

Système de recommandation basé sur la taxonomie et les feedbacks implicites dans un contexte Business-to-Business

THAMER SARA EI¹, MAHA BENALI¹, JEAN-MARC FRAYRET¹

¹ Département de mathématiques et de génie industriel, Polytechnique Montréal
2500 Chem. de Polytechnique, Montréal, QC H3T 1J4, Canada
thamer.saraei@polymtl.ca, maha.benali@polymtl.ca, jean-marc.frayret@polymtl.ca

Résumé

Dans le domaine des systèmes de recommandation, le filtrage collaboratif est une technique couramment employée qui suggère des produits aux utilisateurs actifs sur la base des évaluations antérieures faites par des utilisateurs similaires (appelés utilisateurs voisins). Cependant, cette approche performe moins bien si la sparsité est élevée. Ceci est généralement le cas dans un contexte B2B où peu d'évaluations sont générées par les utilisateurs, ce qui entraîne une diminution de la précision dans la prédiction des préférences des utilisateurs, et génère donc des recommandations de moindre qualité. Pour résoudre ce problème, nous proposons une approche innovante qui intègre l'utilisation de la taxonomie des produits et l'extraction de règles d'association. Notre approche implique la génération d'une nouvelle taxonomie explicite basée uniquement sur les descriptions textuelles des produits, suivie de l'amélioration de la matrice utilisateur-produit grâce à l'utilisation des résultats de l'extraction de règles d'association. Nous incorporons ensuite des produits complémentaires pour les utilisateurs en fonction de leurs habitudes d'achat de produits de base selon les règles d'association extraites. Finalement, nous utilisons la méthode des moindres carrés alternatifs implicites (i-ALS) et initialisons les matrices de facteurs latents avec des valeurs obtenues par décomposition en valeurs singulières (BLS-SVD++). Pour valider l'approche que nous proposons, le cas d'un distributeur québécois a été considéré. Les résultats obtenus à partir des commentaires déduits implicitement des données d'achat ont indiqué une amélioration des performances par rapport aux approches conventionnelles.

Abstract

In the field of recommender systems, collaborative filtering is a commonly used technique that suggests products to active users based on previous reviews by similar users (called neighboring users). However, this approach performs less well if sparsity is high. This is usually the case in a B2B context where few reviews are generated by users, which leads to a decrease in the accuracy in predicting users' preferences, and thus generates lower quality recommendations. To address this problem, we propose an innovative approach that integrates the use of product taxonomy and association rule mining. Our approach involves the generation of a new explicit taxonomy based solely on textual product descriptions, followed by the improvement of the user-product matrix using the results of association rule mining. We then incorporate complementary products for users based on their basic product buying habits according to the extracted association rules. Finally, we use the implicit alternative least squares (i-ALS) method and initialize the latent factor matrices with values obtained by singular value decomposition (BLS-SVD++). To validate our proposed approach, the case from a Quebec distributor has been considered. The results obtained from the feedbacks implicitly deduced from the purchase data indicated an improvement of the performances compared to the conventional approaches.

Mots clés - Filtrage collaboratif, Sparsité, Extraction de règles d'association, Taxonomie, Feedback implicite.

Keywords – Collaborative Filtering, Sparsity, Association Rules Mining, Taxonomy, Implicit Feedback.

1 INTRODUCTION

Les sites de commerce électronique offrent une grande variété de produits à leurs clients, ce qui rend le choix difficile pour les utilisateurs. Les systèmes de recommandation (SR) aident à résoudre ce problème en filtrant les produits et en présentant seulement ceux qui sont susceptibles d'intéresser l'utilisateur. L'une des pratiques de recommandation les plus populaires est la recommandation Top-N, dans laquelle le système propose une liste de N produits qui sont les plus susceptibles d'attirer l'attention de l'utilisateur [Nia et al., 2019]. Le système de recommandation Top-N génère les N éléments qui sont les plus susceptibles d'intéresser l'utilisateur actuel u.

Les SR peuvent être classés en deux catégories en fonction du type de données utilisées pour faire des recommandations [Herlocker et al., 2004] : le filtrage basé sur le contenu (FBC)

et le filtrage collaboratif (FC). Les FBCs exploitent les métadonnées relatives aux utilisateurs, telles que les données démographiques, ainsi que les métadonnées associées aux produits (ex., le genre du film), ou autres caractéristiques pertinentes pour prédire les préférences des utilisateurs en fonction de ces attributs. Cette technique de recommandation est conçue pour établir des profils utilisateur à partir des données disponibles et ainsi proposer des recommandations personnalisées qui répondent aux goûts et aux attentes des utilisateurs. Les FBCs sont largement utilisés dans les domaines de l'e-commerce et des médias, où les systèmes ont démontré leur efficacité pour améliorer l'expérience utilisateur et renforcer la fidélisation de la clientèle.

