



## Une méthodologie de recommandations produits fondée sur l'actionnabilité et l'intérêt économique des clients

Thomas Piton, Julien Blanchard, Fabrice Guillet, Henri Briand

### ► To cite this version:

Thomas Piton, Julien Blanchard, Fabrice Guillet, Henri Briand. Une méthodologie de recommandations produits fondée sur l'actionnabilité et l'intérêt économique des clients. Ali Khenchaf et Pascal Poncelet. Extraction et gestion des connaissances (EGC'2011), Jan 2011, Brest, France. Hermann, E-20, pp.203-214, 2011, Revue des Nouvelles Technologies de l'Information. <hal-00555796>

**HAL Id: hal-00555796**

**<https://hal.archives-ouvertes.fr/hal-00555796>**

Submitted on 14 Jan 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Une méthodologie de recommandations produits fondée sur l'actionnabilité et l'intérêt économique des clients

Thomas Piton<sup>\*,\*\*</sup>, Julien Blanchard<sup>\*</sup>, Fabrice Guillet<sup>\*</sup>, Henri Briand<sup>\*</sup>

\* LINA équipe COD - UMR 6241 CNRS, 2 rue de la Houssinière, 44322 Nantes  
{prenom.nom}@univ-nantes.fr, <http://www.polytech.univ-nantes.fr/COD>

\*\* Groupe VM Matériaux, Route de la Roche sur Yon, 85260 L'Herbergement  
{tpiton}@vm-materiaux.fr, <http://www.vm-materiaux.fr/>

**Résumé.** Dans un contexte économique difficile, la fidélisation des clients figure au premier rang des préoccupations des entreprises. En effet, selon le *Gartner*, fidéliser des clients existants coûterait beaucoup moins cher que prospecter de nouveaux clients. Pour y parvenir, les entreprises optimisent la marge et le cycle de vie des clients en développant une relation personnalisée aboutissant à de meilleures recommandations. Dans cet article, nous proposons une méthodologie pour les systèmes de recommandations fondée sur l'analyse des chiffres d'affaires des clients sur des familles de produits. Plus précisément, la méthodologie consiste à extraire des comportements de référence sous la forme de règles d'association et à en évaluer l'intérêt économique et l'actionnabilité. Les recommandations sont réalisées en ciblant les contre-exemples les plus actionnables sur les règles les plus rentables. Notre méthodologie est appliquée sur 12 000 clients et 100 000 produits de VM Matériaux afin d'orienter les commerciaux sur les possibilités d'accroissement de la valeur client.

## 1 Introduction

Face au contexte économique actuel, de nombreuses entreprises sont amenées à réévaluer leur stratégie. Bon nombre d'entre elles cherchent à recentrer leurs services sur les besoins de leurs clients et à établir des relations personnalisées (Venkatesan, 2004). La Gestion de la Relation Client (ou CRM pour *Customer Relationship Management*) vise à comprendre et à influencer le comportement des clients au moyen de communications utiles, en vue d'améliorer la fidélisation et la rentabilité des clients (Ling et Yen, 2001). Dans ce contexte, les systèmes de recommandations (Resnick et Varian, 1997) permettent d'offrir des suggestions personnalisées et adaptées aux besoins du client en fonction de son comportement (préférences, historique, etc.).

En pleine émergence aujourd'hui, l'utilisation de la fouille de données pour gérer la relation client se concentre principalement sur l'étude des comportements des clients au travers de nombreuses applications (Kenneth D. Lawrence, 2007) telle que l'optimisation des campagnes marketing (Piton et al., 2009). De plus, la rentabilité d'un

## Des recommandations actionnables et économiquement intéressantes

client peut être développée par des ventes croisées (*cross selling*) ou des montées en gammes (*up selling*) : les systèmes de recommandations s'appuyant sur des techniques de fouille de données ont pour objectif d'y répondre.

Cependant, ces systèmes réalisent majoritairement des recommandations fondées sur le produit acheté par le client au moment présent (Adomavicius et Tuzhilin, 2005). Ce modèle convient parfaitement aux sites de commerce en ligne, mais s'avère très réducteur pour les processus de vente assistés par un commercial. Nous proposons d'adopter une philosophie différente : dans le cadre d'un démarchage par un commercial, quel(s) produit(s) recommander au client pour déclencher l'acte d'achat ? Effectuer des recommandations de ce type nécessite (a) d'identifier des comportements typiques d'achat sur le long terme, et (b) de les appliquer au client en veillant à respecter son comportement. Le risque de mauvaise recommandation s'avère plus élevé dans le cadre d'un démarchage commercial que d'un site Internet (Linden et al., 2003). En effet, le commercial peut refuser d'utiliser le système s'il ne juge pas les recommandations suffisamment pertinentes. Ensuite, de nombreux systèmes de recommandations sont fondés sur des règles d'association ou séquentielles (Liu et al., 2009). Si une règle  $A \rightarrow B$  est exploitée, alors le principe est le suivant : lorsqu'un client s'intéresse au produit  $A$ , on lui propose  $B$ . La limite d'un tel raisonnement est qu'il ne prend pas en compte les différences qui peuvent exister entre la population des exemples de la règle (clients qui achètent  $A$  et  $B$ ), et la population des contre-exemples (clients qui achètent  $A$ , mais pas  $B$ ). Plus ces populations diffèrent, moins il y a de raisons que ce qui s'appliquait aux exemples puissent s'adapter facilement aux contre-exemples.

