

État de l'art sur les Systèmes de Recommandation

Nicolas Béchet

nicolas.bechet@inria.fr

Projet AxIS de l'INRIA, dans le cadre du projet Addictrip

Résumé

Ce document fait état des différents types de systèmes de recommandation de la littérature. Nous nous intéressons dans un premier temps aux origines de ces systèmes ainsi que leurs objectifs.

Sont alors présentées différentes approches de recommandation en distinguant les trois principaux types de systèmes : ceux basés sur le contenu, ceux basés sur un filtrage collaboratif et finalement les approches hybrides. Nous concluons en présentant les tendances actuelles en matière de recommandation ainsi qu'en insistant sur le rôle des documents à recommander.

Keywords: Recommandation, filtrage collaboratif, analyse de contenu, personnalisation

Contents

1	Introduction	3
1.1	Pourquoi de tels systèmes ?	3
1.2	Historique	4
1.3	Comment classer les différents systèmes de recommandation ?	5
2	Le Filtrage par le Contenu	6
2.1	Définition	6
2.2	La notion de profil dans les systèmes d'informations	6
2.3	Le modèle vectoriel	7
2.4	Exemple de systèmes de recommandation par le contenu	8
2.5	Synthèse	11
2.5.1	Conclusions	11
2.5.2	Les avantages et limites de ces systèmes	11
3	Le Filtrage Collaboratif	12

3.1	Définition	12
3.2	La notion de profil à usage collaboratif	12
3.3	Exemple de systèmes de recommandation collaboratifs	13
3.4	Synthèse	15
4	Les méthodes Hybrides	16
4.1	Définition	16
4.2	Exemple de systèmes de recommandation hybrides	16
5	D'autres approches de recommandation	17
5.1	Hybridation	17
5.2	Les méthodes à bases de connaissances psychologiques	18
6	Conclusion	19

1 Introduction

1.1 Pourquoi de tels systèmes ?

Un **système de recommandation** a pour objectif de fournir à un utilisateur des ressources pertinentes en fonction de ses préférences. Ce dernier voit ainsi réduit son temps de recherche mais reçoit également des suggestions de la part du système auxquelles il n'aurait pas spontanément prêtées attention. L'essor du Web et sa popularité ont notamment contribué à la mise en place de tels systèmes comme dans le domaine du e-commerce. Citons par exemple les sites Web populaires Amazon¹ dans le e-commerce mais également CiteSeer², outil référençant des articles de recherche. Les systèmes de recommandation peuvent être vus initialement comme une réponse donnée aux utilisateurs ayant des difficultés à prendre une décision dans le cadre d'utilisation d'un système de recherche d'information "classique".

La **recherche d'information** [van Rijsbergen and D, 1979] sur fonde sur un principe d'indexation des données afin de répondre aux requêtes d'utilisateurs. Plus spécifiquement, la recherche documentaire, sous discipline de la recherche d'information, consiste à interroger une base de connaissance par le biais de requêtes écrites en langues naturelles ou bien sous forme de mots clefs (nommées requêtes ad hoc). Un exemple bien connu d'outils utilisant les fondements de cette discipline sont les moteurs de recherche Web comme *Google* ou *Yahoo!* où l'utilisateur formule son besoin par une requête en soumettant au moteur un ensemble de mots clefs. Ces mots clefs sont alors comparés à l'ensemble des indexes des documents présents dans la base de données du moteur de recherche. Dès lors, l'utilisateur se voit retourner un ensemble de résultats plus ou moins pertinents par rapport à sa requête initiale. Le problème inhérent à ce type de résultats et la surcharge d'informations que l'utilisateur doit filtrer. Ainsi, la plupart du temps, un utilisateur se contente de consulter la première page de résultats d'un moteur de recherche en ignorant les autres, faisant l'impasse sur des contenus potentiellement pertinents. Le lien avec la recommandation est ici trivial, l'un des objectifs des systèmes de recommandation étant de faire le tri dans cette masse d'informations pour l'utilisateur de manière transparente. La construction de profils d'utilisateurs constitue une première avancée dans le domaine de la recherche documentaire en terme de recommandation. En effet, comme le propose Google notamment, l'utilisateur se voit maintenant pondérer les résultats obtenus suite à sa requête en fonction des précédentes recherches qu'il a effectuées et des pages qu'il a consultées. Il s'agit d'un système d'identification de l'utilisateur permettant alors d'analyser ses actions. Cette approche peut être qualifié d'analyse de traces et constitue l'un des fondements du **filtrage d'informations** sur lequel s'appuient les systèmes de recommandations [Belkin and Croft, 1992].

C'est ainsi que l'on différencie les systèmes de recherche d'informations, pour lesquels la demande de l'utilisateur de l'orienter et le guider vers des choix appropriés est explicite,

¹<http://www.amazon.com>

²<http://citeseer.ist.psu.edu> et plus récemment <http://citeseerx.ist.psu.edu>

et les systèmes de recommandation où la participation de l'utilisateur du système est non volontaire [Maltz and Ehrlich, 1995], induite notamment par des moteurs de traces [Resnick and Varian, 1997]. Notons cependant que [Burke, 2002] propose de rapprocher les moteurs de recherche aux systèmes de recommandations en les définissant comme suit. Un profil d'utilisateur se voit caractérisé par les mots clés de la requête qu'il émet, et la recommandation par la liste de résultats ordonnés en fonction du profil de l'utilisateur. Avant de présenter plus en détail ces types de recommandation, revenons sur l'historique de la mise en place de systèmes de recommandation.

1.2 *Historique*

Les préjudes des systèmes de recommandation découle de recherches menées dans la construction de modèles représentant les choix d'utilisateurs. Ces recherches sont issues de domaines distincts tels que la recherche documentaire, les sciences de gestion et marketing, les sciences cognitives et les théories d'approximation [Adomavicius et al., 2005].

La recommandation peut-être comparée à un dialogue entre une personne experte d'un domaine et l'autre désireuse d'acquérir des informations dans ce domaine. Plus concrètement, un bibliothécaire va pouvoir, en fonction des goûts d'un de ses clients, proposer une liste d'ouvrages à ce dernier qui ne sera autre qu'une recommandation au sens des systèmes de recommandation. En considérant cette analogie, le bibliothécaire, compte tenu de sa connaissance des différents ouvrages qu'il propose peut être vu comme la base de connaissances des items à recommander. Il connaît ainsi les items de manière individuelle et est capable d'effectuer des associations d'item suivant différents critères caractérisés par le profil d'un utilisateur. Ce dernier est en fait conscient de ses goûts et peut les soumettre au bibliothécaire. Nous pouvons même aller plus loin en supposant que le bibliothécaire connaisse les goûts des différents clients. Il serait alors en mesure de proposer des ouvrages à un client, qui ont été aimés par d'autres clients similaires. La notion de recommandation induite par cet exemple est sans doute à la base des principes de systèmes de recommandation.

Cette vision de la recommandation conforte la définition de celle-ci comme étant un moyen d'aider un utilisateur à choisir un certain nombre d'items parmi un ensemble de ressources où il n'a pas assez de connaissances de ces dernières afin de trier et d'évaluer la pertinence des documents disponibles.