Le FC est une des techniques les plus utilisées dans les SR [Park et al., 2015]. Elle se base uniquement sur l'historique

d'achat des utilisateurs et leurs interactions passées avec les produits. L'approche FC est confrontée à plusieurs défis, tels que le problème de sparsité qui se pose en raison du nombre limité de produits avec lesquels les utilisateurs interagissent [Abbas et al., 2015]. Les données utilisées dans l'approche FC correspondent aux commentaires donnés par les utilisateurs sur les produits peuvent être explicites ou implicites. Nous parlons de feedback explicite lorsque l'utilisateur donne directement son avis, le plus souvent par le biais d'un système de notation [Liu et al., 2010]. Les commentaires implicites, quant à eux, reflètent indirectement l'intérêt de l'utilisateur pour un produit, comme l'historique de navigation ou d'achat [Ricci et al., 2010].

Un autre problème auquel sont confrontés les SR basés sur le FC est le problème de synonymie, lorsque des produits similaires ont des noms ou des descriptions différents, ce qui entraîne l'échec de la découverte de l'association latente [Liphoto et al., 2016]. La plupart des systèmes de FC ont du mal à distinguer les éléments similaires, ce qui rend difficile le calcul de leur similarité.

Dans cet article, nous nous concentrons sur les SR pour des clients B2B. Dans un tel contexte, généralement on utilise les données implicites disponibles en grands volumes à partir de l'historique des interactions des utilisateurs avec divers produits. Pour surmonter les problèmes de sparsité et de synonymie, nous proposons une approche de FC basée sur la taxonomie de produits et une technique de factorisation matricielle adaptée aux retours implicites, appelée Implicit Alternating Least Squaring (i-ALS) [Koren, 2008]. Nous commençons par créer une matrice utilisateur-produit plus dense en agrégeant les préférences des utilisateurs à l'aide d'une taxonomie de produits générée automatiquement et en améliorant la matrice à l'aide de la technique d'extraction de règles d'association [Sharma & Mann]. Ensuite, nous utilisons l'algorithme i-ALS pour générer les matrices de facteurs latents. Afin d'améliorer les performances, nous avons recours à une nouvelle stratégie d'initialisation basée sur BLS-SVD++ [Wang et al., 2020].

La section 2 présente les travaux connexes. La méthode proposée et ses composants clés sont détaillés dans la section 3. Les résultats et la comparaison avec d'autres approches conventionnelles sont discutés dans la section 4. Enfin, la section 5 conclut l'article et suggère des directions pour les travaux futurs.

2 REVUE DE LA LITTÉRATURE

2.1. UTILISATION DE LA TAXONOMIE DE PRODUITS

La littérature a largement étudié l'utilisation de taxonomies dans les systèmes de recommandation. Maidel et al. [2008] ont proposé une méthode basée sur la taxonomie pour la classification de la pertinence dans le domaine du papier électronique. Ziegler [2004] a utilisé une taxonomie hiérarchique de produits pour améliorer la précision des recommandations de produits. Dans un travail plus récent [Gicquel & Lenne, 2013], les auteurs ont recommandé des visites de musées en utilisant ICONCLASS, une classification complète des thèmes de l'art occidental. Mnih et al. [2012] ont développé des modèles de facteurs latents basés sur la taxonomie pour les recommandations musicales.

Les modèles hiérarchiques bayésiens ont également été utilisés pour apprendre les préférences des utilisateurs pour des

attributs de produits. Par exemple, Ahmed et al. [2013] ont proposé un modèle hiérarchique bayésien personnalisé pour apprendre les préférences en relation avec la marque et le prix. Lawrence et al. [2001] ont utilisé une taxonomie de produits pour représenter l'affinité entre divers produits dans un système de recommandation de supermarché. Kanagal et al. [2012] ont introduit des facteurs pour chaque nœud interne de la taxonomie et ont appliqué des priors de taxonomie sur les facteurs latents des utilisateurs et de produits.

Cependant, ces méthodes reposent sur des taxonomies prédéfinies, ce qui limite leur efficacité. Pour résoudre cette limitation, Zhang et al. [2014] ont proposé un modèle paramétrique qui découvre automatiquement la structure de la taxonomie sur les produits et l'utilise pour générer les préférences des utilisateurs. Le modèle a obtenu des performances supérieures à celles des approches conventionnelles sur deux ensembles de données réelles.