Dans cet article, nous proposons une méthodologie pour les systèmes de recommandations fondée sur l'analyse des chiffres d'affaires des clients sur des produits. Plus précisément, la méthodologie consiste à extraire des comportements de référence sous la forme de règles d'association et à en évaluer l'intérêt économique et l'actionnabilité. Les recommandations sont réalisées en ciblant les contre-exemples actionnables sur les règles les plus rentables. Les contributions principales de ce papier sont :

- Nous proposons une méthodologie pour la recommandation de produits à partir de chiffres d'affaires.
- Nous présentons une nouvelle mesure de l'actionnabilité des règles pour la recommandation, fondée sur la similarité entre les exemples et les contre-exemples.
- Nous proposons une mesure originale de l'intérêt économique des règles d'association pour la recommandation, fondée sur des critères définis par les experts métier.
- Nous nous intéressons aux règles d'association non pas seulement en tant qu'implication révélatrice d'un comportement d'achat, mais davantage comme modèle détecteur de contre-exemples à prospecter.
- Nous expérimentons notre méthodologie sur une base de données composée de plus de 12 000 clients et 100 000 produits.

La suite de l'article est structurée comme suit : en premier lieu, nous présentons l'utilisation des règles d'association dans le contexte des systèmes de recommandations. En second lieu, nous proposons notre méthodologie pour l'extraction de comportements de référence sous la forme de règles évaluées préalablement selon leur intérêt économique et leur actionnabilité. Enfin, nous appliquons notre méthodologie sur un jeu de données réel de VM Matériaux.

## 2 État de l'art : recommandations et règles d'association

La croissance explosive d'Internet et l'émergence du commerce électronique ont conduit à l'essor des systèmes de recommandations (Resnick et al., 1994). Ces derniers permettent aux entreprises de filtrer l'information, puis de recommander de manière proactive des produits à leurs clients en fonction de leurs préférences. Recommander des produits et des services peut renforcer la relation entre l'acheteur et le vendeur, et donc augmenter les bénéfices (Zhang et Jiao, 2007). Ces dernières années sont révélatrices de l'utilisation des systèmes de recommandations sur le Web à travers les films<sup>1</sup>, les livres<sup>2</sup> et la musique<sup>3</sup>. Schafer et al. (2001) présentent une classification détaillée des systèmes de recommandations dans le commerce électronique, et élucident la façon dont ils peuvent être utilisés pour fournir un service personnalisé fidélisant le client. De nombreux systèmes élaborés tel que *TrustWalker* (Jamali et Ester, 2009) soulignent l'intérêt actuel de la recommandation. Actuellement, les moteurs de recommandation reposent sur deux paradigmes :

- L'un basé sur le contenu (*content-based filtering*) permettant de construire un profil thématique pour le client (composé par exemple de mots-clefs) afin de comparer les caractéristiques de chaque produit avec ce profil.
- L'autre sur la collaboration (*collaborative filtering*) exploitant l'historique des actions des clients (comportement d'achats) pour prédire la note d'un nouveau produit pour un client sur la base de ce que les clients proches de lui en ont pensé.

Ces approches s'appuient généralement sur des méthodes d'apprentissage ou de fouille de données. Les règles d'association sont devenues application courante des systèmes de recommandations (Leung et al., 2008). Leung et al. (2006) présentent un modèle basé sur les règles d'association floues en tirant parti des similitudes de produits dans des taxonomies produits existantes. Liu et al. (2009) proposent d'extraire des comportements d'achats de clients à l'aide de règles d'association séquentielles, prenant ainsi en compte le comportement d'achat du client dans le temps. La plupart des systèmes de recommandations fondés sur des règles se concentrent principalement sur l'intérêt actuel du client (métaphore du panier de la ménagère) et non sur son historique, s'adaptant seulement à des comportements d'achats éphémères.

## 3 Notre méthodologie

Notre méthodologie, instanciée dans la section 4, se décline en quatre étapes :

1. L'extraction de comportements d'achats sous forme de règles (Section 3.1)
2. La mesure de l'actionnabilité des contre-exemples (Section 3.2)
3. La mesure de l'intérêt économique des cohortes de règles (Section 3.3)
4. Les recommandations actionnables et rentables (Section 3.4)

Tout au long de notre méthodologie, des exemples issus du contexte de VM Matériaux, i.e. de la vente de matériaux de constructions, illustrerons notre démarche.

