

Recommandation interactive, une solution au démarrage à froid

CAMELIA DADOUCHI^{1,2}, BRUNO AGARD^{1,3}, BENOIT MONTREUIL⁴

¹ Laboratoire en Intelligence des Données, Département de mathématiques et génie industriel, École Polytechnique de Montréal, CP 6079, succursale Centre-Ville, Montréal, Québec, Canada, Centre Interuniversitaire de Recherche sur les Réseaux d'Entreprise, la Logistique et le Transport (CIRRELT)

²camelia.dadouchi@polymtl.ca, ³bruno.agard@polymtl.ca,

⁴ School of Industrial and Systems Engineering, Georgia Institute of Technology, Physical Internet Center, United States, Supply Chain and Logistics Institute, Coca-Cola Chair in Material Handling et Distribution, United States, benoit.montreuil@isye.gatech.edu

Résumé

L'un des problèmes les plus connus dans les systèmes de recommandation (RS) est le problème du démarrage à froid [Lika et al., 2014]. Dans un contexte de vente stationnaire, le même problème se produit lorsqu'un client se présente pour la première fois. Le vendeur doit déterminer quels produits montrer au client afin d'évaluer ses intérêts et de faire une recommandation dans un temps limité. Afin d'accroître la satisfaction de la clientèle, nous proposons une méthode de recommandation de produits basée sur le contexte et le contenu permettant d'évaluer les préférences de la clientèle en quatre étapes, menant à une recommandation de produit basée sur des questions directes et le profil du client. La collection de produits présentées au client pour appréciation est dynamique et basée sur les appréciations précédentes.

Abstract

One of the most known problems in Recommender systems (RSs) is the cold start problem [Lika et al., 2014]. In a brick and mortar context, the same problem occurs when a client walks in for the first time. The salesman needs to determine which product to show the customer in order to evaluate his interests and make a recommendation in a limited time. In order to increase client satisfaction, we propose a method that evaluate clients preferences through four stages, leading to product recommendation based on direct questions and customer's profile. The products presented to the customer for appreciation collection is dynamic and based on previous appreciations.

Mots clés

Système de recommandation, démarrage à froid, construction de profil, vente stationnaire.

Keywords

Recommendation system, cold start, profile construction, brick and mortar.

1 INTRODUCTION

Les systèmes de recommandations (SRs), en plus d'être une excellente solution à la surcharge d'information des clients, sont des outils de marketing et de vente puissants [Pathak et al., 2010]. Cependant, pour les mettre en place, des données d'entrées exploitables sont nécessaires. Ceci peut s'avérer compliqué dans plusieurs contextes où il est difficile de créer une base d'apprentissage solide. Parmi les secteurs touchés, on trouve ceux avec des biens de valeur monétaire élevé.

Dans un tel contexte, le client favorise l'appréciation physique des produits afin de juger de leurs adéquations à ses besoins.

Le vendeur, faisant face à un choix d'inventaire de taille conséquente et à un ensemble de produits limités en magasin, se doit de déterminer les intérêts de son client. Des contraintes de temps disponibles avec le client et de nombre de produits à présenter à celui-ci sans le surcharger d'informations sont à prendre en compte.

Dans le présent article, nous proposons un outil qui vise à définir les intérêts des clients afin de leurs proposer les produits qui répondront le mieux à leurs besoins, en évitant de les surcharger d'information, de façon à rendre l'expérience client agréable et également de créer une base d'apprentissage sur les intérêts des clients.

La structure de l'article se décompose de la manière suivante : la section 2 présente l'état de l'art, qui contient une définition des systèmes de recommandation (2.1), les catégories de systèmes de recommandation (2.2), les différentes données utilisées (2.3) par ces systèmes et une des limites les plus persistantes des SRs qu'est le démarrage à froid (2.4). La section 3 présente l'outil de recommandation proposé incluant une mise en situation (3.1), les données d'entrées (3.2) et la méthodologie (3.3) présentant, le formatage des données (3.3.1), le filtrage préliminaire (3.3.2), l'exploration et la création du profil du client (3.3.3) et finalement la phase de recommandation d'un ensemble de produit (3.2.4). La section 4 exposera la mise en œuvre détaillée de la méthode à travers un cas d'étude, la section 5 clôture l'article en rappelant les résultats, les limites de la méthode présentée et propose quelques perspectives.

2 ÉTAT DE L'ART

2.1 Systèmes de recommandation

Les systèmes de recommandation ont été présenté par [Jie et al., 2015] comme étant des outils qui assistent les utilisateurs avec les problèmes de surcharge croissante d'informations et qui améliorent la gestion de la relation client en fournissant aux

utilisateurs des recommandations personnalisées de produits et/ou de services.

Plus formellement, [Adomavicius et Tuzhilin, 2005] présente la définition de la façon suivante : « $\forall c \in C, s'c = \operatorname{argmax}_{s \in S} u(c,s)$ ». Tel que u est la fonction d'utilité qui mesure l'utilité d'un item s pour un utilisateur c , $u: C \times S \rightarrow R$, R est un ensemble ordonné (un ensemble d'entiers non négatifs ou un ensemble de réels dans un certain intervalle). On veut donc choisir un sous ensemble d'items $s' \in S$ qui maximise l'utilité pour l'utilisateur $c \in C$.

