

Utilisation de l'intégrale de Choquet dans un système de recommandation objet basé sur un filtrage actif multicritères, mise en œuvre et perspectives : Application à la recommandation de jeans

Use of a Choquet Integral in an item-based recommendation system and a multicriteria active filtering, application to jeans recommendation and perspectives

Christophe Terrien¹

¹Laboratoire CRIISEA, Centre de Recherche sur les institutions, l'industrie et les systèmes économiques d'Amiens, France, christophe.terrien@u-picardie.fr

RÉSUMÉ. Bien qu'ancienne, l'intégrale de Choquet fait l'objet d'une littérature récente et abondante en recherche opérationnelle. Ces travaux ouvrent des perspectives d'applications intéressantes dans le domaine marketing, notamment dans la mise en œuvre des systèmes de recommandations s'appuyant sur des critères multiples. L'intégrale de Choquet est intéressante dans l'élaboration d'un modèle d'attitude car elle permet de prendre en compte les relations de dépendance entre critères. Ce papier propose une utilisation pratique de l'intégrale de Choquet dans un configurateur de choix ou configurateur de recommandations basé sur un filtrage actif multicritères.

ABSTRACT. Though far from recent, the Choquet integral is at present the subject of an abundant literature in operational research. That tool opens perspectives of interesting applications in the marketing field, in particular in the implementation of recommendations systems based on multiple criteria. The integral of Choquet is interesting in the development of an attitude model because it makes it possible to take into account the relations of dependence between criteria. This paper proposes a practical use of the Choquet integral in an item-based recommendation system based on a multicriteria active filtering.

MOTS-CLÉS. Systèmes de recommandation, Recommandation Objet, Configurateur, AHP, MACBETH, Intégrale de Choquet, Jeans, Analyse Hiérarchique, Analyse multicritères.

KEYWORDS. Recommendation system, item-based recommendation system, Configurator, AHP, MACBETH, Choquet integral, Jeans, Multi-criteria Analysis.

Introduction

Les systèmes de recommandation sont des systèmes de filtrage de l'information utilisés en webmarketing pour proposer aux utilisateurs des produits susceptibles de les intéresser. Les objectifs sont différents selon que l'on se trouve dans une relation B to C ou B to B par nature plus technique. Dans une relation B to C, il s'agit souvent d'améliorer le panier moyen (et, par conséquent, le chiffre d'affaires) en tentant une montée en gamme, c'est à dire en proposant des produits plus chers et/ou plus qualitatifs, ou de rechercher des ventes additionnelles en proposant des produits complémentaires ou, tout simplement de déclencher des achats impulsifs non prévus avant la navigation. Dans les relations B to B, les systèmes de recommandation privilégient le conseil aux utilisateurs. Il s'agit notamment d'éviter la commande de pièces incompatibles qui pénalisent les utilisateurs ou d'orienter les utilisateurs sur les solutions techniques adéquates. En B to C on retrouve cette démarche dans la vente de pièces détachées automobiles aux consommateurs finaux. La justesse de la recommandation crédibilise ou non le système et influence la satisfaction de l'utilisateur fréquentant le site, laquelle influence le réachat et, par ricochet, le chiffre d'affaires. Les systèmes à visée principalement commerciale s'appuient généralement sur une recom-