2.2. EXTRACTION DE REGLES D'ASSOCIATION

La technique d'extraction de règles d'association (Association Rules Mining *ARM*) a été appliquée pour représenter les intérêts des utilisateurs dans divers domaines afin de fournir des modèles de recommandation puissants. Cela est dû à sa capacité à s'adapter à de grands ensembles de données et à atteindre une précision élevée [Tyagi & Bharadwaj, 2013]. Cette méthode sert à décrire les données et non à prédire leur évolution. Dans le cadre de la recherche en matière de systèmes de recommandation, plusieurs études ont exploré l'utilisation d'ARM pour améliorer les performances de ces systèmes. Par exemple, Leung et al. [2011] ont présenté un modèle basé sur des règles d'association floues qui exploitent les similarités entre les produits dans les taxonomies de produits existantes pour améliorer les recommandations. Kardan et Ebrahimi [2013] ont quant à eux utilisé la technique ARM dans un système de recommandation hybride pour calculer la similarité entre les utilisateurs à partir de données implicites collectées dans un groupe de discussion. Dans une étude plus récente, Guo et al. [2017] ont proposé une méthode d'extraction du comportement d'achat des clients en utilisant des règles d'association séquentielles pour tenir compte de l'évolution du comportement d'achat des clients dans le temps. Cette approche permet de mieux comprendre le comportement d'achat des clients et de proposer des recommandations plus pertinentes. Dans une autre étude, Feng et al. [2015] ont proposé une nouvelle technique de recommandation en développant une méthode de détection de communauté à chevauchement temporel basée sur une règle d'association pondérée dans le temps. Cette approche novatrice permet de mieux tenir compte de l'évolution des comportements d'achat des utilisateurs dans le temps, en offrant ainsi des recommandations plus précises et plus adaptées à leurs besoins.

Il est important de noter que, dans les études précédentes, les règles d'association générées n'ont pas été utilisées directement pour améliorer la matrice utilisateur-produit, mais plutôt dans une approche hybride pour fournir des recommandations plus puissantes. Dans cet article, nous proposons d'aller plus loin en améliorant directement la matrice utilisateur-élément en exploitant les résultats des règles d'association. En effet, sur la base de ces règles, nous pouvons compléter les produits des utilisateurs en fonction de leurs achats de produits de base des règles, ce qui peut permettre de proposer des recommandations encore plus pertinentes et personnalisées.

2.3. INITIALISATION DE LA METHODE DE LA FACTORISATION MATRICIELLE

Selon Smilde et al. [2005] une bonne initialisation peut améliorer la vitesse et la précision des algorithmes de factorisation matricielle en produisant une convergence plus rapide vers un minimum local. Dans les travaux [Johnson, 2014; Koren, 2008], l'approche de la factorisation matricielle implique une initialisation aléatoire des matrices de facteurs latents et la mise en œuvre d'une méthode d'optimisation répétée sur les matrices. À chaque itération, les matrices sont mises à jour en minimisant l'erreur d'une fonction de coût. Après quelques itérations, ces méthodes convergent vers un optimum local qui dépend du point de départ (l'initialisation).

Un bon choix des valeurs initiales affecte non seulement la propriété de convergence de l'algorithme mais en particulier la qualité de la solution vers laquelle il converge [Donoho & Grimes, 2003]. Dans notre travail, nous proposons une autre technique de décomposition plus avancée qui s'appelle Recherche Linéaire par Retour arrière (appelé aussi Backtracking Line Search SVD++ ou BLS-SVD++ en anglais) [Wang et al., 2020]. La motivation pour l'utilisation d'une initialisation basée sur la variante BLS-SVD++ de la technique SVD est que Wang S et al. (2020) montrent avec des expériences sur les ensembles de données réelles que l'algorithme BLS-SVD++ proposé surpasse le SVD et d'autres algorithmes de recommandation en termes de performance et de temps de convergence.

3 MÉTHODOLOGIE

La méthodologie proposée vise à résoudre deux problèmes majeurs associés au filtrage collaboratif : le problème de synonymie et le problème de sparsité. Les techniques de filtrage

collaboratif ont été

utilisées pour recommander des produits aux utilisateurs en fonction de leurs préférences, mais deux des principaux défis dans ce domaine sont la synonymie et la sparsité.

Pour résoudre le problème de synonymie, nous utilisons une taxonomie de produits générée automatiquement et agrège les préférences des utilisateurs en fonction de cette nouvelle taxonomie. Ainsi, nous pouvons nous assurer que les éléments de nature similaire sont regroupés, ce qui peut conduire à des recommandations plus précises.

Pour résoudre le problème de sparsité, nous utilisons la technique d'extraction des règles d'association [Sharma & Mann] pour construire une matrice utilisateur-produit plus dense. Les règles d'association (ARs) sont extraites sur la base de l'algorithme À priori, puis les valeurs manquantes de la matrice utilisateur-produit sont remplies en ajoutant des produits complémentaires aux utilisateurs. Cette approche permet de s'assurer que la matrice utilisateur-élément est aussi dense que possible, ce qui conduit à des recommandations plus précises.

Enfin, nous proposons d'utiliser une stratégie d'initialisation basée sur l'algorithme de décomposition de la valeur singulière appelé BLS-SVD++.

Pour résumer, comme le montre Figure 1, la méthodologie proposée se compose de trois étapes. Dans la première étape, une nouvelle taxonomie de produits est générée et utilisée pour générer les préférences des utilisateurs. Dans la deuxième étape, l'exploration de règles d'association est utilisée pour fournir de produits complémentaires aux utilisateurs et pour améliorer la matrice utilisateur-produit. Dans la dernière étape, une initialisation de l'algorithme i-ALS est effectuée en utilisant l'algorithme BLS-SVD++ pour améliorer la précision.