Les définitions pour les systèmes de recommandation diffèrent d'une communauté à une autre, mais restent tout de même toutes axées sur l'établissement d'une liste d'items à recommander aux utilisateurs. Cette recommandation se base sur des algorithmes issus de différentes disciplines qui permettent de prédire l'appréciation d'un utilisateur pour une liste de produits inconnus et de lui suggérer, parmi un large éventail d'items, ceux qui sauront répondre à des besoins spécifiques. [Dadouchi et Agard, 2017]

2.2 Les catégories de systèmes de recommandation :

La typologie traditionnelle des systèmes de recommandation a été établie par [Adomavicius et Tuzhilin, 2005]. Les types de recommandations qu'ils proposent sont basés sur les données d'entrées disponibles et les techniques internes du système, les trois méthodes de recommandations sont présentées ci-dessous :

Recommandations collaboratives (Collaborative recommendations/filtering- CF) :

Le filtrage collaboratif (CF) est la méthode la plus répandue dans le monde de la recommandation [Jamali et Ester, 2010]. Comme son nom l'indique, cette approche est à caractère collaboratif. Elle se base sur l'hypothèse du « stéréotype » qui implique qu'un utilisateur faisant partie d'un groupe de personnes ayant eu des préférences semblables dans le passé, auront des goûts semblables aux leurs dans le futur.

Cette méthode permet de faire des recommandations sans connaître les caractéristiques des items en se basant sur les préférences passées d'autres utilisateurs qui sont considérés semblables à lui. La liste de recommandation est construite avec les N items ayant les meilleures prédictions [Adomavicius et Tuzhilin, 2005].

La méthode collaborative présente des avantages par rapport au filtrage basé sur le contenu. En effet, elle est plus efficace dans la pratique et simple à mettre en oeuvre [Poza et al., 2015].

Recommandations basées sur le contenu (Content-based recommendations/filtering- CB) :

De façon très simple, la recommandation basée sur le contenu (content-based filtering - CBF) tente de définir les préférences des utilisateurs et de les comparer avec les caractéristiques des items disponibles. Pour ce faire les profils des items sont composés de leurs caractéristiques, tandis que les profils des utilisateurs sont composés par les caractéristiques des items précédemment appréciés par les utilisateurs. La prédiction des appréciations de l'utilisateur pour tous les items non évalués par celui-ci est ensuite calculée [Adomavicius et Tuzhilin, 2005].

Les approches CBF nécessitent des évaluations faites par l'utilisateur lui-même, contrairement aux modèles CF qui ne peuvent pas obtenir un résultat efficace sans les évaluations d'autres utilisateurs [Lika et al., 2014].

Approches hybrides (Hybrid approaches/filtering):

Ces méthodes combinent les deux méthodes précédentes.

2.3 Calcul de similarité

Les systèmes de recommandation requièrent souvent des calculs de similarité que ce soit pour calculer les similarités entre l'utilisateur et le profil des items ou entre un utilisateur cible et tous les autres utilisateurs.

La méthode la plus simple est le calcul des distances. Parmi les distances les plus utilisées, on trouve la distance de Minkowski qui est une généralisation de la distance euclidienne :

$$d(x, y) = \sum_{k=1}^n (|x_k - y_k|^r)^{\frac{1}{r}} \quad (1)$$

Avec :

n : nombre de dimensions,

x_k et y_k : les k^{th} composants des objets x et y respectivement.

r : degré de distance. Ex : $r = 2$ pour la distance euclidienne.

La « cosine similarity » [Bhattacharyya, 1946] est la méthode la plus répandue pour les calculs de similarité. Elle se base sur l'évaluation du cosinus de l'angle entre deux vecteurs à n dimensions, chacun représentant les items ou les utilisateurs (x, y) . Elle est définie comme suit :

$$\operatorname{Cos}(x, y) = \frac{(x \cdot y)}{\|x\| \|y\|} \quad (2)$$

La similarité peut aussi être donnée par la corrélation. La corrélation entre deux objets est la mesure de la relation de linéarité entre ses deux objets. La mesure la plus utilisée est la corrélation de Pearson [Pearson, 1896], donnée par l'équation 3.

$$\operatorname{Pearson}(x, y) = \frac{\operatorname{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \quad (3)$$

Il existe différentes variantes des mesures énoncées précédemment, ainsi que d'autres mesures de similarité pour les données binaires comme « The Simple Matching coefficient SMC, the Jaccard coefficient, the Extended Jaccard (Tanimoto) coefficient. » [Amatriain et al., 2015].

Les mesures de similarités ont évolué avec le temps pour tenir compte de la variance dans les évaluations entre les usagers. Par exemple, la cosine similarité ajustée soustrait de l'évaluation de l'utilisateur la moyenne de ses évaluations passées avant de calculer les similarités sur lesquelles se basent la recommandation. Dans le cas général, il a été démontré dans la littérature que la précision des prédictions pour les recommandations était indépendante du choix de la mesure de similarité à utiliser [Lathia et al., 2008].

2.4 Données d'entrée

Il existe deux types de données communément utilisées dans la recommandation de produits : les données implicites et les données explicites.

Les données implicites sont recueillies de façon indirectes contrairement aux données explicites que l'utilisateur fournit à travers ses rétroactions.

Il existe trois niveaux de données catégorisées par [Bobadilla et al., 2013] : les données de contenu, données démographiques, caractéristiques et descriptions de produits. Les données sociales découlant des activités des utilisateurs sur les réseaux sociaux (amis, tweet, appréciation, postes, ...) et finalement les données contextuelles comme les données tirées des RFID, des capteurs de température de précision, de mouvement ou autre.

Des systèmes de recommandation interactifs plus récents recueillent des informations explicites à travers une interface où l'utilisateur est sollicité à fournir des renseignements sur ses

préférences en répondant à des questions [He, Parra et Verbert, 2016].

2.5 Démarrage à froid

Le problème de démarrage à froid (Cold Start) est le phénomène qui survient quand un SR peine à fournir des recommandations dues à un manque d'informations initial sur les appréciations des utilisateurs [Schafer et al., 2007]. Les problèmes de démarrage à froids se divisent en trois catégories : nouvelle communauté (New community), nouvel utilisateur (new user) et nouvel item (new item).