mandation sociale. On parle également de filtrage collaboratif. Le principe consiste à s'appuyer sur les achats des pairs (ou "proches voisins" dans le vocabulaire ad-hoc) pour établir les recommandations. Des algorithmes basés sur des calculs de distances ou de corrélations (ou k-moyennes...), établissent des segments d'utilisateurs dont on connaît les préférences et auxquelles l'utilisateur cible est associé. On suppose donc que les membres des segments ont des goûts similaires et que leurs préférences sont stables dans le temps. Il s'agit de l'approche dite "centré-utilisateur" (User-based nearest neighbor). Inversement, on peut chercher à regrouper les objets plutôt que les utilisateurs (item-based nearest neighbor). Les systèmes employés dans un contexte très technique, et donc, généralement B to B, se basent sur la recommandation objet. La recommandation ne repose plus sur les pratiques des pairs mais sur un profil de caractéristiques propres à l'objet. L'objet recommandé repose alors sur une similarité avec les attributs recherchés par l'utilisateur. L'exemple des mots-clé est souvent rapporté mais il peut s'agir de tout autres attributs comme le genre, le titre, l'auteur... Un livre est ainsi recommandé parce que l'on sait que l'utilisateur du site aime cet auteur, certains thèmes identifiés par des mots-clé, un genre de littérature... Les algorithmes de recommandation utilisés recherchent alors des profils (patterns) proches ou similarités. Des méthodes hybrides [10] sont désormais courantes. Elles sont basées sur une combinaison de l'approche sociale, objet, et personnalisée, laquelle se fonde uniquement sur les transactions passées de l'utilisateur lui-même. De ce fait, on tient compte de l'historique des achats de l'utilisateur, de ses pairs, et des caractéristiques des objets. Quelque soit le type de système de recommandation utilisé (social, objet, hybride) la fiabilité des prédictions dépend de l'aptitude de l'algorithme à rendre compte des préférences. Soit les préférences sont une sorte de boîte noire inaccessible et l'on s'en remet à des techniques statistiques. C'est le cas pour des calculs de similarités reposant sur les corrélations. Soit on cherche à modéliser les préférences en agrégeant des mesures de caractéristiques d'utilisateurs ou d'objets. Cette opération d'agrégation est souvent linéaire, conjonctive ou disjonctive. Elle peine à rendre compte de manière convaincante des mécanismes mentaux qui sous-tendent les choix car les attributs considérés sont, dans ce cas supposés indépendants. Ce qui est rarement le cas. Un prix élevé n'est pas forcément rédhibitoire pour un auteur que l'on adore, ou, la vitesse fera l'objet d'une attention particulière de la part d'un acheteur potentiel de Ferrari mais pas d'un acheteur potentiel d'une 2CV Citroën. Des effets de synergie ou au contraire d'antagonisme sont à l'œuvre lors de l'évaluation conjointe d'attributs. Pour remédier à ce problème des opérateurs non additifs comme l'intégrale de Choquet ou de Sugeno ont été suggérés. Reste à les rendre opérationnels dans un système de recommandation. C'est le propos de ce papier. Dans les systèmes de recommandation, les propositions se font par paquet sans distinction entre-elles. Les alternatives sont donc proposées avec plus ou moins d'acuités. Une analyse hiérarchique est donc associée à la proposition dont l'exposé se déroule de la façon suivante. Les concepts théoriques servant de support à la proposition sont d'abord présentés : la recommandation objet, l'analyse multicritères hiérarchique, l'intégrale de Choquet. La mise en œuvre pratique appliquée à la recommandation de jeans est ensuite détaillée. S'ensuit une analyse des limites et perspectives.

1. Bases théoriques

1.1. *La recommandation objet par filtrage actif*

Les systèmes de recommandation font l'objet de différentes approches. Les recommandations personnalisées (Personalized Recommendation) qui se fondent sur l'historique des transactions des utilisateurs,

