d'apprentissage. Dans le cas où les données manquantes doivent être entières, nous arrondissons les prédictions à l'entier le plus proche. De plus, nous remplaçons les valeurs inférieures à la note minimale par la note minimale, et les valeurs supérieures à la note maximale par la note maximale.

4. Expériences

La tâche Une application répandue du FC est la recommandation de films : dans ce problème, des utilisateurs expriment leurs goûts cinématographiques en attribuant des notes à des films. Les bases existantes référencant un grand nombre de films, chaque utilisateur n'en a en général noté qu'une petite partie. La tâche consiste alors à prédire les notes restantes, c'est à dire celles qui n'ont pas été renseignées par les utilisateurs. Pour faire de la recommandation, il suffit alors d'utiliser ces prédictions pour ordonner les films non vus, et de proposer à l'utilisateur les films les mieux placés. La prédiction des notes est donc une étape essentielle du processus de recommandation, et c'est à ce problème que nous appliquons notre approche.

La base Nous utilisons la base de films MovieLens 100K. Elle contient les notes de 943 utilisateurs sur un ensemble de 1 682 films, représentant un total de 100 000 notes. Les films sont notés sur une échelle de 1 à 5, et chaque utilisateur a noté au moins 20 films ; au total, 93.7% des notes sont manquantes.

Evaluation et mesure d'erreur Pour générer les bases d'apprentissage et de test, nous utilisons le protocole "all-but-one" : il consiste, pour chaque utilisateur, à retirer une note et à apprendre sur toutes les autres. Ainsi, en généralisation faible, on retire une note par utilisateur pour constituer la base d'apprentissage ; on calcule ensuite l'erreur moyenne de prédiction sur les notes qu'on a retirées. En itérant ce processus 10 fois, nous avons créé au total 10 bases d'apprentissage et les 10 bases de test correspondantes. En généralisation forte, on considère un utilisateur non présent dans la base d'apprentissage ; on retire aléatoirement une de ses notes, qu'on essaye de prédire à partir des notes restantes et de l'algorithme décrit en section 3.2. L'erreur de prédiction est moyennée sur plusieurs de ces tirages aléatoires.

Pour calculer cette erreur en prédiction sur les bases de test, nous définissons un coût entre la note $v_{y^u}^u$ attribuée par l'utilisateur u à l'article y^u et la note $\hat{v}_{y^u}^u$ prédite par le modèle. Plus précisément, en notant n le nombre d'utilisateurs et $v_{y^u}^u$ la note cachée, on calcule l'erreur absolue moyenne $MAE = \frac{1}{n} \sum_u |\hat{v}_{y^u}^u - v_{y^u}^u|$. Comme le suggère [MAR 04a], nous adoptons une mesure normalisée NMAE qui permet de comparer les erreurs obtenues sur des bases différentes. Cette mesure normalisée vaut $MAE/E[MAE]$, où $E[MAE]$ est la moyenne de MAE en supposant que les notes observées et les notes prédites sont distribuées uniformément. Par exemple, en choisissant une échelle de notes de 1 à 5, on obtient $E[MAE] = 1.6$.

Nous avons comparé nos résultats avec ceux de la DSV pondérée décrite précédemment ([SRE 03]), et dont [MAR 04a] a montré qu'elle permettait d'obtenir de bons résultats, notamment sur la base MovieLens. A noter que les résultats que nous présentons ici ne concernent que la généralisation faible, nous laissons les expériences en généralisation forte pour des travaux futurs.

Paramètres des expériences Dans nos expériences, nous avons choisi $L = 100$ pour le rang initial, et $k = 6$ pour le rang final. Nous avons également stoppé la descente de rang à des moments intermédiaires de la factorisation, pour $L = 10, 20, 30, 40, 50,$

afin de les continuer à rang fixé. Dans tous les cas, lorsque la factorisation est terminée, nous avons calculé l'erreur NMAE en test à partir de la reconstruction de la matrice au rang d'arrêt.