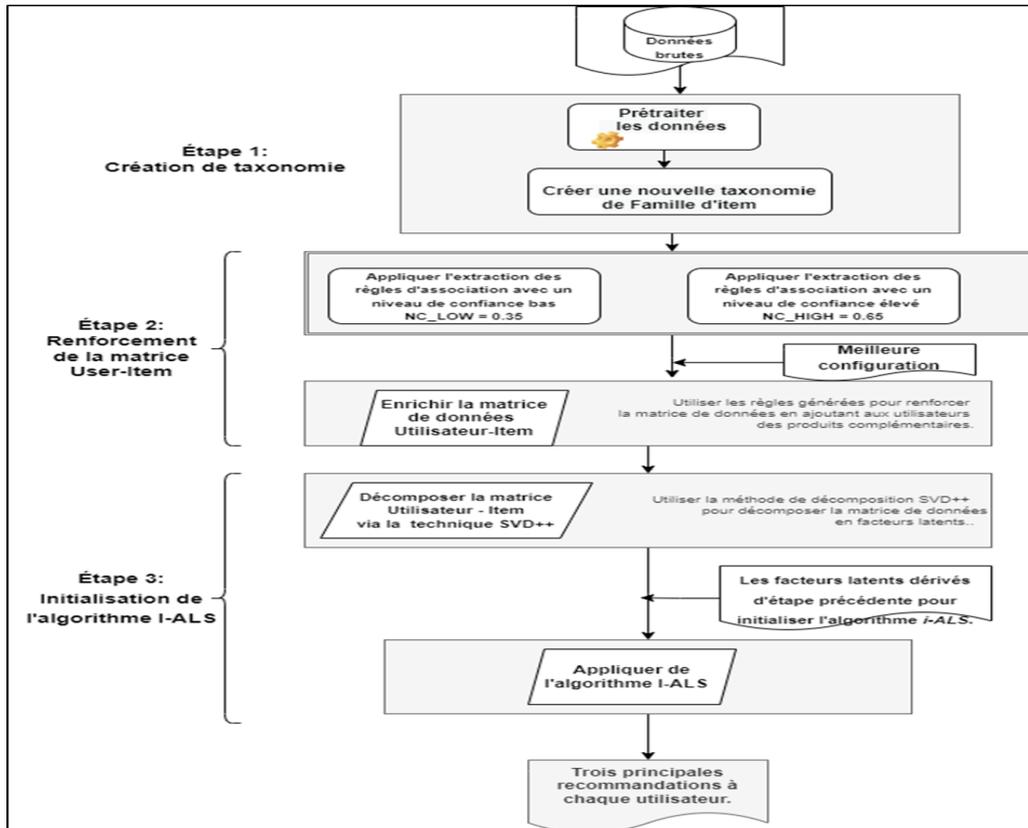


Figure 1. Étapes de la méthodologie

3.1 Création d'une nouvelle taxonomie de produits

Le processus de création d'une nouvelle taxonomie de produits consiste à identifier et à regrouper les produits similaires en familles afin d'améliorer la précision des recommandations du SR à filtrage collaboratif (FC). À cette fin, nous avons procédé comme suit : premièrement, une analyse des caractéristiques des produits est effectuée, qui comprend un examen des descriptions textuelles afin d'identifier les similitudes et les différences entre les produits. Deuxièmement, sur la base de cette analyse, les similitudes et les différences sont identifiées en examinant les descriptions, afin de déterminer quels éléments sont de nature similaire. Troisièmement, une fois les similitudes identifiées, des règles sont créées pour regrouper les produits en familles en fonction de leurs similitudes. Ces règles sont basées sur les descriptions textuelles des produits. Quatrièmement, les produits sont affectés à la famille appropriée en fonction des règles créées à l'étape précédente, les produits ayant des caractéristiques similaires étant regroupés dans la même famille. Enfin, la taxonomie qui en résulte est exploitée par les algorithmes de FC pour améliorer la précision des recommandations. En prenant en compte les produits qui sont de nature similaire mais qui peuvent avoir des identifiants différents, le FC peut identifier les synonymes et fournir des recommandations plus pertinentes aux utilisateurs.

Notons que nous avons développé un outil permettant d'extraire automatiquement les différentes taxonomies à partir de descriptions des produits en utilisant deux modules de traitement nommés Fragmentation du texte (*Text Splitting*) et Regroupement des phrases (*Sentence Clustering*). Le premier divise chaque texte d'une description en mots. Le second module identifie et regroupe les phrases similaires. Le résultat est un ensemble de clusters qui regroupe les textes similaires.