Le problème du nouvel utilisateur, fait partie des grandes difficultés rencontrées par les systèmes de recommandation en action. Lorsque les utilisateurs s'inscrivent, ils n'ont pas encore émis de vote et, par conséquent, ils ne peuvent recevoir aucune recommandation personnalisée [Rashid, Karypis et Riedl, 2008].

Généralement, la solution à ces problèmes est de combiner plusieurs méthodes de recommandations (CF-content based RS, CF-demographic based RS, CF-social based RS). En guise d'exemple, la combinaison du CBF et du CF peut être utilisé pour inclure les caractéristiques des items et les évaluations simultanément dans le calcul des recommandations. Des questions explicites peuvent aussi être posées aux utilisateurs pour avoir des informations générales sur ces derniers. [Bucklin et Sismeiro, 2009]. L'exploration des interactions transitives peut aussi être une solution au démarrage à froid [Leung et al., 2008].

3 L'OUTIL PROPOSE

3.1 Mise en situation

Les systèmes de recommandations ont besoin pour leurs fonctionnements d'un ensemble de données d'entrées tel que présenté à la section 4. Cependant, dans certains cas, l'accès à l'information pertinente et récente pour la recommandation de produits peut s'avérer laborieux ou impossible.

Ce problème s'apparente au problème de démarrage à froid présenté à la (section 2). Dans certaines industries, comme l'industrie de l'automobile ou la valeur monétaire des biens est élevé, ce problème est plus persistant, car les clients sont plus insistants sur la notion d'« assurance » [Rajamma et al., 2007] lors de leurs prise de décision d'achat. Les clients ont tendance à favoriser les achats en magasins, pour pouvoir apprécier les caractéristiques physiques des produits. Les données sur l'activité en ligne d'un client peuvent donc être difficiles à lier à son achat final. Cela rend les données du click-stream, même quand elles sont disponibles, non exploitables pour de la recommandation personnalisée.

Techniquement, n'ayant pas le profil des clients ni le profil des communautés, il est impossible d'utiliser des techniques de filtrage collaboratif. Les techniques de filtrage basées sur le contenu quant à elles, peuvent être utilisées en considérant les caractéristiques des produits. Cependant, un profil de client se basant sur son appréciation des caractéristiques de différents produits, devrait être bâti.

Allier les systèmes collaboratifs aux systèmes basés sur le contenu ou poser des questions prédéfinies directes aux clients sur leurs intérêts sont des façons utilisées dans la littérature pour faire face au problème de démarrage à froid [Bucklin et Sismeiro, 2009]. Cela implique la disponibilité d'informations suffisantes pour la

mise en place d'un SR ou que le client ait une idée claire de ses intérêts.

Déterminer des questions fixes à l'avance peut s'avérer long et peut rendre l'expérience client moins agréable et non personnalisée.

Nous proposons donc une approche interactive permettant de recueillir les intérêts des clients travers une recommandation alliant les questions directes à la déduction implicite des intérêts des clients à travers leurs rétroactions.

La rétroaction du client permet d'adapter la sélection de produit qui lui sera présentée. Une fois les appréciations des clients pour un sous ensemble de produits recueillis, des appréciations les attributs des caractéristiques en sont déduits. Ces appréciations permettent de construire le profil client et de l'utiliser pour faire une recommandation de produit basée sur le contenu. La section 3.3 présente la méthode proposée en détails.

3.2 Données d'entrées

Les données d'entrées nécessaires à la mise en place de cette méthodologie sont présentées ci-dessous :

- Inventaire des produits disponibles et délai de livraison du produit.
- Ensemble des caractéristiques représentant les produits.
- Matrice de similarité entre les produits.

La matrice de similarité des produits est une matrice représentant le degré de ressemblance entre un produit et l'ensemble des produits disponibles. La matrice de similarité entre les caractéristiques est souvent utilisée pour la construction de similarité entre les produits. Il peut y avoir un ensemble de matrices de similarité entre les caractéristiques. Chaque attribut est comparé à l'ensemble des attributs que l'on peut donner à la caractéristique. (Montreuil & Derhami, 2019)

3.3 Méthodologie

Comme présenté dans la mise en situation, l'objectif de la méthodologie est de recueillir de l'information sur les intérêts du client de façon interactive afin de remédier au problème de démarrage à froid et de faire de la recommandation de produit personnalisée en magasin.

La suite de la méthodologie s'articule autour de quatre phases, la première phase consiste au formatage des données (3.3.1), s'en suit la phase de filtrage préliminaire (section 3.3.2), puis, la phase d'exploration interactive et de création du profil du client (section 3.3.3) et se clôt par la phase de recommandation (section 3.3.4).

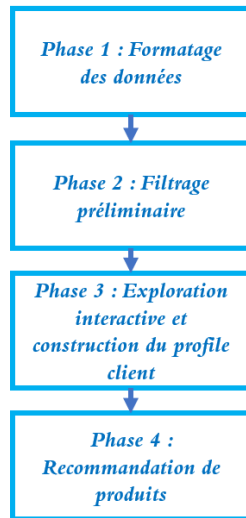


Figure 1 : Méthodologie de recommandation interactive de produits

3.3.1 Phase 1 : Formattage des données

Le formattage de données est une étape nécessaire à l'élaboration de toute valorisation de données. La recommandation de produit basée sur le contenu proposé à la section 3.3.4 nécessite des calculs de similarité. Il est donc nécessaire de formater les données de façon à pouvoir les exploiter. La création du profil de l'utilisateur et des profils de produits se fait grâce à la création de vecteurs se composant des attributs qui représentent l'ensemble des caractéristiques des produits.