5. Résultats

5.1. Performances

Nous présentons ici les erreurs NMAE moyennées sur les bases de test, et calculées après convergence de l'algorithme EM pour différentes valeurs du rang final k :

k	6	10	20	30	40	50
DVS pond.	0.4678	0.5012	0.5704	0.6306	0.6568	0.7404
FMN pond.	0.4821	0.4692	0.5008	0.5163	0.5314	0.5380

D'après nos résultats, l'erreur NMAE en test est optimale à $k_{DSV} = 6$ pour la DSV pondérée, et à $k_{FMN} = 10$ pour la FMN pondérée. La dégradation des performances de la DVS pondérée lorsque k augmente semblent être due à un phénomène de surapprentissage ; on a en effet observé une diminution rapide des erreurs en apprentissage parallèlement à la dégradation des performances en généralisation ; si FMN fait moins bien en apprentissage, en revanche ses performances en test restent bonnes, et semble ne pas souffrir de ce défaut. On peut également remarquer que le rang optimal pour FMN est supérieur au rang optimal pour DVS ; ceci peut s'expliquer intuitivement : avec la DVS, un poids positif pouvait en compenser un négatif ; ce degré de liberté disparaît avec la contrainte de non-négativité de FMN, rendant plus difficile l'approximation pour un même rang k . Pour atteindre une qualité d'approximation égale, on pouvait donc s'attendre à ce que k_{FMN} soit supérieur à k_{DSV} .

Enfin, remarquons qu'en choisissant un rang adéquat, les performances de la DSV et de la FMN sont tout à fait comparables ; grâce aux propriétés de la FMN, notre approche offre l'avantage supplémentaire de fournir des résultats interprétables et visualisables, comme nous l'expliquons dans la sous-section suivante.

Temps de calculs Nous avons comparé les temps d'apprentissage des deux algorithmes sur la base 100k. Si l'on considère que l'algorithme EM a convergé au bout de 1000 itérations, alors la NMF pondérée est en moyenne 18 fois plus rapide que la DVS pondérée : 1h30 en moyenne pour notre approche, contre plus d'une vingtaine d'heures pour la DVS pondérée.

5.2. Interprétation et visualisation

Description La FMN nous a permis de représenter les profils utilisateurs comme des sommes pondérées des k vecteurs de bases, tous à coefficients positifs. Chacun de ces vecteurs peut alors être interprété comme un comportement type, c'est à dire

comme le profil d'un utilisateur imaginaire spécialisé sur un sous-ensemble de films, et dont les goûts sont représentatifs d'une communauté d'utilisateurs.

Une première étape dans l'analyse des utilisateurs de la base est l'interprétation des comportements types. Dans ce but, nous proposons trois techniques permettant de les visualiser, et les illustrons pour $k = 10$. La première visualisation, issue d'une analogie avec les modèles d'aspects de [HOF 04] et de [POL 04], consiste à donner les films les plus représentatifs par comportement type (tableau 1). Bien qu'imparfaite, elle permet d'avoir un aperçu rapide de la différence entre les comportements types (en particulier le fait qu'ils se focalisent sur des films différents). Nous proposons alors d'exploiter, comme dans [POL 04], la répartition par genre des films selon les comportements types, ainsi qu'une représentation considérant la date de réalisation des films. Deux exemples de répartition par genres sont donnés en figure 1. Nous observons que le comportement 0 apprécie les films mystérieux et les documentaires, tandis que le comportement 3 préfère les films fantastiques et les comédies. D'une façon générale, les comportements tendent à se focaliser sur des genres différents. Enfin, la répartition des films par période, illustrée en figure 2 pour les comportements 0 et 7, nous permet de remarquer que le comportement 0 est principalement focalisé sur des films récents, alors que le 7 l'est sur des films anciens.