3.2 Renforcement de la matrice utilisateur-produit

La préparation de la matrice utilisateur-produit représente une étape critique dans la mise en place du filtrage collaboratif. Cette matrice est généralement construite en se basant sur les données d'achat historiques des clients, qui fournissent des feedbacks implicites sur leurs préférences en matière de produits. La matrice ainsi obtenue est représentative de la fréquence d'achat des utilisateurs pour un produit spécifique, permettant ainsi l'établissement de liens entre les préférences des utilisateurs et les produits qu'ils achètent. En résumé, cette étape implique la transformation des données brutes en une matrice, qui est utilisée pour générer des recommandations personnalisées pour les utilisateurs, en se basant sur les fréquences d'achat.

Afin de renforcer la matrice utilisateur-produit, nous proposons de remplir les facteurs associés à un utilisateur particulier en prenant en compte les produits que l'utilisateur a achetés et leurs produits complémentaires qui n'ont pas été achetés. Plus précisément, en utilisant une règle $A \rightarrow B$ avec un niveau de confiance nc , où A et B représentent respectivement un produit de base et un produit complémentaire, nous pouvons déterminer la fréquence d'apparition du produit B dans les transactions où le produit A apparaît. Cette information peut ensuite être appliquée à un utilisateur spécifique u qui a acheté le produit A n_{uiA} fois mais n'a pas acheté le produit B . Dans ce cas, nous attribuons à l'utilisateur un score spécifique n_{uiB} , qui est calculé comme la fréquence d'achat du produit de base A multiplié par le niveau de confiance de la règle nc [Schoinas & Tjortjais, 2019]. Cette méthode prend en compte les relations entre les produits de

base et les produits complémentaires, ainsi que les habitudes d'achat des utilisateurs, afin d'améliorer la précision de la matrice utilisateur-produit. En remplissant les facteurs manquants avec des produits complémentaires qui ont une forte association avec les produits de base des règles générées, nous pouvons fournir des recommandations plus pertinentes aux utilisateurs.

3.3 Initialisation de l'algorithme i-ALS

Les techniques d'initialisation conventionnelles sont basées sur des méthodes purement aléatoires, où les matrices sont initialisées avec des nombres aléatoires uniformément distribués, par exemple entre 0 et 1. Ce type d'initialisation ne fournit pas toujours une bonne solution. Une initialisation appropriée de ces vecteurs permettrait aux méthodes de factorisation matricielle de produire des vecteurs plus précis et donc de donner des prédictions plus précises. Pour cela, nous utilisons l'algorithme de décomposition en valeurs singulières BLS-SVD++ [Wang et al., 2020] dans le but de décomposer la matrice utilisateur-produit et extraire les vecteurs latents des facteurs utilisateur et produit qui sont ensuite utilisés comme valeurs initiales pour l'algorithme i-ALS au lieu de commencer avec des valeurs initiales aléatoires.

4 EXPERIMENTATION

4.1 Données

Dans cet article, nous avons réalisé une étude de cas réelle, en utilisant les jeux de données d'un distributeur québécois. La base de données contient plus de 5 000 clients industriels, plus de 15 000 produits, et l'historique des transactions entre 2007 et 2020. Dans cette étude, nous nous concentrons sur le nombre de fois qu'un client a acheté chaque produit.

Nous proposons de diviser l'ensemble de données en un ensemble d'apprentissage et un ensemble de test à l'aide d'une fonction personnalisée appelée *python_stratified_split*, où le fractionnement est stratifié de manière que le même groupe d'utilisateurs et de produits apparaisse dans les ensembles d'apprentissage et de test.

4.2 Plan d'expérience

Trois séries d'expériences seront à réaliser pour évaluer et comparer différentes configurations. La première série vise à évaluer si l'algorithme i-ALS donne de meilleurs résultats en considérant des différents niveaux de taxonomies. La deuxième série d'expérience sera dédiée à explorer si la performance de l'algorithme i-ALS avec la matrice sélectionnée dans la série précédente pourrait être améliorée par l'utilisation de règles d'association ou/et par l'initialisation l'algorithme i-ALS. Finalement, l'approche retenue sera comparée à des approches conventionnelles en SR, soit :

- **Aléatoire** : Tout simplement, une liste aléatoire de N produits est recommandée à l'utilisateur.
- **Le plus populaire** : C'est un algorithme plus compétitif que l'algorithme aléatoire. Il renvoie tout simplement à l'utilisateur les N produits les plus populaires. Il s'agit d'un point de référence couramment utilisé pour la tâche de recommandation des produits basée sur le feedback implicite.
- **Bayesian Personalized Ranking** [Rendle & Schmidt-Thieme, 2008]: Nous parlons d'un algorithme de recommandation spécialement conçu pour le feedback implicite. Cette technique optimise le modèle de factorisation matricielle avec une technique bayésienne de perte de

classement par paires (pairwise ranking loss). Il utilise la descente de gradient stochastique pour apprendre les préférences des utilisateurs à partir des données de feedback implicite.