Prenant l'exemple de produits aux formes géométriques différentes. Parmi les caractéristiques qui permettent de les représenter on trouve la largeur, la longueur, la profondeur, le rayon, le centre, la couleur. Chaque caractéristique à plusieurs attributs possibles. Les attributs définissant le vecteur présentent l'ensemble des choix possibles pour chaque caractéristique. Les caractéristiques représentées par des attributs continues sont discrétisées.

Les attributs sont par la suite représentés de façon binaire. Les attributs caractérisant le produit sont notés 1, tandis que les attributs ne représentant pas le produit en question sont notés 0. Nous nous retrouvons donc avec un vecteur d'attributs binaires. Si l'on prend une table rectangulaire marron rectangulaire, le vecteur la représentant serait formaté tel que présenté au tableau 1.

Tableau 1 : Table des attributs de la table rectangulaire marron et d'une table ronde jaune.

Attribut	Bleu	Marron	Jaune	Largeur (0-2)	Largeur (2-5)	Profondeur (0-5)	Diamètre (0-5)	Centre (0,0)-(5,5)
Table	-	1	-	1	0	1	-	-
Table ronde	-	-	1	0	0	0	1	1

Les deux vecteurs sont de tailles et de caractéristiques identiques. Ce formattage permet de calculer les distances et les similarités entre les vecteurs.

3.3.2 Phase 2 : Filtrage préliminaire

Cette phase se base sur le concept du niveau de discrimination des caractéristiques. En effet, nous proposons deux types de caractéristiques :

Caractéristiques discriminantes : menant à un choix d'attributs impliquant l'annulation des autres choix possibles. Comme exemple de caractéristiques discriminantes nous proposons la fonctionnalité (se traduisant par un ensemble de caractéristiques spécifiques au métier), disponibilité en fonction du temps d'attente et la gamme de prix. Les caractéristiques discriminantes peuvent servir de filtrage préliminaire des produits à présenter au client. Elles sont généralement facilement déductibles grâce aux connaissances métiers.

Caractéristiques appréciables : Les caractéristiques dont l'appréciation est non exclusive. Le client peut avoir des préférences dont l'importance est variable en fonction des caractéristiques. Le type de moteur peut être une caractéristique importante tandis que la couleur est une caractéristique sur laquelle le compromis peut être plus facile.

Les caractéristiques appréciables représentant les goûts du client seront déduites lors de la création du profil de celui-ci. La moyenne pondérée de l'appréciation que l'utilisateur a donnée au sous-ensemble de produits qui lui ont été présentés est utilisée pour déduire l'appréciation moyenne du client pour les caractéristiques des produits.

Cette première étape se base sur les connaissances métiers qui permettent de déterminer les caractéristiques discriminantes. Les caractéristiques discriminantes résultent en un ensemble de questions directes dont la réponse est catégorique, cela mène à un filtrage des choix initiaux.

Nous considérons également qu'en plus du filtrage préliminaire sur les caractéristiques discriminantes, les contraintes logistiques de disponibilité des produits sont prises en compte en fonction de la date à laquelle le produit est requis. (Montreuil & Derhami, S., 2019)

La phase 3 traite le sous-ensemble de produits (S) disponibles en magasin, découlant du filtrage préliminaire. La phase 4 est appliquée au sous-ensemble de produit (S-R) découlant du filtrage préliminaire qui peuvent être disponible dans les délais acceptables par le client.

3.3.3 Phase 3 : Exploration et création du profil du client

La phase d'exploration permet de sélectionner un ensemble de produits à présenter au client afin d'en recueillir les appréciations. Ces appréciations sont utilisées pour la construction du profil du client (PC). Les intérêts des clients sont rassemblés dans un profil unique à celui-ci. Ce profil représente l'ensemble des appréciations recueillies sur un ensemble de produits pour en déduire l'appréciation implicite des caractéristiques qui leurs sont associées. Le premier problème pour cette phase est au niveau de la sélection du sous-ensemble à présenter au client pour recueillir son appréciation.

En effet, présenter des produits au client lors de la phase d'exploration est sensiblement semblable à les lui recommander étant donné la mise en relief que cela implique. Par conséquent, un compromis entre une large couverture de l'ensemble des caractéristiques présents dans les produits présentés au client et la convergence rapide vers les produits ayant une appréciation élevée est essentiel. Donc, afin d'éviter la surcharge d'information et de converger rapidement, l'exploration de produits se base sur

la rétroaction du client. Les appréciations recueillies définissent le prochain produit à présenter au client.

La figure 2 présente l'algorithme de construction du profil du client. Les 7 étapes de celle-ci sont expliquées ci-dessous :

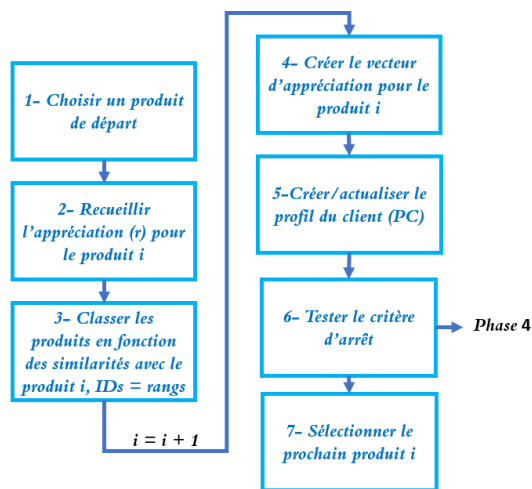


Figure 2 : Algorithme de construction du profil du client

Paramètres :

Y : Nombre minimal de produits à présenter en phase d'exploration.