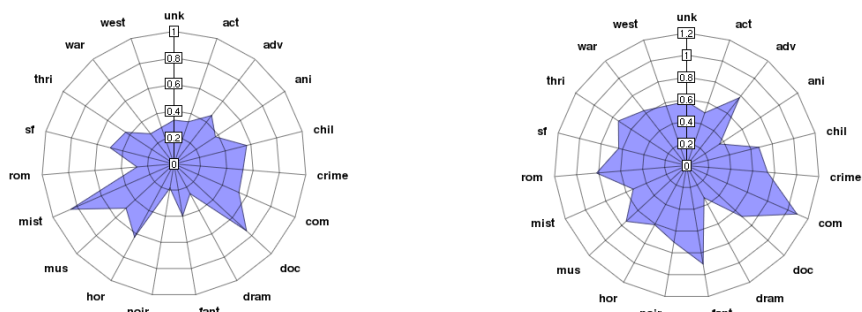
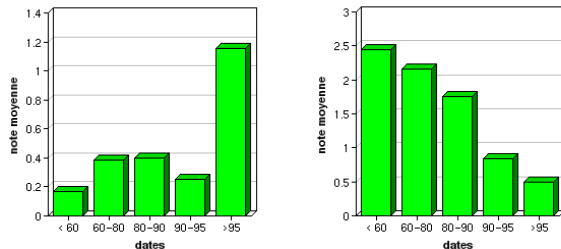


Figure 1. Visualisation pour les comportements types 0 et 3, des notes moyennes par genre : inconnu(unk), action (act), aventure (adv), animation (ani), enfant (chil), policier (crime), comédie (com), documentaire (doc), drame (dram), fantastique (fant), noir (noir), horreur (hor), musical (mus), mystérieux (mist), romantique (rom), science fiction (sf), thriller (thri), guerre (war) et enfin western (wes).

Discussion On peut imaginer plusieurs applications de l'interprétation des vecteurs de base en terme de comportement type : les visualisations, qui ont un intérêt en soi pour le créateur d'un site de recommandation, peuvent être rendues publiques pour que les utilisateurs puissent avoir accès aux comportements des autres utilisateurs du même site. Les regroupements des utilisateurs en communautés peuvent être utilisés pour créer des forums de discussion, et la liste des films favoris par comportement type peut être présentée aux utilisateurs, ce qui permettrait d'une part de collecter plus de jugements pour ces films (permettant ainsi d'affiner les comportements types), mais aussi pour permettre aux utilisateurs d'avoir des listes de films "à voir absolument",



*Périodes considérées :
<1960, entre 1960 et
1970, entre 1980, entre
1980 et 1990, entre
1990 et 1995, > 1995.*

Figure 2. Notes moyennes par période pour les comportements types 0 et 7.

CT 0	CT 1	CT 2	CT 3	CT 4
Sum of us	Wild things	The Joy Luck Club	Schizopolis	les boys
Fresh	the Child Murders at Robin Hood Hills	Men with Guns	Santa with Muscles	They Made Me a Criminal
Hotel of Love	Lost in Space	To live (Huozhe)	Underground	Shadows and Light
Now and then	the Big Bang Theory	Widow's Peak	American Dream	Crossfire
I can't sleep	Night Watch	The big Blue	the Butcher Boy	Aparajito
CT 5	CT 6	CT	CT 8	CT 9
The lady who walked in the sea	The Adventure of Yellow Dog	New York Cop	Safe Passage	Shiloh
A Strong Clear Vision	Angel Baby	Golden Earrings	Stripes	The Shooter
You So Crazy	Love Serenade	The Thin Blue Line	Killer : A Journal of Murder	Pie in the Sky
The Quiet Room	Mondo	The Little Princess	The Show	Total Eclipse
Microcosmos	Highlander III	The Innocent	He Walked By Night	Zeus and Roxanne

Tableau 1. Les 5 films les mieux notés par comportement type (CT)

selon la communauté de laquelle est issue la liste, affinant ainsi les recommandations. D'une façon générale, les utilisations possibles d'un tel outil de fouille de données sont immenses, et l'intérêt principal de la méthode FMN proposée ici est d'être accessible à tout créateur de site Web, et applicable à toutes les échelles, même si le nombre d'utilisateurs ou d'articles considérés dans le site sont très importants.