Parmi les premières approches proposées, citons [Goldberg et al., 1992b], [Resnick et al., 1994], [Shardanand and Maes, 1995] et leurs approches respectives Tapestry, GroupLens, Ringo.

[Goldberg et al., 1992b] sont parmi les premiers à émettre l'hypothèse qu'une intervention humaine pourrait améliorer les résultats de filtrage d'items, des e-mails dans leur cas. Leur objectif est donc d'améliorer les résultats de filtrage d'e-mail en proposant le système de filtrage collaboratif Tapestry. Le principe, novateur à l'époque, est de s'appuyer sur les précédentes requêtes d'utilisateurs du système Tapestry et sur leurs appréciations en

fonction des documents qu'ils ont jugés pertinents afin de filtrer les résultats proposer à de futurs utilisateurs qui feront la même requête. L'originalité est ici, dans un système de recherche d'informations, de non plus s'appuyer uniquement sur l'analyse des documents en fonction de la requête émise, mais de prendre en compte les annotations des utilisateurs du système afin de répondre aux besoins d'autres utilisateurs.

[Resnick et al., 1994] proposent une approche assez similaire en se focalisant sur la recherche documentaire par le biais du système GroupLens. Notons cependant que comme lors de l'utilisation de l'outil Tapestry, GroupLens se fonde sur une appréciation humaine active d'article. Le terme d'actif signifie dans ce cas que l'utilisateur doit faire la démarche d'annoter un document afin de construire des classes de documents pertinents. Ces approches ne sont donc pas fondées sur un apprentissage implicite du comportement d'utilisateurs comme dans le cas d'Amazon [Linden et al., 2003].

Le système Ringo de [Shardanand and Maes, 1995] s'intéresse quant à lui aux albums musicaux en utilisant le principe du filtrage collaboratif mais utilise également un système d'annotation manuel des items. Les auteurs proposent alors une comparaison de quatre algorithmes de recommandation, le premier étant une recommandation sans filtrage collaboratif et les trois autres proposant différents scores de similarités s'intéressant à la proximité dite "social" entre individus comparant alors un utilisateur à un autre. Ainsi, pour produire une recommandation, un utilisateur est comparé à l'ensemble des utilisateurs. Les utilisateurs dont le score obtenu est supérieur à un certain seuil sont retenus. Dès lors, une moyenne des scores des utilisateurs (leurs évaluations des items) est alors conservée afin de proposer au nouvel utilisateur une liste ordonnées d'albums en terme de pertinence. Notons que les meilleurs résultats obtenus l'ont été avec l'algorithme r et un seuil de similarité fixé à 0,6 dont la formule est donnée ci-dessous, avec U_x et U_y représentant deux profils d'utilisateurs.

$$sim = \frac{\sum(U_x - 4)(U_y - 4)}{\sqrt{\sum(U_x - 4)^2 \times \sum(U_y - 4)^2}} \quad (1)$$

1.3 Comment classer les différents systèmes de recommandation ?

Plusieurs facteurs entres en considération afin de catégoriser les systèmes de recommandation.

1. La connaissance de l'utilisateur (c.-à-d. son profil en fonction de ses goûts).
2. Le positionnement d'un utilisateur par rapport aux autres (la notion de classes ou réseaux d'utilisateurs).
3. La connaissance des items à recommander.
4. La connaissance des différentes classes d'items à recommander.

De ces facteurs sont produits divers type de recommandations dont les plus utilisés dans la littérature sont le filtrage basé sur le contenu et le filtrage collaboratif. Ce document

présente dans un premier temps ces deux approches ainsi que leur hybridations.

2 Le Filtrage par le Contenu

2.1 *Définition*

Les systèmes de recommandation basés sur le contenu s'appuient sur des évaluations effectuées par un utilisateur sur un ensemble de documents ou items. L'objectif est alors de comprendre les motivations l'ayant conduit à juger comme pertinent ou non un item donné.

Le système peut alors proposer à l'utilisateur un choix parmi de nouveaux items jugés proches des items qu'il a précédemment appréciés. Ces méthodes font donc appels aux facteurs 1., 3. et 4. présentés en section 1.3. Notons que le filtrage par le contenu peu s'appuyer sur les appréciations à la fois positives mais également négatives d'un utilisateur. [Salton and Buckley, 1988] ont montré en effet dans le cadre du filtrage de descripteurs que l'absence d'un mot est à prendre en compte au même titre que sa présence. L'objectif de ces derniers étant de proposer une mesure de pondération de mots afin de faire émerger les plus représentatifs d'un domaine. Leur protocole expérimental est le suivant. Ils disposent d'un certain nombre de requêtes d'utilisateurs et d'un certain nombre de corpus de domaines différents. Alors, un classement de la proximité des requêtes utilisateurs avec les documents des différents corpus est obtenu en utilisant leurs différentes mesures proposées. Les résultats ont montré que l'utilisation de la mesure bien connue de l'IDF (inverse document frequency), pondérant les termes en fonction de leur absence dans un certain nombre de documents, améliorerait significativement la qualité du classement obtenu.

2.2 *La notion de profil dans les systèmes d'informations*

Les approches présentées dans [Salton and Buckley, 1988] ne sont pas considérées comme des systèmes de recommandation car se contentant d'effectuer un classement d'items possibles. Notons en revanche qu'en introduisant un simple seuil au delà duquel les items sont considérés comme cohérent vis à vis du profil de l'utilisateur (dans l'approche de [Salton and Buckley, 1988] étant définis par les requêtes qu'il a formulé), nous pouvons alors le qualifier de recommandation par filtrage basée sur le contenu. En effet, ce type de filtrage est souvent considéré comme une évolution du filtrage d'information issu du domaine de la recherche d'informations (précédemment évoqué en section 1.1).

A l'origine, les travaux sur le filtrage par le contenu à l'instar de ceux sur le filtrage collaboratif (qui sera présenté dans la section 3), proviennent d'un besoin de sélectionner de l'information pertinente dans les groupes de discussions (NewsGroups) pour des utilisateurs. Les méthodes traditionnelles ayant montré leurs limites comme dans les travaux de [Rewari et al., 1992] qui introduisent un certain nombre d'heuristiques afin de réduire le nombre de documents proposées aux utilisateurs en affinant notamment la sélection

de descripteurs. Sont alors apparus les premières approches de construction de profils d'utilisateurs basés sur le contenu dont quelques unes sont présentées dans la section suivante.

2.3 Le modèle vectoriel

Un certain nombre d'approche étant basé sur des modèles vectoriels, employant notamment des approches de clustering, cette section présente brièvement quelques fondement de ces approches.

L'idée d'une représentation de documents dans un espace vectoriel introduit la notion de matrice numérique. Cette dernière va être constituées de $m \times n$ cellules. En d'autres termes, elle possèdera n colonnes de dimension m ou bien m rangs de dimension n . Outre la facilité de l'approche afin de représenter des données, cette matrice va également permettre d'effectuer un certain nombre de calculs vectoriels.