- **NeuMF** [He et al., 2017] : Cette méthode est un modèle de factorisation matricielle avec une architecture de réseau neuronal. Nous utilisons l'implémentation originale qui est déjà destinée au problème de recommandation avec les feedbacks implicite.

5 RESULTATS ET DISCUSSION

Pour comparer la performance de différentes configurations du SR, nous utilisons la précision à 3 ($P@3$). Cette métrique estime à quel point le SR recommande de « bons » produits (soit des produits susceptibles d'être achetés par l'utilisateur) dans le top 3 des items recommandés. Les résultats présentés sont pour toutes les approches et modèles testés représentent les $P@3$ des prédictions sur l'ensemble du jeu de données du test.

5.1 Comparer la performance de l'algorithme i-ALS selon les trois niveaux de taxonomies

Partant de la matrice utilisateur-produit, nous créons les matrices suivantes :

- Matrice **ITEM_FAMILY** considérant la taxonomie « famille de produits » ;
- Matrice **ITEM_SS_CATG** considérant la taxonomie « sous-catégorie du produits » ;
- Matrice **ITEM_ID** considérant les « IDentifiants des produits ».

En comparant les résultats obtenus en appliquant l'algorithme i-ALS sur les trois matrices (voir Tableau 1), il apparaît que l'algorithme i-ALS est plus performant avec la matrice **ITEM_FAMILY**. Par rapport à la matrice standard (**ITEM_ID**), les deux matrices basées sur des niveaux de taxonomie, **ITEM_FAMILY** et **ITEM_SS_CATG**, sont respectivement 23,25% et 14,34% plus performantes. La raison pour laquelle ces deux matrices sont plus performantes n'est pas exclusivement due au fait qu'elles constituent un meilleur modèle. La principale différence entre les trois matrices est le niveau de sparsité des données. En effet, la matrice **ITEM_FAMILY** a un meilleur degré de sparsité (99.34%) que les deux autres matrices (98.70% pour **ITEM_ID** et 91.18% pour **ITEM_SS_CATG**).

Tableau 1. Comparaison la performance de l'algorithme i-ALS en fonction de différents niveaux de taxonomie

	P@3
ALS_ITEM_ID	0.238
ALS_ITEM_SS_CATG	0.295
ALS_ITEM_FAMILY	0.318

Vu que la matrice **ITEM_FAMILY** a la meilleure performance, nous l'utilisons pour la suite des expériences.

5.2 Évaluer l'effet d'introduire l'ARM sur la performance de l'algorithme

Dans cette section, nous évaluons l'effet d'introduire l'ARM sur la performance de l'algorithme i-ALS. Pour cela, nous construisons les algorithmes suivants :

- Algorithme testé avec la **Matrice ITEM_FAMILY Standard** : **MFS**

- Algorithme testé avec la **Matrice ITEM_FAMILY Renforcé** par les règles d'association avec un niveau de confiance faible $NC_LOW = 0.35$: **MFRL**
- Algorithme testé avec la **Matrice ITEM_FAMILY Renforcé** par les règles d'association avec un niveau de confiance élevé $NC_HIGH = 0.65$: **MFRH**

Les résultats sont présentés dans Tableau 2. Les deux configurations avec renforcement (MFRL et MFRH) de la matrice **ITEM_FAMILY** performe mieux que la configuration standard (MFS). Ce résultat s'explique par le fait que l'ajout des produits à la liste des utilisateurs sur la base de règles ayant un bon niveau de confiance diminue significativement la sparsité de la matrice standard. En particulier, le renforcement avec l'ARM configuré avec un faible niveau de confiance a fourni la meilleure précision. Ceci pourrait être expliqué par le fait qu'en utilisant un faible niveau de confiance, nous avons généré plus de règles et donc nous avons plus renforcé la matrice, ce qui permet d'améliorer la sparsité.

Tableau 2. Comparaison de l'effet du niveau de confiance sur la performance de l'algorithme i-ALS

	P@3
MFS	0.318
MFRL	0.333
MFRH	0.324

5.3 Évaluer l'effet de l'initialisation de l'algorithme i-ALS

Dans cette section, nous évaluons l'effet de l'initialisation de l'algorithme i-ALS avec le résultat de la technique de décomposition BLS-SVD++. Pour ce faire, nous comparons les quatre configurations suivantes :

- Configuration Standard : consiste à appliquer l'algorithme i-ALS avec la matrice **ITEM_FAMILY**. L'initialisation dans ce cas sera aléatoire.
- Configuration Améliorée par l'ARM : consiste à appliquer l'algorithme i-ALS avec la matrice **ITEM_FAMILY** renforcée par les règles d'association ;
- Configuration Initialisée par BLS-SVD++ : consiste à appliquer l'algorithme i-ALS initialisée par les résultats de l'algorithme BLS-SVD++ en utilisant la matrice **ITEM_FAMILY** ;
- Configuration Améliorée & Initialisé : cette configuration est la combinaison de deux précédentes méthodes.