X : Nombre maximal de produits à présenter en phase d'exploration.

Z : Niveau de similarité acceptable pour considérer un produit satisfaisant pour répondre aux besoins du client.

W : Nombre de produits à recommander.

Initialisation :

Saut = 0

ID = 1

Étapes :

Étape 1 : choisir un produit de départ

Cette étape vise à déterminer un produit de départ pour l'exploration du sous ensemble de produits (S) découlant de la phase 2. Plusieurs choix de produits peuvent être justifiables, soit-il en fonction des connaissances métier, du temps disponibles avec le client pour converger vers la recommandation, des promotions en cours ou du meilleur vendeur.

Étape 2 : Recueillir l'appréciation (r) pour le produit i

Le produit i est présenté au client et le niveau d'appréciation (r) de celui-ci est recueilli, $r \in (1:10)$.

1 étant l'appréciation minimale, 10 étant l'appréciation maximale.

Étape 3 : Classer les produits

Cette étape se base sur la similarité entre les produits afin de déterminer l'ordre d'exploration des produits. Les produits sont ordonnés du plus similaire au plus dissimilaire par rapport au produit i. Cet ensemble ordonné est nommé O-S. Chaque produit à un indice qui lui est assigné en fonction de son positionnement dans le classement. Le produit de départ i a toujours un indice ID =1. Pour le reste des étapes les indices des produits sont actualisés en fonction de leurs rangs.

La similarité peut être déterminée grâce aux connaissances métier ou être calculée comme étant la distance entre les produits tel que présenté à la section 3.2.

À chaque itération le produit présenté au client est soustrait de l'ensemble des produits O-S.

$O-S = O-S - \text{produit } i$

Étape 4 : Créer le vecteur d'appréciation des attributs

À cette étape, nous émettons l'hypothèse que l'appréciation du produit est équivalente à l'appréciation de chacun de ses attributs. L'apriori ici étant qu'un utilisateur qui apprécie un produit en apprécie les attributs.

Étant donné que l'ensemble des attributs est binaire. L'appréciation du produit i est reporté à ses attributs par une multiplication de l'ensemble du vecteur par le ratio d'appréciation du produit recueillie à l'étape 4.

Étape 5 : Actualiser le profil du client

Le profil du client (PC) est la moyenne des vecteurs d'appréciation des attributs pour l'ensemble des produits évalués.

$$\frac{\sum_1^i vr(\text{Produit}_i)}{i}$$

Étape 6 : Tester le critère d'arrêt

Considérons i étant le nombre de produits dont l'appréciation a été recueillie. Pour chaque itération, i est actualisé

Si $i < Y$, Passer à l'étape 7.

Si $Y > i \geq X$, Calculer la similarité (Sim) entre PC et S.

Si $\text{Sim}(\text{PC}, S) > Z$, Passer à la phase 4.

Sinon Passer à l'étape 7,

Si $i > X$, passer à la phase 4.

Étape 7 : Sélectionner le prochain produit

La sélection du produit se fait grâce à des sauts dans l'ensemble O-S.

Le saut est défini comme étant le nombre de produits par lequel on se décale dans l'ensemble O-S pour sélectionner le prochain produit à présenter au client. Il est calculé par l'équation (1) défini ci-dessous :

$$\text{Saut} = (1 - r) * \text{Longueur}(\text{O-S}) \quad (2)$$

Longueur(O-S) est le nombre de produits dans l'ensemble de produits O-S.

Le prochain produit est le produit associé à l'ID actualisé.

ID = arrondi (Saut),

Le nombre de produit évalué est incrémenté, $i = i + 1$, aller à l'étape 2.

3.3.4 Phase 4 : Recommandation d'un ensemble de produits

La quatrième et dernière phase est celle de la recommandation. Celle-ci suit une méthode classique de recommandation basée sur le contenu.

On utilisera le calcul de similarité entre le profil du client que nous avons construit à la phase trois et le vecteur de caractéristiques des produits du sous-ensembles (S-R) disponibles dans les temps voulus.

Une fois que le profil du client PC est construit, un ensemble de produits est recommandé à celui-ci. Cet ensemble de produits est sélectionné à partir de l'ensemble des produits disponibles. En fonction de leurs similarités avec le profil du client.

La similarité est calculée grâce à la formule de similarité basée sur le cosinus (formule 2).

4 CAS D'ÉTUDE

Le cas d'étude présente l'application de la méthode proposée à la section 3 sur un ensemble de huit produits ayant huit attributs chacun. Nous nous positionnons dans le contexte de vente de voitures.

Phase 1 : Formatage de données

Les données d'entrée nécessaires à la présentation du cas d'étude sont : les modèles de voiture disponibles ainsi que leurs caractéristiques et la matrice de similarité entre les produits en inventaire.

Les vecteurs des attributs des produits sont formatés tel qu'indiqué à la section 3.3.1.

Afin de simplifier l'exemple, le formatage des données sera présenté à la phase 2, dans les tableaux 3 et 4 représentant la matrice de similarité entre les produits et les vecteurs représentant les produits respectivement.

Phase 2 : Filtrage préliminaire

Nous considérons que le filtrage préliminaire a été effectué suite aux réponses recueillies en posant les questions définies plus bas.

Questions discriminantes :

- Combien de temps pouvez-vous attendre pour recevoir votre produit ?

Je souhaite avoir le produit aujourd'hui.

- Pour quel type d'activité comptez-vous utiliser votre véhicule ?

Je souhaite utiliser mon véhicule pour des courses sur circuit.

- Combien de passagers souhaitez-vous transporter ?

Je ne compte pas transporter de passagers, je me déplace seul.