---

1. <http://www.netflix.com>

2. <http://www.amazon.com>

3. <http://www.last.fm>

Des recommandations actionnables et économiquement intéressantes

### 3.1 Extraction de comportements d'achats sous forme de règles

#### 3.1.1 Préparation des données

- Soit, un ensemble  $C$  de  $n$  clients  $\{C_1, C_2, \dots, C_n\}$  décrits par deux types de variables :
- Un ensemble  $CA$  de  $m$  variables numériques  $\{CA_1, CA_2, \dots, CA_m\}$  relatives au comportement d'achats sur des Chiffres d'Affaires (CA) extraits sur un ensemble  $P$  de  $m$  produits disjoints  $\{P_1, P_2, \dots, P_m\}$ .
  - Un ensemble  $D$  de  $q$  variables numériques ou catégoriques  $\{D_1, D_2, \dots, D_q\}$  caractéristiques du profil client.

**Exemple 1** Les variables  $CA$  représentent des chiffres d'affaires de clients sur les familles de produits telles que les ciments gris, les ciments blancs, la chaux hydraulique, la chaux aérienne ou les liants colles. Les clients artisans  $C$  peuvent être décrits par un profil  $D$  composé de caractéristiques telles que le métier, le nombre de salariés, l'encours assuré, la région, etc.

Ensuite, la prochaine étape consiste à rendre commensurable pour chaque client de  $C$ , ses chiffres d'affaires  $CA$  en transformant la valeur effective en euros en proportion de chiffre d'affaires. Soit  $CA(c, p)$  avec  $c \in C$  et  $p \in P$ , le chiffre d'affaires réalisé par le client  $c$  pour le produit  $p$ . Nous définissons la part de chiffre d'affaires  $CA\%$  du client  $c$  pour le produit  $p$  comme suit :

$$CA\%(c, p) = \frac{CA(c, p)}{\sum_{p \in P} CA(c, p)} \quad (1)$$

Enfin, la dernière étape de pré-traitement consiste à discrétiser les  $CA\%(p)$  en un ensemble  $J(p)$  de  $q$  intervalles de valeurs. Sans nuire à la généralité de l'approche, nous illustrerons avec trois intervalles distincts et de même fréquence ( $q = 3$  et  $J(p) = \{a(p), b(p), c(p)\}$ ). Pour chaque produit  $p$ , les intervalles  $a(p)$ ,  $b(p)$  et  $c(p)$  identifient trois niveaux de part de chiffre d'affaires : faible, moyen et fort.

**Exemple 2** Soient deux entreprises de maçonnerie  $c_1$  et  $c_2$  réalisant 100 000 € et 1 000 € de chiffre d'affaires dans la famille de produits  $p_1$  des ciments.  $c_1$  et  $c_2$  réalisent 1 000 000 € et 5 000 € de chiffre d'affaires total, i.e.  $CA\%(c_1, p_1) = 10\%$  et  $CA\%(c_2, p_1) = 20\%$ . Pour la famille de produits  $p_1$  :  $a(p_1) = [0; 10\%]$ ,  $b(p_1) = [11; 40\%]$  et  $c(p_1) = [41; 100\%]$ . Nous pouvons en déduire que  $CA\%(c_1, p_1) \in a(p_1)$  et  $CA\%(c_2, p_1) \in b(p_1)$ . Par conséquent,  $c_1$  achète faiblement et  $c_2$  moyennement dans la famille des ciments.

#### 3.1.2 Extraction de règles d'association

Étant donné un produit  $p$ , un item est une variable booléenne de la forme  $CA\%(p) \in v$  où  $v \in J(p)$ . Une règle d'association (Agrawal et al., 1993) est une liaison orientée de la forme  $X \rightarrow Y$ , où  $X$  et  $Y$  sont des itemsets qui n'ont pas d'items en commun. Chaque règle est caractérisée par deux mesures : le support et la confiance. Dans notre méthodologie, nous notons abusivement  $p_1 = b \rightarrow p_2 = c$  pour la règle d'association :

$$CA\%(p_1) \in b(p_1) \rightarrow CA\%(p_2) \in c(p_2) \quad (2)$$

où  $p_1$  et  $p_2$  sont des produits achetés dans des intervalles respectifs  $b(p_1)$  et  $c(p_2)$ .

**Définition 1** *Les clients exemples d'une règle sont les individus qui vérifient la prémisse et la conclusion. En revanche, les clients contre-exemples sont les individus qui vérifient la prémisse et non la conclusion à hauteur du chiffre d'affaires souhaité.*

Remarquons qu'un client peut être contre-exemple de une ou plusieurs règles d'association présentant la même conclusion. Pour éviter de recommander plusieurs fois le même produit à un même client, nous introduisons la notion de *cohorte* de règles.