Appliqué aux documents textuels, nous nous fonderons sur le principe de *sac de mots*. Ainsi, les approches de ce type considèrent les mots comme de simples séquences non ordonnées. Le principe général est de représenter un corpus, contenant un ensemble de documents formés de descripteurs (souvent le mot) par une matrice de type documents/descripteurs telle que représentée dans la figure 1. Le principe est de définir

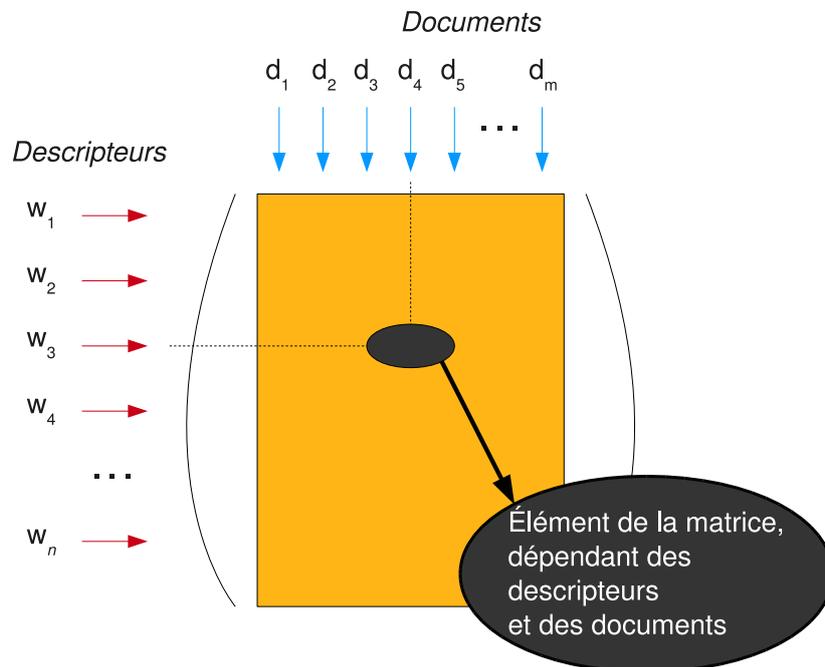


Figure 1: Représentation vectorielle d'un document.

chaque vecteur de descripteurs en fonction des documents dans lesquels ils apparaissent. La définition inverse peut également être utilisée suivant le type de besoin à savoir

un document est représenté par les descripteurs qu'il possède. Cette représentation est fréquemment nommée la représentation ou matrice de *Salton* et fut pour la première fois utilisée dans le cadre du système de recherche d'information *SMART* [Salton and Lesk, 1965]. Le modèle vectoriel de Salton est décrit précisément dans [Salton et al., 1975a]. Notons que ce paragraphe est extrait de [Béchet, 2009] et que des informations complémentaires peuvent y être trouvée concernant les modèles vectorielles et les algorithmes de classification.

2.4 Exemple de systèmes de recommandation par le contenu

Les systèmes de recommandation utilisant uniquement un système de filtrage par le contenu sont assez rares et anciens. En effet, la plupart des systèmes recommandation utilisent le filtrage collaboratif. Ce dernier est d'ailleurs à l'origine de la notion de "recommandation". Ainsi, sont présentées dans cette section deux approches visant à faire de la recommandation d'articles en fonction de profils d'utilisateurs.

[Fischer and Stevens, 1991] proposent le système INFOSCOPE afin de faire de la recommandation à des utilisateurs de groupes de discussions en fonction des requêtes qu'ils soumettent. Le principe est de construire une représentation sémantique d'un utilisateur en fonction de ses préférences en s'appuyant sur une structure d'arbre afin de catégoriser les articles des NewsGroups. L'idée est alors de construire des NewsGroups "virtuelles" correspondant aux profils d'utilisateurs. Plus concrètement, le système peut se décomposer en plusieurs étape :

1. L'utilisateur interagit avec le système de NewsGroups géré par INFOSCOPE sans recommandation.
2. Si l'utilisateur a consulté plus de 15% des messages disponibles dans un groupe particulier, ce groupe est défini comme pertinent pour l'utilisateur
3. Quand un nombre suffisant de groupes ont été rassemblés dans un sous groupe virtuelle représentant l'utilisateur, une recommandation d'articles peut lui être proposée.
4. Si l'utilisateur juge non pertinente la recommandation, le système s'adapte pour les futures recommandations.

D'autres approches comme celle de [Sheth and Maes, 1993] se sont également intéressées à cette problématique. L'auteur présente le système NEWT, se basant sur le retour de pertinences d'utilisateurs (*relevance feedback*) et d'un algorithme de classification afin d'effectuer un filtrage personnalisé d'informations. Le retour de pertinence utilisateur a été choisi afin de permettre une mise à jour dynamique des choix de l'utilisateur. L'algorithme de classification, un algorithme génétique, a été quant à lui utilisé pour sa capacité à sortir des minimums locaux lors des phases de mutations. Ces dernières

permettent en effet de proposer à l'utilisateur des choix plus éloignés que ceux présents dans son profil. Par cet algorithme, un des problèmes majeur inhérent aux systèmes de filtrages par le contenu peut être résolu comme précisé ci-dessous.

Outres les approches filtrages de NewsGroups, les recommandations basées sur le contenu ont également été employées dans le cadre d'applications Web. Les travaux de [Lieberman, 1995] présentent le système LETIZIA destiné à assister les utilisateurs dans la navigation Web. L'assistant personnel LETIZIA présente l'originalité d'être l'un des premiers à proposer des recommandations à un utilisateur via un système de filtrage basé sur le contenu sans l'intervention explicite de l'utilisateur (son opinion sur la page qu'il visite ne lui est pas demandée). Le principe est d'enregistrer les URLs visitées par l'utilisateur, lire ces pages et de construire dynamiquement un profil d'utilisateur. Dès lors, des pages sont recommandées à l'utilisateur dans une fenêtre indépendante de celle destinée à la navigation Web. Ce système propose également d'explicitier son choix en informant l'utilisateur avec par exemple un message du type : *Vous avez visités des pages contenant les thématiques sport et voyage, ces liens sont fortement corrélés à ces thématiques..* Plus formellement, le système utilise des techniques statistiques afin de construire un profil utilisateur en se fondant sur des outils d'extraction de descripteurs ou mots clés. Les informations ciblées sont les liens des pages visitées par les utilisateurs. Notons que la méthode d'extraction de descripteurs n'est pas précisément décrite dans [Lieberman, 1995].

Sur la même thématique de recherche, [Pazzani et al., 1997] proposent d'identifier des sites pertinents pour un utilisateur sur un domaine fixé, en montrant la qualité du classificateur Bayésien [Duda and Hart, 1973] comparativement à d'autres systèmes de classification. Les données "appries" sont dans un premier temps projetées dans un espace vectoriel de co-occurrences des mots en fonction des pages Web dans lesquels ils apparaissent et ceux pour un utilisateur donné.

Il en ressort ainsi un profil. Dès lors le classificateur bayésien est employé afin de construire des classes d'individus. Les expérimentations ont montré qu'à partir de 30 exemples appris, la qualité des résultats de recommandations augmentait significativement. Les auteurs proposent également l'introduction d'informations sémantiques en utilisant WordNet [Miller et al., 1990], ce qui accroît encore les performances.