6. Conclusion et Perspectives

Nous avons présenté une nouvelle approche pour le FC basée sur la méthode de réduction dimensionnelle appelée factorisation en matrices non négatives. Elle permet de représenter les vecteurs de notes de chaque utilisateur sur une base de petite dimension, où chaque vecteur de base peut être interprété comme un ensemble de notes typiques associées à un sous-ensemble de la collection. En terme de prédiction de notes, l'approche a une performance comparable avec la méthode de réduction dimensionnelle la plus utilisée dans l'état de l'art, à base de décomposition en valeurs singulières. Comme cette dernière, notre algorithme est très simple à implanter, et peut en particulier être utilisé par des non-spécialistes pour créer des systèmes performants et rapides grâce à la faible complexité de la méthode. De plus, les regroupements d'articles dont les notes sont corrélées qu'effectue la méthode possèdent des visualisations

simples qui permettent de comprendre les comportements principaux des utilisateurs. La méthode peut donc être utilisée comme outil de fouille de données pour le FC et peut être utilisée pour améliorer le système, l'interface utilisateur, ou simplement pour mieux comprendre le comportement des utilisateurs. Ainsi, nous pensons avoir fourni une alternative intéressante aux algorithmes de FC utilisant la DVS.

Par manque de temps, nous n'avons pas testé les performances de notre méthode en généralisation forte. De plus, les expériences n'ont pour l'instant été faites que sur la plus petite des bases MovieLens. Une direction de travail futur consiste donc à faire des expériences à plus grande échelle, mais aussi sur d'autres tâches que la recommandation de films. Par ailleurs, l'aspect visualisation des comportements types est encore peu étudié en filtrage collaboratif. Une autre direction de travail est l'étude précise, soit des améliorations dans les recommandations, soit dans l'interface utilisateur qu'une telle compréhension des utilisateurs permet.

Remerciements Ce travail a été partiellement financé par le programme IST de la Communauté Européenne, dans le cadre du réseau PASCAL, IST-2002-506778.

7. Bibliographie

- [BAS 04] BASILICO J., HOFMANN T., « A joint framework for collaborative and content filtering », *SIGIR '04*, New York, NY, USA, 2004, ACM Press, p. 550–551.
- [BER 04] BERRY M., SHAHNAZ F., PAUCA P., PLEMMONS R., « Document Clustering using Nonnegative Matrix Factorization », *IPM Journal*, , 2004.
- [HER 04] HERLOCKER J. L., KONSTAN J. A., TERVEEN L. G., RIEDL J. T., « Evaluating collaborative filtering recommender systems », *ACM Trans. Inf. Syst.*, vol. 22, 2004.
- [HOF 04] HOFMANN T., « Latent semantic models for collaborative filtering », *ACM Trans. Inf. Syst.*, vol. 22, n° 1, 2004, p. 89–115, ACM Press.
- [LEE 99] LEE D. D., SEUNG H. S., « Learning the parts of objects by non-negative matrix factorization », *Nature*, vol. 401, 1999, p. 788–791.
- [LEE 00] LEE D. D., SEUNG H. S., « Algorithms for Non-negative Matrix Factorization », *NIPS*, 2000, p. 556–562.
- [MAR 04a] MARLIN B., « Collaborative filtering : A machine learning perspective », PhD thesis, University of Toronto, 2004.
- [MAR 04b] MARLIN B., « Modeling user rating profiles for collaborative filtering », 2004.
- [POL 04] POLCICOVÁ G., « Topographic Organization of User Preference Patterns in Collaborative Filtering », PhD thesis, Slovak University of Technology in Bratislava, 2004.
- [SAR 00] SARWAR B., KARYPIS G., KONSTAN J., RIEDL J., « Application of dimensionality reduction in recommender systems—a case study », 2000.
- [SHA 05] SHANI G., HECKERMAN D., BRAFMAN R. I., « An MDP-Based Recommender System », *Journal of Machine Learning Research*, vol. 6, 2005.
- [SRE 03] SREBRO N., JAAKKOLA T., « Weighted Low-Rank Approximations », *ICML*, 2003.
- [XU 03] XU W., LIU X., GONG Y., « Document clustering based on non-negative matrix factorization », *SIGIR '03*, 2003, p. 267–273.