les recommandations Objet (Content-based filtering) qui se fondent sur l'analyse des caractéristiques de l'objet, les recommandations sociales (ou collaboratives, Collaborative filtering) qui se fondent sur les consommations des pairs, les recommandations hybrides qui s'appuient sur un mix des précédentes. Les données utilisées peuvent être simplement observées par l'analyste (analyse des historique de navigation, d'achats...) ou déclarées par l'utilisateur à sa demande (feed-back sollicité). Dans le premier cas on parle de filtrage passif et l'on risque un biais déclaratif, dans le second cas on parle de filtrage actif et l'on risque un biais d'attribution ne sachant avec certitude qui est l'utilisateur réel (même IP pour toute la famille). Notre proposition de système de recommandation se place dans la catégorie des systèmes de recommandations Objet par filtrage actif. Les recommandations (dans l'exemple présent des jeans) aux shoppers, se feront à partir des caractéristiques propres aux objets et de la connaissance de la hiérarchie et du poids des critères utilisés par les shoppers. La recommandation objet présente des avantages et des inconvénients. Contrairement à la recommandation sociale, elle ne s'appuie pas sur des mesures statistiques nécessitant des données sur un nombre important d'utilisateurs. Un seul utilisateur suffit. Les évaluations des caractéristiques quantitatives (taille, poids, prix) peuvent être extraites automatiquement. A contrario, certaines caractéristiques subjectives sont inaccessibles sans interactions avec l'utilisateur (aspects esthétiques d'un objet par exemple). Les prédictions se basent sur un modèle d'attitude agrégeant des critères supposés être indépendants et qui font rarement l'unanimité. L'exhaustivité étant impossible et l'agrégation difficile, la fiabilité est aussi très relative. La recommandation objet se prête facilement à l'analyse multicritère tandis que la recommandation sociale s'établit sur la base d'une évaluation globale d'objets (notations de 0 à 5 étoiles).

1.2. *L'analyse multicritères hiérarchique*

L'analyse multicritère hiérarchique¹ tire son nom de l'existence d'une hiérarchie ordonnée des critères de manière similaire à la méthode du chaînage cognitif [5] ou chaînes moyens-fins. Des critères hiérarchiquement inférieurs (souvent des alternatives ou des actions) sont subordonnés à des critères supérieurs. Plusieurs dénominations sont utilisées pour désigner ces catégories de critères. Dans la théorie du chaînage cognitif les « moyens » sont subordonnés aux « fins » Olson (77) les sépare en attributs intrinsèques et extrinsèques, la théorie des niveaux de représentation [13] utilise la notion d'attribut concret et d'attribut abstrait. Pour recommander un objet, nous allons nous baser sur un modèle d'attitude prenant en considération des attributs de manière isolé ainsi que des combinaisons d'attributs. Bien souvent les attributs ne sont pas comparables directement. Ils sont souvent reliés entre eux parce qu'ils participent à la production d'un même bénéfice. Pour un jean, des attributs comme la couleur, la forme, la matière ou la marque servent à produire des bénéfices comme l'image que l'on renvoie aux autres ou le confort. L'attribut « matière » (imaginons de la soie) peut être utilisé pour produire un bénéfice en termes d'image et/ou un bénéfice « confort ». L'attribut est donc subordonné au bénéfice qu'il participe à produire. On peut donc concevoir à la manière de Saaty une structure hiérarchique (Fig. 1) pour un produit qui se décline en bénéfices puis en attributs enfin en différentes modalités. Nous appellerons les critères de premier niveau les bénéfices, ceux de deuxième niveau les attributs. Les attributs se déclinent ensuite en modalités. Ainsi, la forme d'un jean peut se définir par les modalités, « slim », « regular », « patte d'éléphant »... Un attribut peut être évalué différemment selon le bénéfice recherché. Par exemple dans le cas des jeans, la forme peut être vue par l'utilisateur comme un facteur d'esthétisme ou de confort.