Bien que les systèmes de recommandation à base collaboratives soient les plus populaires, nous trouvons dans la littérature des systèmes de recommandation montrant que les approches basées sur le contenu peuvent permettre de produire des systèmes de recommandation très efficaces. Citons par exemple [Mooney and Roy, 1999] qui prône justement l'utilisation de ce type d'approches. Il s'appuie sur un exemple concret pour lequel les systèmes de recommandations collaboratifs sont limités. Il s'agit de la recommandation d'ouvrages. Les limites rencontrées viennent du fait que les livres populaires empruntés dans les bibliothèques le sont par un nombre conséquent d'individus, comme c'est par exemple le cas dans les bibliothèques universitaire [Kent and of Pittsburgh., 1979]. Ainsi,

il est difficile de s'appuyer sur la connaissance de classes d'individus similaires (à la base des systèmes collaboratifs, *cf. section 3*). Il y aurait en effet une classe dominante dans laquelle une grande majorité d'individus seraient présents, rendant la recommandation superflue. En partant de cette constatation, les auteurs proposent le système LIBRA destiné à faire de la recommandation de livres à partir du site Web d'Amazon.

Le déroulement du processus de recommandation de LIBRA peut se résumer de la manière suivante.

1. Extraction d'informations. Dans un premier temps, le système interroge Amazon sur une thématique particulière et obtient une liste de livres. L'information pertinente du livre est extraite à partir du titre de l'ouvrage et le profil de l'auteur en utilisant des patrons.
2. Évaluation utilisateur. L'utilisateur doit donner une note de 1 à 10 aux livres proposés par le système.
3. Apprentissage de profils. Les choix effectués par les différents utilisateurs sont appris par un classificateur Bayésien.
4. Recommandation. Le système propose de nouveaux livres ordonnés en termes d'adéquation avec les profils des utilisateurs.

Notons que parallèlement aux travaux de [Kent and of Pittsburgh., 1979], Amazon proposaient un système de recommandation de livres sur leur site internet par une méthode collaborative. Bien que les résultats de cette approche soient pertinents, on aurait aimé une comparaison avec le système d'Amazon. Notons cependant que les auteurs discutent en fin d'article sur la possibilité de combiner les deux approches.

Le dernier article présenté dans cette section traite d'un domaine connexe à la recommandation : déterminer si des documents contiennent des informations redondantes ou nouvelles parmi un ensemble de documents préalablement étiqueté ou "appris". Ces travaux d'une certaine manière assez proches des tâches à réaliser lors d'un processus de recommandation par le contenu. L'objectif d'un tel système est en effet de construire un profil d'utilisateur en fonction de ses goûts. Dès lors, la tâche revient à savoir si le profil de l'utilisateur est redondant avec d'autres documents.

[Zhang and Callan, 2002] s'intéressent ainsi aux problématiques rencontrés lors de la mise en place d'un système de recommandation par le contenu. Notons cependant que les auteurs se situent plus dans le domaine du filtrage d'informations que de la recommandation, leurs objectifs étant l'identification de documents redondants par opposition aux documents contenant de l'information nouvelle. Les auteurs utilisent un classificateur Bayésien en expérimentant différentes mesures de similarités. Leurs résultats montrent cependant que la mesure *Cosinus* reste une des plus performantes permettant d'obtenir un taux d'erreur de 18,7%.

2.5 Synthèse

2.5.1 Conclusions

Les approches présentées dans cette section ont une structure similaire s'appuyant sur la démarche suivante.

1. Constitution de profils utilisateurs
2. Construction de classes d'items à prédire
3. Recommandation d'items aux utilisateur via l'utilisation d'une mesure de proximité
4. Enrichissement des profils utilisateurs : Intervention de l'utilisateur pour valider la recommandation (facultatif)

Parmi les points évoqués, plusieurs approches sont envisagées dans la littérature. Les profils d'utilisateurs nécessitent une première étape de sélection de descripteurs visant en général à construire un espace vectoriel dans lequel sont projetés les profils des différents utilisateurs. L'approche la plus triviale est l'utilisateur du modèle vectoriel de Salton décrit précisément dans [Salton et al., 1975b].

2.5.2 Les avantages et limites de ces systèmes

Un système de recommandation basé sur le contenu ne nécessite pas de connaissances du domaine étudié. En effet, seul la connaissance de l'utilisateur est requise. La caractère dynamique de ces systèmes est également un avantage car plus l'utilisateur va utiliser le système et plus la pertinence des items qui lui seront proposés sera fine.

En revanche, un utilisateur ne se verra jamais proposer d'items qui n'auront pas été jugés similaires à ceux qu'il apprécie. Cela pose le problème de la redondance thématique des propositions soumises à l'utilisateur. En effet, si un utilisateur ne s'intéresse qu'aux articles parlant de sport, il ne se verra jamais proposer un article politique. Bien que cet exemple volontairement exagéré ne se rencontre que rarement, il a l'avantage d'être pédagogique et d'explicitier cette limitation. Une manière d'y remédier est par exemple d'utiliser des algorithmes de classification se permettant des propositions pseudo-aléatoire comme c'est la cas avec les algorithmes génétiques précédemment évoqués [Sheth and Maes, 1993].

Notons pour finir que ces méthodes posent également le problème du nouvel arrivant. Un utilisateur qui n'aura jamais utilisé le système ne pourra se voir proposer des recommandation pertinentes. Le système manquera en effet d'informations. Un certain nombre d'heuristiques peuvent cependant résoudre ce problème comme par exemple ne pas faire de recommandations avant d'avoir recueilli assez d'informations [Fischer and Stevens, 1991].

3 Le Filtrage Collaboratif

3.1 *Définition*

La notion de filtrage collaboratif est à la base de la recommandation, les méthodes de filtrage par le contenu étant plutôt liées aux systèmes de recherche d'informations dits personnalisés. Elle se fonde non plus sur la notion de proximité d'un couple "nouvel item – profil utilisateur" mais cherche à rapprocher l'utilisateur courant avec un ensemble d'utilisateurs existants. L'idée est ici non plus de s'intéresser spécifiquement au nouvel item qui serait susceptible de plaire à l'utilisateur mais de regarder quels items ont appréciés les utilisateurs proches de l'utilisateur courant. Ces méthodes sont ainsi appelées aux facteurs 1., 2. et éventuellement 4. présentés en section 1.3.

Ces approches offrent l'avantage de se dispenser de la connaissance des items que l'on recommande. Il n'est, en d'autres termes, pas nécessaire d'analyser le contenu des items. En effet, il suffit de connaître la proximité des utilisateurs en fonction de leur profil. Toute la difficulté de ce type de recommandation réside ainsi dans la construction de profils utilisateurs.

3.2 *La notion de profil à usage collaboratif*

A l'instar des approches de recommandation basées sur le contenu, il est nécessaire de construire un profil de l'utilisateur courant afin de faire de la recommandation collaborative. Cependant, deux notions diffèrent par rapport à la recommandation sur le contenu.

1. Le profil de l'utilisateur courant doit être connu mais également le profil de tous les autres utilisateurs
2. Il n'est pas nécessaire de construire des profils utilisant le même espace de description que celui des items comme cela doit être le cas lors de la recommandation à base de contenu. En effet, dans le cas de ce dernier, le profil utilisateur et les items doivent être comparables.