Ce filtrage permet d'éliminer tous les produits qui ne sont pas disponibles en magasin, qui ne sont pas des véhicules de type sport et qui ont plus que 2 sièges. La matrice de similarité entre les produits filtrés est présentée au tableau 3. Le tableau 4 présente les produits filtrés ainsi que les attributs les caractérisant.

Pour notre cas d'étude l'ensemble de produits S disponibles en magasin est équivalent à l'ensemble des produits S-R disponibles dans le réseau du client car le client souhaite avoir le produit à l'instant même.

Tableau 2 : Matrice de similarité entre les produits de l'ensemble S.

	00R0110	00R0111	00A0112	00E0113	00E0114	00P0115	00O0116	00W0117
00R0110	1	0	0.258 1989	0.774 5967	0	0.654 6537	0.288 6751	0.288 6751
00R0111	0	1	0.774 5967	0	0.816 4966	0.436 4358	0.577 3503	0.577 3503
00A0112	0.258 1989	0.774 5967	1	0.4	0.632 4555	0.676 1234	0.894 4272	0.447 2136
00E0113	0.774 5967	0	0.4	1	0	0.845 1543	0.447 2136	0.447 2136
00E0114	0	0.816 4966	0.632 4555	0	1	0.267 2612	0.353 5534	0.353 5534
00P0115	0.654 6537	0.436 4358	0.676 1234	0.845 1543	0.267 2612	1	0.566 9467	0.566 9467
00O0116	0.288 6751	0.577 3503	0.894 4272	0.447 2136	0.353 5534	0.566 9467	1	0.5
00W0117	0.288 6751	0.577 3503	0.447 2136	0.447 2136	0.353 5534	0.566 9467	0.5	1

Tableau 3: Identifiants, noms et attributs des produits de l'ensemble S.

ID	Nom du	A1	A2	B1	B2	B3	C	D	E
----	--------	----	----	----	----	----	---	---	---

	Produit								
1	00R0110	1	0	0	1	0	1	0	0
2	00R0111	1	0	0	1	0	1	0	1
3	00A0112	1	0	1	1	0	1	1	0
4	00E0113	0	0	0	1	0	1	0	1
5	00E0114	0	0	1	1	0	1	1	0
6	00P0115	0	1	1	0	0	1	1	1
7	00O0116	0	1	1	0	1	1	0	0
8	00W0117	0	1	1	0	1	0	1	1

Phase 3 : Exploration interactive et création du profil client.

Paramètres :

Y : 3 ; X : 5 ; Z : 80% ; W : 3

Initialisation :

Saut = 0

ID = 1

Itération 1 :

Étape 1 :

Nous considérons le produit le plus populaire comme produit de départ. Avec le niveau d'informations disponibles, produit le plus populaire est celui qui a le plus de chance d'intéresser le client. Le meilleur vendeur pour ce cas d'étude est le produit **00A0112**.

Étape 2 :

L'appréciation que l'utilisateur donne au produit est :

$R(00A0112) = 7/10 = 70\%$

Étape 3 : Classer les produits en fonction de leurs similarités avec le produit i

Le tableau 4 présente l'ensemble des produits classés en fonction de leurs similarités avec le produit 00A0112.

Tableau 4: Ensemble de produits triés (O-S) en fonction de leurs similarités avec le produit 00A0112

ID	Product	00A0112
1	00A0112	1
2	00O0116	0.8944272
3	00R0111	0.7745967
4	00P0115	0.6761234
5	00E0114	0.6324555
6	00W0117	0.4472136
7	00E0113	0.4
8	00R0110	0.2581989

Suppression du produit présenté au client de l'ensemble d'exploration

$O-S = O-S - 00A0112$.

$O-S = \{00O0116, 00R0111, 00P0115, 00E0114, 00W0117, 00E0113, 00R0110\}$

Étape 4 : Vecteur d'appréciation des attributs

Pour le produit 00A0112 ayant eu une appréciation de 70% le vecteur d'appréciation est :

$Vr(00A0112) = (0.7, 0, 0.7, 0.7, 0, 0.7, 0.7, 0)$

Étape 5 : Créer le profil du client

Le profil du client est la moyenne des vecteurs des appréciations donné à chacun des produits présentés au client. Ici un seul lui a été présenté. Le profil du client est donc équivalent à l'appréciation que le client a donné au produit 00A0112.

Profil client (PC) :

$$PC = \frac{(vr(00A0112))}{1}$$

Le tableau 5 présente un vecteur représentatif de l'appréciation de chacune des caractéristiques.

Tableau 5 : Vecteur représentant le profil du client à la première itération

PC	0.7	0	0.7	0.7	0	0.7	0.7	0
----	-----	---	-----	-----	---	-----	-----	---

Étape 6 : Tester le critère d'arrêt

Un seul produit a été présenté au client, la limite inférieure pour passer à la phase 4 n'a pas été atteinte

La première partie de la condition est vraie, car $n < 3$.

On passe donc au produit suivant dont le d'indice est calculé si dessous :

$$ID = \text{arrondi}(2.4) = 2,$$

Retourner à l'étape 2.

Le produit de départ est le produit 00R0111 associé à l'ID = 2.