1. A ne pas confondre avec la classification hiérarchique

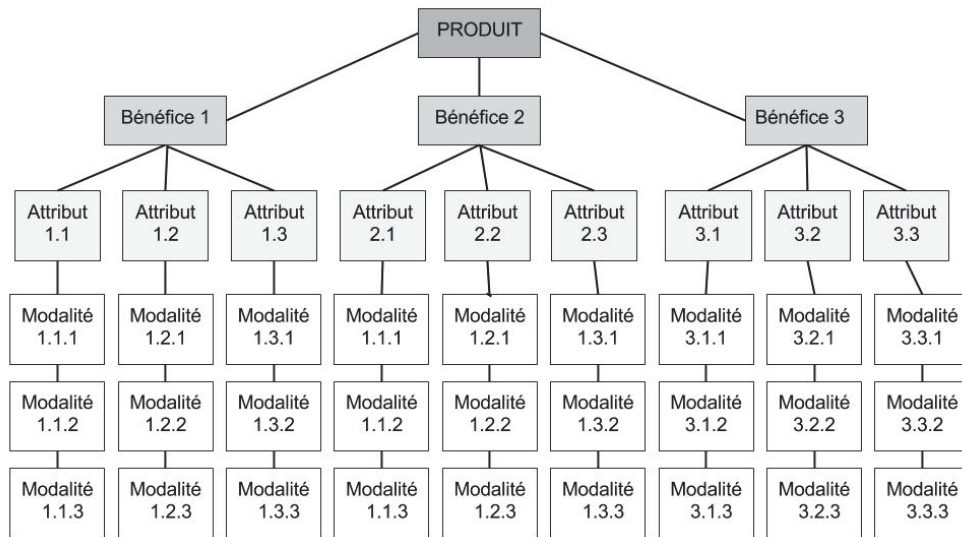


Figure 1. Structure hiérarchique

L'évaluation de cet attribut en fonction des bénéfices est subjective et peut différer. On peut trouver un jean très large confortable mais qui va renvoyer de nous une mauvaise image. Dans le modèle AHP, Saaty fait une évaluation à chaque nœud. Le nombre de nœuds peut être très important. Si il y a n critères il y a : $\frac{n(n-1)}{2}$, ou C_n^2 comparaisons. Ce qui limite dans la mise en œuvre le nombre de bénéfices évalués et le nombre d'attributs. D'un point de vue pratique, soit on limite le nombre de bénéfices, attributs, modalités à un ou deux items et l'on se permet de conserver un configurateur de recherche sur la page vue, soit on propose un configurateur de recherche plus sophistiqué dans un espace dédié à la manière des configurateurs automobiles. Les évaluations subjectives ou jugements peuvent se faire simplement par des questions de classement.

1.3. L'intégrale de choquet

L'intégrale de Choquet est, avec l'intégrale de Sugeno, un opérateur d'agrégations non additives très connu. Si sa création n'est pas récente [6], elle a fait l'objet d'une attention particulière depuis une quinzaine d'années en recherche opérationnelle notamment dans le cadre de la sélection d'alternatives à partir de critères multiples [3] [8] [15]. Les sciences de gestion commence à s'y intéresser pour étudier la formation des choix. Par exemple, [12] utilise l'intégrale de Choquet pour mieux comprendre la structure des préférences des voyageurs. On s'y intéresse également pour exploiter à des fins commerciales les importantes masses d'informations sur le comportement du consommateur générées par le commerce électronique. L'utilisation de ces données dans les systèmes de recommandation permet, en effet, de stimuler les chiffres d'affaires grâce à des ventes complémentaires ou additionnelles. Concrètement, l'intégrale de Choquet permet d'agréger les scores d'un item réunis dans une capacité selon la définition 1.1.

DEFINITION 1.1– Soit $N = e_1, \dots, e_n$ un ensemble d'éléments. Une capacité sur N est une fonction $\mu : 2^N \rightarrow [0, 1]$ vérifiant :

$\mu(\emptyset) = 0, \mu(N) = 1$ et $\mu(A) \leq \mu(B)$ si $A \subseteq B$ (respect de la monotonie)

où A et B sont des parties de N

Le score global de l'item à partir d'un ensemble de critères est établi de la manière suivante (Définition 1.2) :