Ainsi, ce second point permet de représenter un profil utilisateur avec des techniques plus proches du domaine cognitif comme par exemple la prise en compte de traits de caractère, de personnalité ou encore la prise en compte des émotions. Ces nouvelles approches permettant de construire des profils peuvent être vues d'une manière plus générale comme la prise en compte du profil psychologique de l'utilisateur dans des systèmes de recommandation, travaux sur lesquels a particulièrement travaillé Gustavo González comme par exemple [Gon, 2007]. Ces techniques sont également applicables avec des approches basées sur le contenu dès lors que l'on est en mesure de lier une émotion à un item particulier.

3.3 Exemple de systèmes de recommandation collaboratifs

Les premiers systèmes de recommandation fondés sur un filtrage collaboratif apparaissent dans les années 1990. Les problématiques de l'époque sont axées autour notamment de l'émergence des documents électroniques rendant ainsi la distribution plus aisée. Ainsi, [Goldberg et al., 1992a] proposent par exemple un des premiers systèmes de ce type en l'implémentant dans leur outil TAPESTRY. Ce système comprend plusieurs modules permettant l'indexation des données, un outil de filtrage de l'information, le stockage d'une grande quantité de documents électroniques, et un système d'émission de courrier. L'idée de TAPESTRY est de supposer qu'un système de filtrage d'information, destiné à cibler des utilisateurs, sera plus efficace en faisant intervenir les utilisateurs dans le processus de filtrage. Le principe de fonctionnement de TAPESTRY est de proposer à l'utilisateur d'annoter les documents qu'il consulte, constituant ainsi son profil. Sont alors comparés les profils des utilisateurs ainsi que les documents qu'ils ont consulté afin de proposer des documents proches de leurs attentes. La proximité des profils est réalisée via leur système de filtrage s'appuyant sur un système de requêtes, employant un langage propre à TAPESTRY, le TQL (Tapestry Query Language).

Une question récurrente lors de la mise en place d'un système de recommandation basé sur un filtrage collaboratif est la définition du profil initial. En effet, lors de la mise en place d'un système basé sur le contenu, il est aisé de proposer une première recommandation à un utilisateur. Nous connaissons alors la base contenant les données à proposer à l'utilisateur et les avons préalablement regroupés par catégories prédéfinies. Ainsi, un utilisateur choisit un document, et nous pouvons lui proposer des documents proches de celui qu'il a choisi. Mais comment faire une proposition similaire en ne se basant que sur le filtrage collaboratif. En d'autres termes, si un utilisateur n'a jamais donné son avis sur un document, comment connaître ses préférences ?

Plusieurs approches de la littérature sont proposées afin de résoudre ce problème. La méthode la plus triviale consiste à partir avec un profil vide pour les utilisateurs. Citons par exemple l'approche de [Chen and Sycara, 1998], qui proposent le système WEBMATE permettant à ses utilisateurs de rendre la navigation Web plus efficace, en proposant notamment un agent les guidant dans leurs recherches. La méthode utilisée est vectorielle et propose de construire, dans un premier temps, les profils utilisateurs avec un système Saltonien reposant sur une pondération de type Tf-Idf. Les mots clés représentatifs des choix des utilisateurs sont alors extraits. Dès lors, l'originalité de la méthode est de réintroduire dans la requête de l'utilisateur des résultats jugés pertinents par d'autres utilisateurs proches de lui, utilisant le principe du retour de pertinence (Relevance Feedback). Finalement, lors de la première utilisation du système par un utilisateur, son profil est alors vide et aucune recommandation ne peut lui être faite. Notons que leur système a été évalué avec les moteurs de recherche Altavista et Lycos.

Dans le même domaine, citons l'article de [Cunningham et al., 2000] proposant également

un système d'aide à la navigation sur le Web mais s'intéressant dans ce cas à la vente en ligne. Le principe est de produire un assistant de vente en ligne pour un utilisateur en ciblant dans un premier temps ces choix et en lui recommandant dans un second temps des produits relatifs à son profil. La méthode repose sur le projet WEBSSELL utilisant notamment un filtrage collaboratif.

Le profil est construit à partir d'informations diverses :

1. Les informations personnelles relatives à un utilisateur (age, nom, adresse, etc.)
2. Les domaines de préférences comme voyager en avion plutôt qu'en train, préférer une compagnie aérienne particulière, le budget moyen pour un voyage, etc.
3. La liste des produits qu'il a précédemment sélectionné, constituant la base de connaissance la plus importante dans la mise en place d'un système de recommandation basé sur un filtrage collaboratif.

Notons que ce système ne part ainsi plus d'un profil vide mais de connaissances *a priori* sur l'utilisateur (les points 1 et 2).

Les systèmes évoqués précédemment font apparaître deux types de filtrage collaboratif, et plus particulièrement deux méthodologies permettant la construction d'un profil utilisateur :

1. L'approche dite passive. Elle consiste à construire un profil sans faire intervenir un utilisateur de manière spécifique. En d'autres termes, l'utilisateur ne doit pas annoter telles ou telles ressources, ou bien donner une note à un document.
2. L'approche dite active. Cette dernière, par opposition à la précédente, met pleinement l'utilisateur à contribution en l'interrogeant par exemple sur la qualité d'un film ou d'un livre.

[Nichols, 1998] s'est notamment intéressé à cette problématique en défendant l'évaluation implicite de l'utilisateur, c'est-à-dire une approche passive, en interprétant les actions d'utilisateurs afin de construire leurs profils. L'auteur discute alors des qualités et des inconvénients de ces approches en mettant en avant le temps passé par un utilisateur à annoter un document qui devient nul avec un système passif.

Notons que les approches proposant de ne pas partir de profils vides et reposant sur un modèle de filtrage collaboratif ne peuvent se baser sur une approche active (tout du moins lors de l'initialisation). Une des approches fréquemment rencontrée dans la littérature est dite de "stereotyping" se fondant sur l'emploi de stéréotypes. Celle-ci repose sur le fait que la création d'un profil initial d'utilisateur peut être vu comme un problème de classification. Cette approche introduite par [Rich, 1979] via son système GRUNDY visait à l'époque la construction de profils utilisateurs sans parler nécessairement de recommandation mais plutôt de personnalisation. Le principe est de construire des profils "stéréotypé"

et d'activer ces derniers en fonction du comportement d'un utilisateur, selon le type de requête qu'il formule. Citons par exemple le stéréotype (traduit de l'anglais) "personne n'utilisant pas de télévision" induisant le fait que cette dernière soit sérieuse, éduquée, etc. selon l'auteur. Ainsi, cette approche est toujours utilisée afin de construire des profils initiaux d'utilisateurs comme dans [Kobsa et al., 2001]. Leurs propositions visent à favoriser les relations entre clients de sites en ligne via une approche collaborative. Les auteurs présentent un large panel d'approches afin de résoudre leurs problématiques dans ce domaine allant de la construction de profils actifs et passifs aux méthodes hybrides, en passant par les méthodes fondées sur le contenu.