Étape 7 : Sélectionner le prochain produit

$$\text{Saut} = (1-0.7) * \text{Longueur} (O-S) = 0.3*8 = 2.4.$$

Itération 2

Étape 2 :

L'Appréciation que l'utilisateur a donnée au produit est de :

$$R(00R0111) = 3/10 = 30\%$$

Étape 3 : Classer les produits

Tableau 6 : Ensemble de produits triés en fonction de leurs similarités avec le produit 00R0111

ID	Nom du produit	00R0111
1	00R0111	1
2	00E0114	0.8164966
3	00O0116	0.5773503
4	00W0117	0.5773503
5	00P0115	0.4364358
6	00R0110	0
7	00E0113	0

Suppression du produit présenté au client de l'ensemble d'exploration

$$O-S = O-S - 00R0111.$$

$$O-S = \{00O0116, 00P0115, 00E0114, 00W0117, 00A0112, 00R0110\}$$

Étape 4 : Vecteur d'appréciation des attributs

Pour le produit 00R0111 ayant eu une appréciation de 30% le vecteur d'appréciation est :

$$Vr(00R0111) = (0.3, 0, 0, 0.3, 0, 0.3, 0, 0.3)$$

Étape 5 : Actualiser le profil du client

Le profil du client est la moyenne des vecteurs des appréciations.

Profil du client (PC) :

$$PC = \frac{(vr(00A0112) + vr(00R0111))}{2}$$

Le tableau 7 présente un vecteur représentatif de l'appréciation de chacune des caractéristiques.

Tableau 7 : Vecteur représentant le profil du client à la deuxième itération

PC	0.5	0	0.35	0.5	0	0.5	0.35	0.15
----	-----	---	------	-----	---	-----	------	------

Étape 6 : Tester le critère d'arrêt

Deux itérations effectuées, $n < 3$, la première partie de la condition est vraie, car $n < 3$.

On passe donc au produit suivant.

Étape 7 : Sélectionner le prochain produit

$$\text{Saut} = (1-0.3) * \text{Longueur} (O-S) = 0.7 * 7 = 4.9$$

$$ID = \text{arrondi}(4.9) = 5,$$

Retourner à l'étape 2.

Le prochain produit est le produit 00E0114 dont l'ID=5.

Itération 3

Étape 2 : Appréciation 9/10 = 90%

Étape 3: Classer les produits

Tableau 8 : Ensemble de produits triés en fonction de leurs similarités avec le produit 00E0114

ID	Nom du produit	00E0114
1	00E0114	1
2	00O0116	0.3535534
3	00W0117	0.3535534
4	00P0115	0.2672612
5	00R0110	0
6	00E0113	0

Suppression du produit présenté au client de l'ensemble d'exploration

$$O-S = O-S - 00R0114.$$

$$O-S = \{00O0116, 00P0115, 00W0117, 00A0112, 00R0110\}$$

Étape 4: Vecteur d'appréciation des attributs

Pour le produit 00E0114 ayant eu une appréciation de 90% le vecteur d'appréciation est :

$$Vr(00E0114) = (0, 0, 0.9, 0.9, 0, 0.9, 0, 0)$$

Étape 5 : Actualiser le profil du client

Le profil du client est la moyenne des vecteurs des appréciations

Profil du client (PC) :

$$PC = \frac{(vr(00A0112) + vr(00R0111)) + vr(00E0114)}{3}$$

Le tableau 10 présente un vecteur représentatif de l'appréciation de chacune des caractéristiques.

Tableau 9 : Vecteur représentant le profil du client à la troisième itération

PC	0.5	0	0.65	0.65	0	0.35	0.65	0.15
----	-----	---	------	------	---	------	------	------

Étape 6 : Tester le critère d'arrêt

Trois itérations effectuées, $n = 3$,

$n > 3$, $n < 5$, donc on calcule la similarité entre PC et l'ensemble S, Sim (PC, S) présenté ci-dessous dans le tableau 11.

Nous avons deux produits ayant une similarité de 80% ou plus au profil du client **Sim (PC, S) > 0**,

Le critère d'arrêt est rempli.

Passer à la **Phase 4**.

Tableau 10 : Similarité entre le profil du client et l'ensemble S.

	CP
00A0112	0.9711632
00E0114	0.8919017
00R0110	0.6716605
00R0111	0.6398426
00P0115	0.6243192
00E0113	0.5149397
00W0117	0.502923
00O0116	0.3877834

Phase 4 : Recommandation d'un ensemble de produits

On calcule la similarité entre le sous-ensemble de produits S-R et le profil du client. Le sous-ensemble S-R est le même que le sous-ensemble S. La similarité entre PC et S-R sont équivalentes aux similarités représentées au tableau 11.

Les W=3 produits de la liste Sim (PC, S-R) ayant la similarité la plus élevée avec le profil du client sont présentés ci-dessous :

00A0112, 00E0114, 00R0110

Cet ensemble de produits représente la recommandation qui sera proposée au client. Ici, si l'ensemble de produits S était différent de l'ensemble S-R, des produits qui n'ont pas été présentés au client en magasin auraient pu être recommandé si la similarité d'un de ses produits avec le profil du client faisait partie des W premier produit de la liste ordonnée de recommandation.

5 CONCLUSION

Lors du processus d'achat, le client peut se retrouver face à un éventail de choix qui rend la prise de décision longue et pénible. Les systèmes de recommandation ont été mis en place, entre autres, afin d'aider l'utilisateur face au problème de surcharge d'information. Cependant, malgré la disponibilité de données sur les produits proposés, on peut se retrouver face à un manque d'information sur les utilisateurs, surtout dans un contexte de vente stationnaire où il n'existe généralement pas de données sur les comportements des utilisateurs. La méthode proposée dans cet article permet de créer une base d'apprentissage sur les préférences des utilisateurs à travers l'élaboration d'une méthode de recommandation interactive sur 4 phases. À travers la collecte de rétroaction explicite (Phase de filtrage) ou implicite (Phase d'exploration interactive et création du profil client), il est possible de pallier au problème du démarrage à froid.