DEFINITION 1.2– Soit μ une capacité sur N , et $f : N \rightarrow R$ une fonction représentant les scores d'un objet sur les n critères. L'intégrale de Choquet de f par rapport à μ (score global de l'objet) est donné par :

$$C_{\mu}(f) = \sum_{i=1}^n [f(\sigma(i)) - f(\sigma(i-1))] \mu(A_i)$$

avec $A_i = \{\sigma(i), \dots, \sigma(n)\}$ et $f(\sigma(0)) = 0$

où σ est une permutation sur N telle que $f(\sigma(1)) \leq f(\sigma(2)) \leq \dots \leq f(\sigma(n))$

2. Mise en œuvre de l'intégrale de Choquet dans un système de recommandation Objet

L'utilisation de l'intégrale de Choquet associée à une analyse hiérarchique, qui plus est, dans un système de recommandation est une suggestion récente [7] [1] [9] [16]. Mais elle n'a pas encore fait l'objet d'une mise en œuvre pratique dans un système de recommandation. Son application est en effet difficile car elle requiert une interaction relativement importante avec l'utilisateur actif. En tout cas plus importante que le simple tri en fonction d'un critère comme cela est souvent proposé. Toutefois la méthode n'a pas de problèmes de cold-start (démarrage à froid), ne nécessite pas une quantité de données très importantes pour être fiable, ni de l'utilisateur ni de voisins. La méthode nécessite quatre étapes principales (détaillées dans la figure 2) :

- un filtrage actif de l'utilisateur final permettant le calcul d'un modèle d'attitude grâce à des questions interactives simples présentées dans un "configurateur".
- une analyse des critères des objets pouvant être recommandés
- un calcul des scores globaux des objets établissant leur utilité pour la recommandation
- une recommandation sur seuil top-10 par exemple ou k-moyenne

2.1. Le configurateur de choix (ou de recommandations)

L'objectif du configurateur est d'obtenir les scores de la capacité, en d'autres termes, les coefficients des caractéristiques du produit et des combinaisons de caractéristiques au regard de l'intérêt (autrement dit "bénéfice") recherché par le consommateur. On suppose que les bénéfices recherchés pour l'achat d'une catégorie de produit, les attributs saillants, ainsi que les caractéristiques (ou modalités) ont été identifiés au préalable au travers de focus groupes, dans la littérature et/ou l'analyse de l'offre. En prenant l'exemple de l'achat de jeans, nous pouvons établir le tableau ci-dessous (Tab. 2.1., pour une référence à la littérature en ce qui concerne ce produit voir par exemple [17]) :

Dans notre tableau il y a quatre bénéfices principaux recherchés par les consommateurs, soit quatre capacités définissables. Ceux-ci peuvent être présentés dans une liste déroulante permettant la sélection des attributs susceptibles de concourir à la réalisation du bénéfice choisi. Si, pour un consommateur actif, le confort est la motivation principale dans le choix d'un jean (le principal bénéfice recherché), alors, il devra évaluer l'importance qu'il attache à la marque, la matière, la forme, et toute combinaison de ces attributs, pour parvenir à cet objectif (avoir un jean confortable). Avec trois attributs, excepté les cas