**Définition 2** *Étant donné un produit  $p$  acheté à hauteur de  $v$ ,  $v \in J(p) = \{a(p), b(p), c(p)\}$ , une cohorte est un ensemble de règles d'association concluant sur le même niveau d'achat  $p = v$ , i.e sur la même recommandation. L'ensemble des exemples d'une cohorte, noté  $C^+$ , est l'union des ensembles d'exemples des règles de la cohorte. L'ensemble des contre-exemples d'une cohorte, noté  $C^-$ , est l'union des ensembles des contre-exemples des règles de la cohorte. L'ensemble des règles extrait est partitionné en autant de cohortes que de conclusion de règles.*

**Définition 3** *Étant donné une cohorte  $R$  extrait par l'algorithme, une  $R$ -Recommandation consiste à proposer la conclusion de la cohorte à un client contre-exemple.*

**Exemple 3** *Considérons la cohorte  $R$  composée de l'unique règle Ciments gris =  $c$  et Sable de Rivière =  $c \rightarrow$  Graviers de Carrière =  $c$ . 177 clients artisans vérifient cette règle (exemples) et 50 sont contre-exemples. Une  $R$ -Recommandation consiste à proposer aux 50 clients contre-exemples d'acheter du gravier de carrière à hauteur de  $c$ .*

## 3.2 Mesure de l'actionnabilité des contre-exemples

Une analyse factorielle permettant d'inclure à la fois des variables numériques et catégoriques en tant qu'éléments actifs d'une même analyse a été proposée par ESCOPIER en 1979 dans le cadre de l'Analyse des Correspondances Multiples. Pagès (2004) souligne que cette approche se confondait avec les travaux de SAPORTA de 1990 autour de l'Analyse en Correspondances Principales. L'ensemble de ces points de vue confère une méthode à part entière dotée de plusieurs bonnes propriétés : l'Analyse Factorielle de Données Mixtes (AFDM). Pour mesurer l'actionnabilité des clients contre-exemples dans un espace à variables mixtes (numériques et catégoriques), nous utilisons la mesure de distance de l'AFDM. Le principe général est le suivant : un contre-exemple est considéré comme actionnable s'il ressemble suffisamment aux exemples.

### 3.2.1 Variables d'achats

Afin d'être semblable aux exemples  $C^+$  d'une cohorte, un contre-exemple  $e^-$  doit présenter des parts de CA non aberrantes sur l'ensemble des prémisses de la cohorte.

**Définition 4** *Étant donné la répartition des sommes de CA% des exemples  $C^+$  sur les prémisses de la cohorte, nous définissons  $M$  la médiane et  $(Q_3 - Q_1)$  l'intervalle interquartile de la distribution. Un contre-exemple, noté  $e^-$ , appartenant à  $C^-$  est pré-actionnable si :*

$$M - (Q_3 - Q_1) \leq \sum_{\substack{p \in \text{premisses} \\ \text{de la cohorte}}} CA\%(e^-, p) \leq M + (Q_3 - Q_1) \quad (3)$$

Des recommandations actionnables et économiquement intéressantes

**Exemple 4** Soit la règle *Ciments gris = c* et *Sable de Rivière = c* → *Graviers de Carrière = c*. Les achats en prémisses des exemples de  $C^+$  sont caractérisés par  $M = 40\%$  et  $Q_3 - Q_1 = 20\%$ . Pour être pré-actionnable, un contre-exemple de  $C^-$  devra avoir acheté entre 20% et 60% de sa part de CA dans les ciments gris et le sable de rivière.

### 3.2.2 Variables caractéristiques des profils clients

Les caractéristiques des profils clients sont séparées en deux groupes distincts :

- Les variables numériques  $DN = \{DN_1, DN_2, \dots, DN_s\}$  centrées et réduites.
- Les variables catégoriques  $DC = \{DC_1, DC_2, \dots, DC_r\}$  sur lesquelles est appliqué un codage disjonctif complet générant un ensemble d'indicatrices  $I = \{I_1, I_2, \dots, I_t\}$  (une variable binaire par modalité). Les valeurs de ces indicatrices sont divisées par la racine carrée de la fréquence  $F$  de chaque indicatrice, notée  $\sqrt{F(I_j)}$ .

**Exemple 5** Les couvreurs et les carreleurs achètent de nombreux produits en commun. Ils peuvent être rapprochés par leurs achats (cf. section 3.2.1) mais éloignés par les caractéristiques de leurs profils. Par exemple, les carreleurs et couvreurs ont tendances à acheter régulièrement des sacs de ciments gris de 25kg pour couler une chape ou installer une faîte de toiture. Pour la règle *Sac de ciment* → *Carrelage*, la majorité des exemples sont des carreleurs. Si des contre-exemples de la règle sont carreleurs et couvreurs, vous ne recommanderez qu'aux carreleurs l'achat du produit carrelage.