Comme l'illustre Tableau 3, toutes les méthodes appliquées pour augmenter la précision du SR l'ont considérablement amélioré. Ainsi, la combinaison des deux méthodes d'amélioration a réalisé la plus grande amélioration comparée à la configuration standard, soit 12.47% d'amélioration de la précision $P@3$.

Tableau 3 : Évaluation de l'effet de l'initialisation

Configurations	P@3
Configuration Standard	0.318
Configuration Améliorée par l'ARM	0.333
Configuration Initialisée par BLS-SVD++	0.342
Configuration Améliorée & Initialisé	0.358

Pour la suite, nous retenons l'approche de la configuration Améliorée & Initialisé qui a mieux performé. Cette approche consiste donc à utiliser la matrice « ITEM FAMILY » automatiquement créée, puis renforcée avec l'ARM, et à appliquer l'algorithme i-ALS initialisé par l'algorithme BLS-SVD++.

5.4 Comparaison avec les approches conventionnelles

Comme dernière série d'expériences, nous proposons de comparer l'approche retenue dans la section précédente avec les approches conventionnelles. Tableau 4 présente la moyenne de la P@3 des prédictions obtenues avec les approches testées sur l'ensemble du jeu de données du test.

Tableau 4. Comparaison de l'approche proposée avec les approches conventionnelles

Approches testées	P@3
Approche proposée	0.358
BPR-MF	0.316
Random	0.09
Plus Populaire	0.134
NeuMF	0.314

Les résultats montrent que sur l'intervalle de test, l'approche proposée a mieux performé que les approches conventionnelles. Cela signifie que dans une certaine mesure, le SR développé peut produire des recommandations personnalisées plus précises. Nous avons montré, à partir d'une base de données réelle, que la création automatique d'une nouvelle taxonomie basée sur la description textuelle des produits permet de surmonter les deux principaux défis du filtrage collaboratif : le problème de la sparsité et le problème des synonymes. D'autre part, l'introduction de règles d'association pour créer des transactions "artificielles" afin d'améliorer la matrice utilisateur-produit améliore les performances en réduisant la sparsité de la matrice. Une autre hypothèse a également été testée impliquant une initialisation non aléatoire (mais plutôt basée sur une technique de décomposition de la valeur singulière). Notre résultat montre qu'une telle initialisation améliore significativement la précision de la technique de factorisation matricielle.

6 CONCLUSION

Cet article a pour objectif de développer un système de recommandation pour un distributeur offrant ses produits à des clients industriels B2B, en utilisant le filtrage collaboratif basé sur le feedback implicite. Bien que cette technique soit la plus couramment utilisée pour la recommandation, elle est confrontée au problème de la sparsité des données. En effet, la matrice utilisateur-produit dans un contexte de ventes B2B est souvent incomplète et contient des données manquantes, ce qui

pourrait affecter la qualité et la précision des recommandations. Pour pallier ce problème, nous avons proposé une nouvelle approche basée sur l'utilisation de la taxonomie des produits et l'extraction de règles d'association. Premièrement, nous avons proposé de générer automatiquement une nouvelle taxonomie explicite basée sur les descriptions textuelles des produits. Deuxièmement, nous avons renforcé la matrice utilisateur-produit en exploitant les résultats de l'extraction de règles d'association. Nous avons ainsi ajouté des produits complémentaires aux utilisateurs en fonction de leurs achats des produits de base de la règle. Troisièmement, nous avons proposé d'initialiser l'algorithme de factorisation matricielle (i-ALS) en utilisant la technique de décomposition en valeurs singulières (BLS-SVD++).

Les résultats obtenus à la suite d'expériences menées sur les données réelles du partenaire industriel démontrent que l'approche proposée performe mieux que les approches conventionnelles en SR. L'étude met en évidence la pertinence la capacité de la nouvelle taxonomie créée et des techniques d'extraction de règles d'association à surmonter le problème de la sparsité des données. L'étude souligne également l'importance de l'effet de l'initialisation de l'algorithme principal sur la précision.

Dans cette étude, seule une évaluation hors ligne a pu être réalisée sur les données du partenaire. Dès que le système de recommandation proposé serait implémenté sur le site web du partenaire industriel, des expériences en ligne peuvent être menées pour mesurer en plus du pouvoir prédictif du système de recommandation proposé, l'influence du système sur le comportement de l'utilisateur.

7 REMERCIEMENTS

Les auteurs remercient MITACS et l'entreprise JSV Inc. pour le support financier.