Une autre approche permettant la construction de profils à froid se base sur la construction d'une base d'apprentissage, reposant sur un certain nombre d'exemples d'utilisateurs d'autres systèmes. Ainsi, les approches actives de filtrage collaboratif sont souvent employées afin de construire des profils initiaux. Le principe est de proposer à des utilisateurs précis, d'annoter des documents afin de construire leurs profils, et ceci avant le déploiement du système. Ces derniers seront alors utilisés avec le système pour les nouveaux utilisateurs. Notons que le système pourra alors utiliser une méthode passive, la méthode active n'étant utilisée que lors de la phase d'initialisation. Parmi les méthodes utilisant ce type d'initialisation, citons [Schwab et al., 2001] qui s'intéresse à l'apprentissage de profils en se fondant sur un filtrage collaboratif et une analyse du contenu du profil de l'utilisateur. Notons que cette approche peut être apparentée à une méthode hybride, utilisant une méthode de filtrage collaboratif et une méthode basée sur le contenu.

Notons également qu'un certain nombre d'approches n'ont pas été évoquées dans cette section, ne pouvant être exhaustive à la vue du grand nombre d'approches existantes. Citons pour finir les approches novatrices de [Terry, 1993], [Harman, 1994] sur les applications de filtrages collaboratifs, le système MovieLens de recommandation de films, le système Jeter recommandant des farces, et finalement le système FlyCasting recommandant des radio en ligne [Hauver and French, 2001].

3.4 Synthèse

L'avantage des systèmes de recommandation basés sur le filtrage collaboratif est indubitablement la participation active ou passive de l'utilisateur du système. En effet, orienter un utilisateur vers un document semble naturellement plus aisé en connaissant ses préférences. Prenons le cas d'un bibliothécaire. Ce dernier sera à même de conseiller plus efficacement un emprunteur régulier qu'un nouvel arrivant. En effet, le principal inconvénient de ce type de systèmes est le démarrage à froid. Que proposer à un nouvel utilisateur. Comme nous l'avons vu, un certain nombre d'approches permettent de résoudre aujourd'hui cette problématique.

Cependant, il manque la notion de thèmes. Un utilisateur peut par exemple ne s'intéresser qu'à des romans policiers. Mais s'il est jugé proche d'un autre utilisateur aimant également les romans policiers et les romans dramatiques, il pourra se voir proposer des romans dra-

matiques. Cette notion de thèmes peut être résolu par des approches basées sur le contenu, en construisant des classes thématiques. Ainsi, comme nous allons le voir dans la section suivante, les approches prenant en compte tous ces facteurs sont les méthodes dites hybrides, introduisant le collaboratif et le contenu.

4 Les méthodes Hybrides

4.1 Définition

Définies par [Burke, 2002] comme étant une combinaison des méthodes traditionnelles précédemment présentées afin d'en palier les limites, ces dernières sont actuellement les plus représentées dans la littérature, notamment à cause du fait qu'elles soient jugées comme étant les plus efficaces [Schein et al., 2002]. Selon ce dernier, un système hybride est généralement organisé en deux phases :

1. Effectuer de manière indépendante les filtrages des items via des méthodes collaboratives ou par le contenu (ou autres)
2. Combiner ces ensembles de recommandations via des méthodes d'hybridations telles que des pondérations, commutations, cascade, etc.

4.2 Exemple de systèmes de recommandation hybrides

De nombreux types de systèmes hybrides existent dans la littérature, et reposent principalement sur les deux approches présentées précédemment à savoir les méthodes basées sur le contenu et celles fondées sur le filtrage collaboratif. Ainsi, les différentes méthodes produisant des approches par ces systèmes ne seront pas évoquées dans cette section. La principale difficulté d'un système hybride consiste en l'hybridation elle-même. Comment, à partir des connaissances basées sur le contenu et des différents profils d'utilisateurs, pouvons nous obtenir une recommandation efficace ?

Plusieurs approches ont été proposées dont celle de [Claypool et al., 1999] qui proposent d'hybrider les deux méthodologies par l'utilisation d'une combinaison linéaire des deux mesures relatives aux deux types d'approches. Leurs travaux reposent sur ceux de [Vogt et al., 1997] qui a montré qu'une simple combinaison linéaire de scores retournés par différentes approches de recherche d'information pouvait améliorer significativement les résultats. Ainsi, afin de fournir à un utilisateur un document, les auteurs calculent au préalable un score via l'approche collaborative et un score via l'approche basée sur le contenu. La combinaison linéaire des deux scores va déterminer le score final du document pour l'utilisateur. Par ailleurs, [Li and Kim, 2003] s'intéressent également à cette problématique d'hybridation en proposant une approche se fondant sur la construction de classes, en résolvant également le problème du démarrage à froid précédemment évoqué. Le principe de l'approche se décompose en trois points :

1. Utiliser un algorithme de clustering afin de grouper les documents proposer aux utilisateurs.
2. Calculer les distances entre les différents groupes précédemment constitués mais également entre les documents et les groupes, avec une mesure de cosinus améliorée et de Pearson.
3. Fournir une prédiction à un utilisateur en proposant des items proches de son voisinage.

Notons que cette dernière approche ne propose pas tout à fait de méthode se fondant sur une approche collaborative, ne regroupant pas directement des utilisateurs et ne construisant pas de profils.

D'autres approches consiste à proposer deux approches indépendantes de recommandation basées sur le contenu et à base de filtrage collaboratif dont l'hybridation consiste à choisir la meilleure méthode en proposant des mesures de qualité comme dans les travaux de [Tran and Cohen, 2000] qui propose à un utilisateur une recommandation compatible avec ces précédentes évaluations.

Un autre type de combinaison repose sur l'ajout de contenu dans les méthodes à base de filtrage collaboratif. En d'autres termes, la combinaison se fait au niveau du modèle collaboratif auquel on ajoute des ressources acquise via un système basé sur le contenu. Citons par exemple [Soboro and Nicholas, 1999] qui propose de construire un modèle vectoriel contenant les profils utilisateurs se fondant sur le contenu des documents. Dés, ils réduisent la taille de la matrice ainsi produite par la méthode LSI (latent semantic indexing) améliorant les résultats obtenu avec un simple filtrage basé sur le contenu.

Une dernière approche visant à combiner les approches fondées sur le contenu et collaborative visent à) produire un modèle de recommandation unique, c'est-à-dire sans avoir recours à la production de deux modèle devant par la suite être combinés. Citons les travaux de [Basu et al., 2001] qui proposent une méthode à base de règle regroupant dans un même classificateur des notions basées sur le contenu et collaboratives. Notons que la partie collaborative est dans cette approche passive, les auteurs récupérant des informations sur les utilisateur via le Web, sans interaction avec eux.

De même, [Popescul et al., 2001] et [Schein et al., 2002] propose une approche semblable en utilisant un modèle probabiliste afin de combiner les méthodes dans un modèle unique. Les auteurs proposent d'utiliser ici l'analyse latente probabiliste (PLSA).