Ensuite, la distance de l'AFDM est appliquée aux clients  $c_1$  et  $c_2$  dans l'espace des variables caractéristiques du profil :

$$d^2(c_1, c_2) = \sum_{i=1}^s (DN_{i,c_1} - DN_{i,c_2})^2 + \sum_{j=1}^t (I_{j,c_1} - I_{j,c_2})^2 \quad (4)$$

**Définition 5** Un contre-exemple  $e^-$  est actionnable si il est suffisamment proche des exemples. Étant donné deux seuils  $minDist$  et  $\delta$ ,  $e^-$  est considéré comme  $\delta$ -actionnable vis-à-vis d'une cohorte de règles  $R$  si :

1.  $e^-$  est pré-actionnable (cf section 3.2.1).
2.  $e^-$  respecte l'inéquation suivante :

$$\frac{|\{e^+ \in C^+ \mid d(e^-, e^+) \leq minDist\}|}{|C^+|} \geq \delta \quad (5)$$

avec  $minDist$ , un seuil déterminant un voisinage autour du contre-exemple  $e^-$ . Compte tenu de la distribution des exemples dans l'espace des caractéristiques des clients, la distance d'un contre-exemple au barycentre des exemples n'est pas forcément représentative. C'est pourquoi, nous n'utilisons pas le barycentre dans notre mesure.

### 3.3 Mesure de l'intérêt économique des cohortes de règles

L'intérêt économique d'un contre-exemple correspond à la rentabilité en euros de la R-Recommandation appliquée. Pour chaque contre-exemple actionnable (cf. partie 3.2 de notre méthodologie), nous calculons la somme de CA en euros qui aurait dû être dépensée si le contre-exemple s'était comporté comme un exemple de la cohorte.

**Définition 6** *Étant donné un seuil  $\theta \in \mathbb{R}^+$ , un contre-exemple  $e^-$  est défini comme  $\theta$ -rentable pour le produit  $p$  si :*

$$(CA(e^-) * \mu) - CA(e^-, p) \geq \theta \quad \text{avec} \quad \mu = \frac{\sum_{c \in C^+} CA\%(c, p)}{|C^+|} \quad (6)$$

Nous pouvons en déduire que la rentabilité d'une cohorte de règles correspond à la somme de la rentabilité des contre-exemples  $\theta$ -rentables de la cohorte. De cet intérêt économique, pourront être déduits les coûts fixes et variables engendrés par une démarche commerciale. Nous suggérons l'utilisation d'un *scoring* pour affiner l'intérêt économique espéré pour chaque client. La phase d'apprentissage est guidée par les exemples de la cohorte et la phase d'application par les contre-exemples actionnables.

### 3.4 Recommandations actionnables et rentables

Plusieurs possibilités s'offrent aux experts métier pour actionner le système de recommandation de manière économiquement intéressante :

1. Trier les cohortes pour identifier les R-Recommandations les plus intéressantes économiquement. Dès lors, des actions commerciales peuvent être déclenchées pour promouvoir des produits.
2. Trier pour chaque client contre-exemple l'ensemble des cohortes dans lesquelles il est  $\theta$ -rentable. Les experts métier pourront réaliser une ou plusieurs recommandations, créant ainsi une relation personnalisée avec le client ciblé.

## 4 Application

L'activité Négoce du groupe VM Matériaux<sup>4</sup> propose à ses clients professionnels plus de 100 000 références produits. Les clients artisans du bâtiment présentent des comportements d'achats différents et évolutifs ainsi que des caractéristiques de profil souvent atypiques. Dès lors, l'utilisation de notre méthodologie s'avère pertinente pour recommander des produits à forte valeur ajoutée à des clients susceptibles d'en acheter et ainsi créer une relation personnalisée et privilégiée avec le client.

---

4. <http://www.vm-materiaux.fr>



Des recommandations actionnables et économiquement intéressantes

#### 4.1 Préparation des données et extraction des règles d'association

Pour remédier au problème de *sur-spécialisation* (Balabanovic et Shoham, 1997), nous utilisons la taxonomie produits de VM Matériaux. L'intérêt de cette façon de procéder est d'obtenir des règles plus pertinentes, dans lesquelles les produits les plus courants ne dissimulent pas, par leur fréquence, les produits les moins courants.

**Exemple 6** *Les produits les plus rares tels que les panneaux photovoltaïques seront codifiés à un niveau plus général, tandis que les produits les plus courants comme la plaque de plâtre standard de 13 mm seront codifiés à un niveau plus fin.*

Ensuite, 12 102 clients sont sélectionnés dans l'entrepôt de données existant et décrits à travers 461 variables divisées en trois groupes :

1. 21 variables (numériques et catégoriques) internes à l'entrepôt de données telles que le potentiel annuel, le risque, la fidélisation, le magasin principal, etc.
2. 9 variables (numériques et catégoriques) externes au système d'information telles que le nombre d'employés, la catégorie socio-professionnelle, etc.
3. 431 variables de chiffre d'affaires calculées à partir des différents niveaux utilisés de la taxonomie produits.