La création de profils de clients en plus de déjouer le démarrage à froid, permet de générer une base d'apprentissage qui permettra dans le futur de faire des recommandations de plus en plus performantes en utilisant des méthodes de recommandation plus élaborées, tenant compte des similarités entre les clients. Cette base d'apprentissage peut également être utilisée pour étudier les ventes manquées et mieux prédire la demande et établir les caractéristiques de produits que les clients auraient apprécié trouver dans un même produit.

Le problème pour déterminer la demande est souvent lié à un manque d'information liée aux ventes perdues. En mettant en

place un système de recommandation interactif, il sera possible de recueillir de l'information sur les ventes potentiellement manquées et les produits qui auraient pu être présentés au client.

La méthode de création du profil du client que nous proposons se base sur l'hypothèse que l'appréciation que l'utilisateur donne à un produit est représentative de son appréciation de chacun des attributs de celui-ci. Cette hypothèse, malgré le fait qu'elle soit rattrapée par l'évaluation des appréciations des attributs à partir de la moyenne de plusieurs produits, pourrait être améliorée par la collecte d'appréciations pour chacun des attributs des produits présentés au client. En effet, le client pourrait fortement apprécier les sièges et le moteur, mais détester la couleur du véhicule. Cette amélioration dépend du temps que le client est prêt à investir pour son achat.

Aussi, à l'étape 1, nous devons évaluer l'influence du choix du premier produit sur la rapidité avec laquelle on acquière de l'information suffisante pour proposer un produit qui sera intéressant pour l'utilisateur. De même, le parcours d'un produit à l'autre pourra être bonifié en considérant l'éventail des caractéristiques présentés ou non au client.

Notons également que trouver un compromis entre le nombre de produits à présenter au client pour avoir un aperçu de ses préférences et la surcharge d'information peut mener à une couverture incomplète des attributs présentés aux clients à travers les produits évalués. À l'avenir, il faudra tenir compte du gain informationnel des attributs et leurs impacts sur la prise de décision des utilisateurs lors du choix du sous-ensemble de produit à présenter au client.

D'autres variables entrent en jeu pour la phase de déploiement de l'outil proposé par les vendeurs. Une résistance au changement peut résulter lors de la mise en place de ce système en fonction du profil du vendeur. Les profils des vendeurs ainsi que leurs stratégies de vente sont différents. Certains vendeurs favoriseront la satisfaction du client à long terme au dépend du profit monétaire ponctuel résultant de la vente. D'autres, considérerons le profit instantané comme étant la priorité absolue. Cette différence de comportement aura un impact direct sur l'utilisation de l'outil de recommandation de produits. Certains vendeurs pourraient être retissant lorsque les recommandations proposées par la méthode ne sont pas celles qui maximisent leurs profits.

6 REMERCIEMENTS

Nous tenons à remercier nos partenaires industriels, le CRSNG (RDCPJ 492021-15) ainsi que le MITACS (FR29950) pour le soutien financier qu'ils ont mis à notre disposition. Nous tenons également à remercier le « Physical Internet Center », le « Supply Chain and Logistics Institute, Coca-Cola Chair in Material Handling and Distribution » sans qui le projet n'aurait pu voir le jour.

7 REFERENCES

- Adomavicius, G., et Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering, 17(6), pp. 734-749.
- Amatriain, X., Jaimes, A., Oliver, N., et Pujol, J.M. (2015). Chapter 2: Data Mining Methods for Recommender Systems. In Recommender systems handbook, pp. 42-43. Springer, Boston, MA.
- Bhattacharyya, A. (1946). On a measure of divergence between two

- multinomial populations. *Sankhyā: the indian journal of statistics*, pp. 401-406.
- Bobadilla, J., Ortega, F., Hernando, A., et Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, 46, pp. 109-132.
- Bucklin, R. E., Sismeiro, C. (2009). Click here for Internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive marketing*, 23(1), pp. 35-48.
- Dadouchi, C., Agard, B. (2017). État de l'art sur les systèmes de recommandation, 12ème Congrès International de Génie Industriel – GI 2017, Compiègne, France, 3-5 mai, 2017.
- He, C., Parra, D., et Verbert, K. (2016). Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56, pp. 9-27.
- Jamali, M., et Ester, M. (2010). A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*, pp. 135-142. ACM.
- Lathia, N., Hailes, S., et Capra, L., (2008) The effect of correlation coefficients on communities of recommenders. In *SAC '08: Proceedings of the 2008 ACM symposium on Applied computing*, pp. 2000–2005, New York, NY, USA. ACM.
- Lika, B., Kolomvatsos, K., et Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4), pp. 2065-2073.
- Lu, J., Wu, D., Mao, M., Wang, W., et Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, 74, pp. 12-32.
- Montreuil, B., et Derhami, S. (2019). Beyond inventory position: Assessing product availability across a network of interconnected customer-centric retailers. Unpublished Manuscript.
- Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., et Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, 27(2), pp. 159-188.
- Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression heredity and panmixia, *Philos. Trans. R. Soc. London*, 187, pp. 253-318.
- Pozo, M., Chiky, R., et Metais, E. (2015). Extraction de l'intérêt implicite des utilisateurs dans les attributs des items pour améliorer les systèmes de recommandations. In *15ème conférence internationale sur l'extraction et la gestion des connaissances (EGC 2015)*, pp. 329-334. RNTI.
- Rajamma R.K., Paswan A.K., Ganesh G. (2007) Emerald Article: Services purchased at brick and mortar versus online stores, and shopping motivation. *J Serv Mark* 21, pp. 200-212.
- Rashid, A. M., Karypis, G., et Riedl, J. (2008). Learning preferences of new users in recommender systems: an information theoretic approach. *ACM Sigkdd Explorations Newsletter*, 10(2), pp. 90-100.
- Schafer, J. B., Frankowski, D., Herlocker, J., et Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web*, pp. 291-324. Springer, Berlin, Heidelberg.