5 D'autres approches de recommandation

5.1 Hybridation

Outre les approches fondées sur le contenu, filtrages collaboratif ou hybrides, il existe d'autres approches dans la littérature se fondant sur d'autres méthode issues de la fouille

de données, de la recherche d'information ou encore du clustering. Citons par exemple celles se fondant sur des stéréotypes telles qu'évoquées précédemment afin de construire des profils initiaux. Certaines méthodes préconisent l'utilisation de telles approches au delà du processus de démarrage à froid, notamment dans le processus global de recommandation. Ainsi sont proposés des approches hybrides combinant recommandation usuelle et stéréotype comme dans les travaux de [Krulwich, 1997] qui suppose qu'une base initiale d'apprentissage ai été construite en se fondant sur les réponses données par des utilisateurs à un questionnaire.

D'autres approches proposent d'utiliser des données externes comme les pages Web personnelles des utilisateurs comme [Pazzani, 1999]. Ce dernier mesure l'impact de son approche utilisant uniquement les données externes en la comparant à des approches classiques de recommandation. Les approches classiques fournissent de meilleurs résultats mais sa proposition donne cependant de bon résultats montrant ainsi le potentiel d'une hybridation. Une autre approche propose d'utiliser les historiques de navigation des utilisateurs afin d'affiner un profil comme dans les travaux de [Ahmad Wasfi, 1999].

5.2 Les méthodes à bases de connaissances psychologiques

Plus récentes, les méthodes utilisant des connaissances dites psychologiques ont été popularisées par des personnes comme Timo Saari, Gustavo Gonzalez, ou Judith Masthoff depuis une dizaine d'années.

Plusieurs courant existent dans ce domaine.

Les travaux de Masthoff s'intéressent à la "satisfaction d'un utilisateur" dans un processus de recommandation. Ce processus est dynamique et permet de s'adapter à l'humeur des utilisateurs, en introduisant ce modèle dans un système classique de recommandation. Citons par exemple [Masthoff and Gatt, 2006] où elle proposait de sélectionner un certain nombre de programme de télévision en fonction d'un groupe d'utilisateurs. Citons également [Masthoff, 2004] où elle montrait l'influence de l'état affectif d'un utilisateur et son impact dans des systèmes de recommandations.

[Saari et al., 2005] considèrent que les effets psychologiques peuvent être décrits par des états psychologiques comme les émotions, la persuasion, l'attention, la présence, etc. Ces effets sont alors utilisés afin de prédire l'état psychologique de la personne utilisatrice d'un système à un instant précis, influençant les résultats obtenus dans les systèmes de recommandation. Les auteurs appliquent ici leurs approches aux jeux vidéo. Ainsi, en fonction du moment, un utilisateur se voit encouragé à utiliser un jeu plutôt qu'un autre. Finalement, [Gon, 2007] introduisent dans les systèmes de recommandation des notions d'intelligence émotionnelle propres aux utilisateurs. Ainsi, le profil utilisateur est composé d'informations provenant de bases de données sociaux-démographiques et de logs de navigation Web.

6 Conclusion

Cet état de l'art a présenté un certain nombre d'approches visant à produire des systèmes de recommandation. Nous avons ainsi évoqué deux types d'approche et la combinaison des deux telles que listées ci-dessous.

- Les approches basées sur le contenu. Les propositions faites à l'utilisateur sont issues des données, pouvant être des documents, des films, des pages, Web, etc. Ainsi, d'une manière générale, ce type d'approches consiste à construire des classes de documents et à proposer à un utilisateur des items provenant de ces classes, en fonction de la proximité de son profil vis à vis des classes.
- Les approches utilisant un filtrage collaboratif. Le principe est de construire des profils d'utilisateurs en fonction des items qu'ils ont appréciés. Cette méthode repose donc sur la construction des profils pouvant être active ou passive, en fonction de l'intervention explicite ou non de l'utilisateur.
- Les méthodes hybrides. Elles sont la combinaison des deux précédentes. Différentes techniques existent afin de combiner les approches basées sur le contenu et celles reposant sur le filtrage collaboratif comme la combinaison linéaire, les algorithmes de classification, etc.

La tendance actuelle des systèmes de recommandation est plutôt axée sur des méthodes nouvelles, multicritères, multidimensionnelles ou encore se fondant sur des notions psychologiques comme les émotions, les opinions. Notons cependant qu'un système de recommandation doit avant tout s'adapter aux données, celles là mêmes que l'on proposera à un utilisateur. Ainsi, le choix d'une méthode de recommandation doit en premier lieu être dirigé par ce critère.

Prenons l'exemple des données manipulées dans [Béchet N., 2011]. Ces travaux s'inscrivent dans la mise en place d'un système de recommandation pour le site Web **Addictrip**. Le choix d'un algorithme hybride de recommandation peut sembler plus adapté, mais est dans ce cas pénalisé par le manque actuel d'informations relatives aux utilisateurs afin de construire des profils. Ainsi, nous privilégierons dans ce cas l'approche basée sur le contenu.

References

- [Gon, 2007] (2007). Embedding Emotional Context in Recommender Systems.
- [Adomavicius et al., 2005] Adomavicius, G., Sankaranarayanan, R., Sen, S. and Tuzhilin, A. (2005). Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach. *ACM Transactions on Information Systems* 23, 103–145.
- [Ahmad Wasfi, 1999] Ahmad Wasfi, A. M. (1999). Collecting user access patterns for building user profiles and collaborative filtering. In *Proceedings of the 4th international conference on Intelligent user interfaces IUI '99* pp. 57–64, ACM, New York, NY, USA.
- [Basu et al., 2001] Basu, C., Hirsh, H., Cohen, W. and Manning, N. C. (2001). Technical paper recommendation: A study in combining multiple information sources.
- [Béchet, 2009] Béchet, N. (2009). Extraction et regroupement de descripteurs morpho-syntaxiques pour des processus de Fouille de Textes. PhD thesis,.
- [Béchet N., 2011] Béchet N., Aufaure M.-A., L. Y. (2011). Construction et peuplement de structures hiérarchiques de concepts dans le domaine du e-tourisme. In *IC'2011 (Ingénierie des connaissances)*. A paraître.
- [Belkin and Croft, 1992] Belkin, N. J. and Croft, W. B. (1992). Information filtering and information retrieval: two sides of the same coin? *Commun. ACM* 35, 29–38.
- [Burke, 2002] Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction* 12, 331–370.
- [Chen and Sycara, 1998] Chen, L. and Sycara, K. (1998). WebMate: a personal agent for browsing and searching. In *Proceedings of the second international conference on Autonomous agents AGENTS '98* pp. 132–139, ACM, New York, NY, USA.
- [Claypool et al., 1999] Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D. and Sartin, M. (1999). Combining Content-Based and Collaborative Filters in an Online Newspaper.
- [Cunningham et al., 2000] Cunningham, P., Bergmann, R., Schmitt, S., Traphöner, R., Breen, S. and Smyth, B. (2000). WEBSELL: Intelligent Sales Assistants for the World Wide Web. In *Procs. of the Workshop Programme at the Fourth International Conference on Case-Based Reasoning*.
- [Duda and Hart, 1973] Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. John Wiley & Sons Inc.
- [Fischer and Stevens, 1991] Fischer, G. and Stevens, C. (1991). Information Access in Complex, Poorly Structured Information Spaces. pp. 63–70, ACM Press.
- [Goldberg et al., 1992a] Goldberg, D., Nichols, D., Oki, B. M. and Terry, D. (1992a). Using collaborative filtering to weave an information tapestry. *Commun. ACM* 35, 61–70.