Par la suite, en accord avec les experts métier, les variables de chiffre d'affaires sont rendues commensurables et sont discrétisées en trois intervalles  $a$ ,  $b$  et  $c$  de même fréquence  $\frac{1}{3}$ . L'extraction des règles d'association est réalisée à l'aide de l'algorithme CHARM<sup>5</sup> fondé sur la recherche en profondeur d'itemsets fermés (Zaki et Hsiao, 2002) avec un support de 0,004 (i.e. 50 clients) et une confiance de 50 %. À l'aide des mesures objectives pré-sélectionnées, un ensemble de 27 040 règles est extrait (Cf. tableau 1) et partitionné en 180 cohortes de règles d'association.

Cohortes	Support	Confiance	Lift	Règles
1	0.51 %	53.45 %	11.91	SF_COLLE EN PATE = c ET F_CROISILLONS = c → F_OUTILS CARRELAGE = c
2	0.74 %	61.38 %	24.52	SF_LITEAUX SAPIN = c ET SF_TUILES ROMANES = c → SF_ACC TUILES ROMANES = c
3	0.64 %	71.96 %	12.60	SF_CIMENTS GRIS = c ET SF_GRAVIERS RIVIERE = c → SF_SABLE DE CARRIERE = c
2	0.49 %	57.28 %	22.88	SF_CIMENTS GRIS = b ET SF_TUILES ROMANES = c → SF_ACC TUILES ROMANES = c
1	0.64 %	53.42 %	11.91	SF_MORTIERS COLLES = c ET SF_TALOCHE_SCEAU = c → F_OUTILS CARRELAGE = c

TAB. 1 – Quelques règles d'association extraites par l'algorithme CHARM.

Pour la suite de l'expérimentation, nous nous concentrons sur la cohorte présentée dans le tableau 2 permettant de recommander l'achats de plâtres ( $GF\_PLATRES = c$ ). La cohorte est composée de 18 règles d'association, de 210 exemples et 193 contre-exemples.

5. Codé en C++ par son concepteur Mohammed J. ZAKI, <http://www.cs.rpi.edu/~zaki/software/>

Support	Confiance	Lift	Règle
0.61 %	52.86 %	14.00	F_COLLES ET ENDUITS CARREAUX PLATRE E = c → GF_PLATRES = c
1.50 %	66.42 %	17.59	GF_CLOISONS ET PLAFONDS BRIQUES = c → GF_PLATRES = c
0.40 %	51.61 %	13.67	F_PROFILES DE FINITION POUR CLOISONS = b → GF_PLATRES = c
0.46 %	94.92 %	25.13	F_COLLES ENDUITS CARREAUX PLÂTRE = c ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.53 %	74.42 %	19.71	SF_CIMENTS GRIS = a ET GF_CLOISONS ET PLAFONDS BRIQUES = c → GF_PLATRES = c
0.41 %	79.37 %	21.02	SF_LAINE VERRE ROULEAUX COMBLES = b ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.40 %	66.22 %	17.53	SF_TRANSPORT MATERIAUX = b ET GF_CLOISONS ET PLAFONDS BRIQUES = c → GF_PLATRES = c
0.46 %	80.00 %	21.19	SF_PLAQUE DE PLATRE STANDARD = b ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.50 %	82.19 %	21.77	SF_RAILS ET MONTANTS METALLIQUES 48 = b ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.61 %	79.57 %	21.07	SF_COLLES ET ENDUITS POUR PLAQUES = c ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.45 %	78.26 %	20.72	F_FOURRURES ET AUTRES OSSATURES = b ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.50 %	76.92 %	20.37	SF_COLLES ET ENDUITS POUR PLAQUES = b ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.61 %	75.51 %	20.00	SF_PLAQUE DE PLATRE STANDARD = c ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.43 %	86.67 %	22.95	F_PLAQUE DE PLATRE HYDROFUGE = b ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.45 %	73.33 %	19.42	SF_RAILS ET MONTANTS METALLIQUES 48 = c ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.45 %	82.09 %	21.74	SF_LAINE VERRE ROULEAUX COMBLES TOIT=c ET GF_CLOISONS PLAFONDS BRIQUES = c → GF_PLATRES = c
0.48 %	90.63 %	24.00	GF_CLOISONS PLAFONDS BRIQUES = c ET SF_LAINE VERRE PANNEAUX PAROIS INT = c → GF_PLATRES = c
0.40 %	81.67 %	21.63	GF_CLOISONS PLAFONDS BRIQUES = c ET SF_RUBANS ADHESIFS = c → GF_PLATRES = c

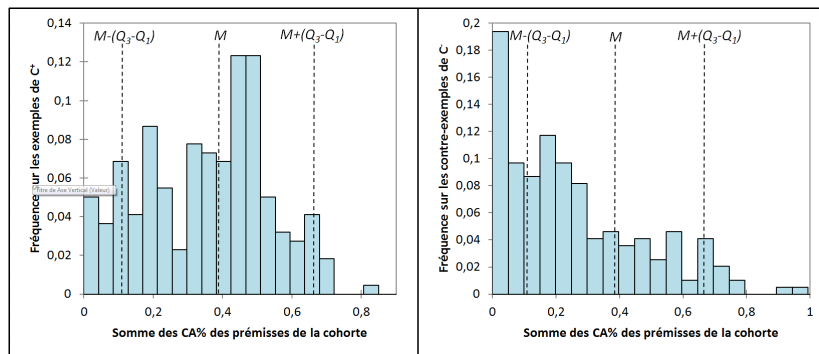
TAB. 2 – Une cohorte de 18 règles d'association recommandant l'achat de GF\_PLATRES = c.