8 REFERENCES

- Abbas, A., Zhang, L., & Khan, S. U. (2015). A survey on context-aware recommender systems based on computational intelligence techniques. *Computing*, 97, 667-690.
- Ahmed, A., Kanagal, B., Pandey, S., Josifovski, V., Pueyo, L. G., & Yuan, J. (2013). Latent factor models with additive and hierarchically-smoothed user preferences. *Proceedings of the sixth ACM international conference on Web search and data mining*.
- Chan, H. L., Pang, K. W., & Li, K. W. (2011). Association rule based approach for improving operation efficiency in a randomized warehouse. *Proc. Int. Conf. Ind. Eng. Oper. Manage.*,
- Donoho, D. L., & Grimes, C. (2003). Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 100(10), 5591-5596.
- Feng, H., Tian, J., Wang, H. J., & Li, M. (2015). Personalized recommendations based on time-weighted overlapping community detection. *Information & Management*, 52(7), 789-800.

- Gicquel, P.-Y., & Lenne, D. (2013). Semantic and Contextual Proximities for Informal Learning: The Case Study of Museum Visits. *International Journal of Computer Applications*, 69(26), <http://www.ijcaonline.org/archives/volume69/number26/12134-18410>.
- Guo, Y., Wang, M., & Li, X. (2017). Application of an improved Apriori algorithm in a mobile e-commerce recommendation system. *Industrial Management & Data Systems*, 117(2), 287-303.
- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). Neural collaborative filtering. Proceedings of the 26th international conference on world wide web,
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5-53.
- Johnson, C. C. (2014). Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems*, 27(78), 1-9.
- Kanagal, B., Ahmed, A., Pandey, S., Josifovski, V., Yuan, J., & Garcia-Pueyo, L. (2012). Supercharging recommender systems using taxonomies for learning user purchase behavior. *arXiv preprint arXiv:1207.0136*.
- Koren, Y. (2008). Factorization meets the neighborhood: a multifaceted collaborative filtering model. Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining,
- Lawrence, R. D., Almasi, G. S., Kotlyar, V., Viveros, M., & Duri, S. S. (2001). *Personalization of supermarket product recommendations*. Springer.
- Liphoto, M., Du, C., & Ngwira, S. (2016). A survey on recommender systems. 2016 International Conference on Advances in Computing and Communication Engineering (ICACCE),
- Liu, N. N., Xiang, E. W., Zhao, M., & Yang, Q. (2010). Unifying explicit and implicit feedback for collaborative filtering. Proceedings of the 19th ACM international conference on Information and knowledge management,
- Maidel, V., Shoval, P., Shapira, B., & Taieb-Maimon, M. (2008). *Evaluation of an ontology-content based filtering method for a personalized newspaper*. Proceedings of the 2008 ACM conference on Recommender systems, Lausanne, Switzerland. <https://doi.org/10.1145/1454008.1454024>
- Mnih, A. (2012). Taxonomy-informed latent factor models for implicit feedback. Proceedings of KDD Cup 2011,
- Nia, A. G., Lu, J., Zhang, Q., & Ribeiro, M. (2019). A framework for a large-scale B2B recommender system. 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE),
- Park, Y., Park, S., Jung, W., & Lee, S.-g. (2015). Reversed CF: A fast collaborative filtering algorithm using a k-nearest neighbor graph. *Expert Systems with Applications*, 42(8), 4022-4028.
- Rendle, S., & Schmidt-Thieme, L. (2008). Online-updating regularized kernel matrix factorization models for large-scale recommender systems. Proceedings of the 2008 ACM conference on Recommender systems,
- Ricci, F., Rokach, L., & Shapira, B. (2010). Introduction to recommender systems handbook. In *Recommender systems handbook* (pp. 1-35). Springer.
- Schoinas, I., & Tjortjjs, C. (2019). MuSIF: a product recommendation system based on multi-source implicit feedback. Artificial Intelligence Applications and Innovations: 15th IFIP WG 12.5 International Conference, AIAI 2019, Hersonissos, Crete, Greece, May 24–26, 2019, Proceedings 15,
- Sharma, M., & Mann, S. (2013). A survey of recommender systems: approaches and limitations. *International Journal of Innovations in Engineering and Technology*, 2(2), 8-14.
- Smilde, A. K., Geladi, P., & Bro, R. (2005). *Multi-way analysis: applications in the chemical sciences*. John Wiley & Sons.
- Tyagi, S., & Bharadwaj, K. K. (2013). Enhancing collaborative filtering recommendations by utilizing multi-objective particle swarm optimization embedded association rule mining. *Swarm and Evolutionary Computation*, 13, 1-12.
- Wang, J., Han, P., Miao, Y., & Zhang, F. (2019). A collaborative filtering algorithm based on svd and trust factor. 2019 international conference on computer, network, communication and information systems (CNCI 2019),
- Wang, S., Sun, G., & Li, Y. (2020). SVD++ recommendation algorithm based on backtracking. *Information*, 11(7), 369.
- Zhang, Y., Ahmed, A., Josifovski, V., & Smola, A. (2014). Taxonomy discovery for personalized recommendation. Proceedings of the 7th ACM international conference on Web search and data mining,
- Ziegler, C.-N., Lausen, G., & Schmidt-Thieme, L. (2004). Taxonomy-driven computation of product recommendations. Proceedings of the thirteenth ACM international conference on Information and knowledge management,