- [Goldberg et al., 1992b] Goldberg, D., Nichols, D., Oki, B. M. and Terry, D. (1992b). Using collaborative filtering to weave an information tapestry. *Communications of the ACM* 35, 61–70.
- [Harman, 1994] Harman, D. (1994). Overview of the Third Text REtrieval Conference (TREC-3). In *TREC'94*.
- [Hauver and French, 2001] Hauver, D. B. and French, J. C. (2001). Flycasting: Using Collaborative Filtering to Generate a Playlist for Online Radio. In *Proc. Int. Conf. Web Delivering of Music* pp. 123–130,.
- [Kent and of Pittsburgh., 1979] Kent, A. and of Pittsburgh., U. (1979). Use of library materials : the University of Pittsburgh study / Allen Kent ... [et al.]. M. Dekker, New York :.
- [Kobsa et al., 2001] Kobsa, A., Koenemann, J. and Pohl, W. (2001). Personalised hypermedia presentation techniques for improving online customer relationships. *Knowl. Eng. Rev.* 16, 111–155.
- [Krulwich, 1997] Krulwich, B. (1997). LIFESTYLE FINDER: Intelligent User Profiling Using Large-Scale Demographic Data. *AI Magazine* 18, 37–45.
- [Li and Kim, 2003] Li, Q. and Kim, B. M. (2003). An approach for combining content-based and collaborative filters. In *Proceedings of the sixth international workshop on Information retrieval with Asian languages - Volume 11 AsianIR '03* pp. 17–24, Association for Computational Linguistics, Stroudsburg, PA, USA.
- [Lieberman, 1995] Lieberman, H. (1995). Letizia: An Agent That Assists Web Browsing. In *INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE* pp. 924–929,.
- [Linden et al., 2003] Linden, G., Smith, B. and York, J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing* 7, 76–80.
- [Maltz and Ehrlich, 1995] Maltz, D. and Ehrlich, K. (1995). Pointing the Way: Active Collaborative Filtering. In *CHI* pp. 202–209,.
- [Masthoff, 2004] Masthoff, J. (2004). Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. *User Modeling and User-Adapted Interaction* 14, 37–85.
- [Masthoff and Gatt, 2006] Masthoff, J. and Gatt, A. (2006). In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. *User Modeling and User-Adapted Interaction* 16, 281–319.
- [Miller et al., 1990] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography* 3, 235–244.
- [Mooney and Roy, 1999] Mooney, R. J. and Roy, L. (1999). Content-Based Book Recommending Using Learning for Text Categorization. In *In Proceedings of the Fifth ACM Conference on Digital Libraries* pp. 195–204, ACM Press.

- [Nichols, 1998] Nichols, D. (1998). Implicit rating and filtering. In Proceedings of 5th DELOS Workshop on Filtering and Collaborative Filtering pp. 31–36, ERCIM.
- [Pazzani et al., 1997] Pazzani, M., Billsus, D., Michalski, S. and Wnek, J. (1997). Learning and Revising User Profiles: The Identification of Interesting Web Sites. In Machine Learning pp. 313–331,.
- [Pazzani, 1999] Pazzani, M. J. (1999). A Framework for Collaborative, Content-Based and Demographic Filtering. *ARTIFICIAL INTELLIGENCE REVIEW* 13, 393–408.
- [Popescul et al., 2001] Popescul, A., Ungar, L. H., Pennock, D. M. and Lawrence, S. (2001). Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments. In Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence UAI '01 pp. 437–444, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [Resnick et al., 1994] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. pp. 175–186, ACM Press.
- [Resnick and Varian, 1997] Resnick, P. and Varian, H. R. (1997). Recommender systems. *Commun. ACM* 40, 56–58.
- [Rewari et al., 1992] Rewari, A., Adler, M., Anick, P. G., Billmers, M. A., Carifio, M., Gunderson, A., Pundit, N. and Swartout, M. W. (1992). AI Research and Applications in Digital's Service Organization. *AI Magazine* 13, 63–75.
- [Rich, 1979] Rich, E. (1979). User modeling via stereotypes. *Cognitive Science* 3, 329–354.
- [Saari et al., 2005] Saari, T., Ravaja, N., Laarni, J. and Turpeinen, M. (2005). Towards Emotionally Adapted Games based on User Controlled Emotion Knobs. In DIGRA Conf.
- [Salton and Buckley, 1988] Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. In *INFORMATION PROCESSING AND MANAGEMENT* pp. 513–523,.
- [Salton and Lesk, 1965] Salton, G. and Lesk, M. E. (1965). The SMART automatic document retrieval systems an illustration. *Communications of the ACM* 8, 391–398.
- [Salton et al., 1975a] Salton, G., Wong, A. and Yang, C. S. (1975a). A Vector Space Model for Automatic Indexing. *Commun. ACM* 18, 613–620.
- [Salton et al., 1975b] Salton, G., Wong, A. and Yang, C. S. (1975b). A Vector Space Model for Automatic Indexing. *Commun. ACM* 18, 613–620.
- [Schein et al., 2002] Schein, A. I., Popescul, A., H., L., Popescul, R., Ungar, L. H. and Pennock, D. M. (2002). Methods and Metrics for Cold-Start Recommendations. In In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval pp. 253–260, ACM Press.

- [Schwab et al., 2001] Schwab, I., Kobsa, A. and Koychev, I. (2001). Learning User Interests through Positive Examples Using Content Analysis and Collaborative Filtering. In 30 2001. Internal Memo, GMD.
- [Shardanand and Maes, 1995] Shardanand, U. and Maes, P. (1995). Social Information Filtering: Algorithms for Automating "Word of Mouth". pp. 210–217, ACM Press.
- [Sheth and Maes, 1993] Sheth, B. and Maes, P. (1993). Evolving agents for personalized information filtering. pp. 345–352,.
- [Soboro and Nicholas, 1999] Soboro, I. M. and Nicholas, C. K. (1999). Combining content and collaboration in text filtering. In Proceedings of the IJCAI vol. 99, pp. 86–91,.
- [Terry, 1993] Terry, D. B. (1993). A tour through Tapestry. In Proceedings of the conference on Organizational computing systems COCS '93 pp. 21–30, ACM, New York, NY, USA.
- [Tran and Cohen, 2000] Tran, T. and Cohen, R. (2000). Hybrid Recommender Systems for Electronic Commerce'. In Knowledge-Based Electronic Markets.
- [van Rijsbergen and D, 1979] van Rijsbergen, C. and D, P. (1979). Information Retrieval.
- [Vogt et al., 1997] Vogt, C. C., Cottrell, G. W., Belew, R. K. and Bartell, B. T. (1997). Using Relevance to Train a Linear Mixture of Experts. In In [Harman pp. 503–515,.
- [Zhang and Callan, 2002] Zhang, Y. and Callan, J. (2002). Novelty and redundancy detection in adaptive filtering. pp. 81–88, ACM Press.