## 4.2 Mesure de l'actionnabilité des contre-exemples

Afin de sélectionner les contre-exemples actionnables de la cohorte  $R$ , nous fixons empiriquement le paramètre  $minDist = 50$  et définissons avec la direction commerciale le seuil  $\delta = 10\%$ . Dès lors, un contre-exemple sera actionnable si et seulement si :

- Il est pré-actionnable, i.e. ses parts d'achats en prémisses de la cohorte sont dans l'intervalle défini dans la section 3.2.1 de notre méthodologie.
- Il possède 10 % des  $C^+$  dans son voisinage à une distance inférieure à 50.

Pour y parvenir, visualisons tout d'abord les distributions des exemples  $C^+$  et contre-exemples  $C^-$  sur les achats des prémisses de la cohorte (cf. figure 1).

FIG. 1 – Distribution des exemples et des contre-exemples sur les achats des prémisses de  $R$ .

## Des recommandations actionnables et économiquement intéressantes

Les achats en prémisses des  $C^+$  de la cohorte sont caractérisés par une médiane  $M = 37,32\%$  et un intervalle interquartile  $Q3 - Q1 = 28,05\%$ . Ces derniers sont reportés sur la distribution des  $C^-$ . Par conséquent, les  $C^-$  en dehors de l'intervalle  $[9,27\%; 65,37\%]$  sont considérés comme non pré-actionnables. Ce filtrage permet d'élaguer 36 contre-exemples pour la suite de notre méthodologie. Ensuite, pour être actionnable, les 157 contre-exemples pré-actionnables doivent posséder 10% de  $C^+$  dans leur voisinage à une distance inférieure ou égale à 50 (cf. section 3.2.2). Dès lors, ce deuxième filtrage permet de sélectionner 103 clients actionnables.

### 4.3 Mesure de l'intérêt économique de la cohorte de règles

À l'aide des experts métier (directeur marketing et équipe force de ventes) de VM Matériaux, nous avons considéré que les coûts fixes de démarchage d'un client par un commercial étaient de 350 €. Dès lors, nous avons pu fixer le seuil  $\theta = 1000$  € à partir duquel un contre-exemple est rentable ou économiquement intéressant pour l'entreprise. En appliquant la formule 6 de notre méthodologie, 79 contre-exemples actionnables sont 1000 €-rentables pour l'achat du groupe famille  $GF\_PLATRES = c$ . Un récapitulatif du filtrage des contre-exemples de la cohorte est illustré en figure 2.

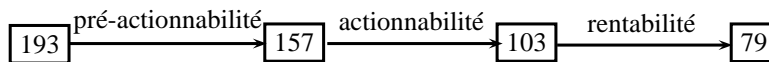


FIG. 2 – Récapitulatif du filtrage des contre-exemples.

Pour les 79 contre-exemples rentables, l'intérêt économique espéré pour la cohorte de règles est estimé à environ 600 000 € de chiffre d'affaires pour VM Matériaux.

### 4.4 Recommandations aux clients de VM Matériaux

Notre méthodologie a permis à l'équipe commerciale d'analyser le comportement de leurs clients de manière à comprendre les raisons pour lesquelles certains clients n'achètent pas certains produits (tarif, concurrence, etc.). De plus, le nombre restreint de recommandations rentables permet d'afficher des indicateurs intelligibles dans un outil de CRM pour aider les commerciaux à développer la valeur de leurs clients. Ensuite, des démarches davantage marketing peuvent être mises en place :

1. Promouvoir des produits sur des actions commerciales ciblées (produit dans le coffre des commerciaux) en sélectionnant les cohortes les plus intéressantes économiquement.
2. Cibler les clients d'une région ou d'un magasin et recommander trois à quatre produits en fonction des cohortes dans lesquelles les clients sont les plus intéressants économiquement.

Enfin, en fonction des recommandations économiquement intéressantes, des partenariats fournisseurs peuvent être convenus pour mettre en valeur des produits sur une période donnée, dans des régions ou magasins ciblés.

## 5 Conclusion et perspectives

Nous avons proposé dans cet article une méthodologie pour les systèmes de recommandations fondée sur l'analyse des chiffres d'affaires des clients sur des familles de produits. En mesurant l'actionnabilité des règles fondée sur la similarité entre les exemples et les contre-exemples, nous avons présenté une mesure originale de l'intérêt économique des règles d'association. Nous nous intéressons aux règles d'association non pas seulement en tant qu'implication révélatrice d'un comportement d'achat, mais davantage comme modèle détecteur de contre-exemples à prospector. L'application de notre méthodologie sur des données réelles de VM Matériaux a permis d'obtenir des recommandations actionnables et rentables à hauteur de plusieurs kilos d'euros, d'échanger avec les experts métier sur les raisons des non achats des clients et de préparer des actions commerciales ciblées sur un nombre de recommandations restreint. Cet article ouvre la perspective à d'autres travaux. En premier lieu, la mesure de l'intérêt économique peut être améliorée. En effet, l'actionnabilité d'une cohorte de règles peut être modulée par le nombre de contre-exemples actionnables, la marge dégagée sur les produits recommandés, le potentiel des clients, etc. En deuxième lieu, le développement d'une application est en cours pour intégrer la recommandation de règles au sein du système d'informations de l'entreprise.

**Remerciements.** Les auteurs remercient Pierrick RICHARD et Gaëtan BLAIN du groupe VM Matériaux pour leurs commentaires et interventions constructives.

## Références

- Adomavicius, G. et A. Tuzhilin (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), 734–749.
- Agrawal, R., T. Imielinski, et A. Swami (1993). Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Washington D.C., pp. 207–216.
- Balabanovic, M. et Y. Shoham (1997). Fab : Content-Based, Collaborative Recommendation. *Commun. ACM* 40(3), 66–72.
- Jamali, M. et M. Ester (2009). TrustWalker : a Random Walk Model for Combining Trust-Based and Item-Based Recommendation. In *KDD '09 : Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, pp. 397–406. ACM.
- Kenneth D. Lawrence, Stephan Kudyba, R. K. K. (2007). *Data Mining Methods and Applications*. Auerbach Publications.
- Leung, C. W.-k., S. C.-f. Chan, et F.-I. Chung (2006). A Collaborative Filtering Framework Based on Fuzzy Association Rules and Multiple-Level Similarity. *Knowledge and Information Systems* 10, 357–381.

Des recommandations actionnables et économiquement intéressantes

- Leung, C. W.-k., S. C.-f. Chan, et F.-l. Chung (2008). An Empirical Study of a Cross-Level Association Rule Mining Approach to Cold-Start Recommendations. *Knowledge-Based Systems* 21(7), 515 – 529.
- Linden, G., B. Smith, et J. York (2003). Amazon.com Recommendations : item-to-item Collaborative Filtering. *Internet Computing, IEEE* 7(1), 76–80.
- Ling, R. et D. C. Yen (2001). Customer Relationship Management : An Analysis Framework and Implementation Strategies. *The Journal of Comp. Inf. Syst.* 41, 82–97.
- Liu, D.-R., C.-H. Lai, et W.-J. Lee (2009). A Hybrid of Sequential Rules and Collaborative Filtering for Product Recommendation. *Information Sciences* 179(20), 3505–3519.
- Pagès, J. (2004). Analyse Factorielle de Données Mixtes. *Revue Statistique Appliquée LIII*, 93–111.
- Piton, T., J. Blanchard, H. Briand, et F. Guillet (2009). Domain Driven Data Mining to Improve Promotional Campaign ROI and Select Marketing Channels. In *CIKM '09 : Proc. of the 18th ACM Conf. CIKM*, New York, NY, USA, pp. 1057–1066. ACM.
- Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, et J. Riedl (1994). GroupLens : an Open Architecture for Collaborative Filtering of Netnews. In *CSCW '94 : Proceedings of the 1994 ACM conference on Computer supported cooperative work*, New York, NY, USA, pp. 175–186. ACM.
- Resnick, P. et H. R. Varian (1997). Recommender Systems. *Communications of the ACM* 40(3), 56–58.
- Schafer, B. J., J. A. Konstan, et J. Riedl (2001). E-Commerce Recommendation Applications. *Data Mining and Knowledge Discovery* 5(1 - 2), 115–153.
- Venkatesan, R., K.-V. (2004). A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal of Mark. Research* 68, 106–125.
- Zaki, M. J. et C.-J. Hsiao (2002). CHARM : An Efficient Algorithm for Closed Itemset Mining. In *Proceedings of the Second SIAM DM*, Arlington, VA. SIAM.
- Zhang, Y. et J. R. Jiao (2007). An Associative Classification-Based Recommendation System for Personalization in B2C e-commerce Applications. *Expert Syst. Appl.* 33(2), 357–367.

## Summary

In modern day management, there is a demand to optimize the clients life time value by developing a personal relationship between the company and its clients. This is often accomplished while taking account preferences and buying behavior into regard. In this paper we propose a method for recommending systems based on analysis of the turnover of customers for specific product families. The methodology extracts the typical behavior of customers in the form of association rules, the estimate of actionability and the economic interest. Recommendations are made by targeting the most actionable counter-examples on the most profitable association rules. Our methodology is applied to the customer relationship management in VM Matériaux company, helping the salesperson increase their customer value.