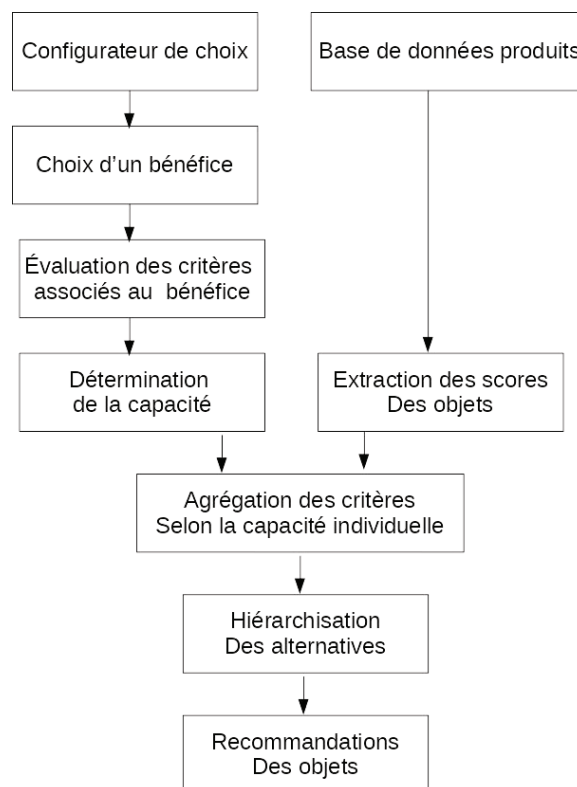


Figure 2. Schéma global

Bénéfices	Attributs	Modalités
Image renvoyée	Couleur	Bleu jean, Bleu électrique, Blanc, Chocolat, Gris, Noir, ...
	Forme	Bootcut, Taper, Skinny, Regular, Slim, Patte d'éph...
	Prix	10€, 20€, 30€, 40€, 50€, ..., 250€
	Matière	100% Coton, Coton + Élasthane, Polyester
	Marque	Jules, H&M, Devred, Diesel, Gstar, Boss, Aber, Kiabi,...
	Style	Patché, Classique, Délavé, Déchiré, Usé.
Qualité-prix	Prix	10€, 20€, 30€, 40€, 50€, ..., 250€
	Matière	100% Coton, Coton + Élasthane, Polyester
	Marque	Jules, H&M, Devred, Diesel, Gstar, Boss, Aber, Kiabi,...
Beauté	Couleur	Bleu jean, Bleu électrique, Blanc, Chocolat, Gris, Noir,...
	Forme	Bootcut, Taper, Skinny, Regular, Slim, Patte d'éph...
	Matière	100% Coton, Coton + Élasthane, Polyester
	Style	Patché, Classique, Délavé, Déchiré, Usé.
Confort	Marque	Jules, H&M, Devred, Diesel, Gstar, Boss, Aber, Kiabi,...
	Matière	100% Coton, Coton + Élasthane, Polyester
	Forme	Bootcut, Taper, Skinny, Regular, Slim, Patte d'éph...

Tableau 2.1. Tableau des bénéfices et de leurs attributs associés

extrêmes (aucun attribut présent et, tous les attributs sont présents), l'utilisateur devra évaluer six items. Soit on utilise une question de classement (Fig. 4) soit des curseurs ou des questions à échelles. L'étendue de l'échelle doit être suffisamment grande pour mettre en évidence les effets de sous ou sur-additivité. Pour trois attributs, une échelle de MACBETH [2] à neuf positions fonctionne parfaitement (Fig. 3).

Quel est votre principale motivation d'achat ?

📌 Veuillez sélectionner une réponse ci-dessous

Le confort

Evaluer les critères et combinaisons de critères nécessaires à vos choix

	1	2	3	4	5	6	7	8	9
La marque seule	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La matière seule	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La forme seule	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La marque et la matière	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La marque et la forme	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La matière et la forme	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 3. Exemple de question à échelle

Effectuez un double-clic ou glissez/déposez les éléments de la liste de gauche à la liste de droite. L'élément avec le rang le plus élevé est situé le plus haut jusqu'à celui du rang le moins élevé.

Vos choix	Votre classement
	Forme seule
	Marque seule
	Matière seule
	Marque et forme
	Forme et matière
	Marque et matière

Figure 4. Exemple de question à classement

Les réponses obtenues permettent d'établir les scores de la capacité μ_{CO} définie sur une partition de l'ensemble $\{Marque, Forme, Matière\}$ par :

$$\begin{aligned}
 &\mu_{CO}(\emptyset) \\
 &\mu_{CO}(\{Marque\}) = 0.3 \qquad \mu_{CO}(\{Forme\}) = 0.2 \qquad \mu_{CO}(\{Matière\}) = 0.4 \\
 &\mu_{CO}(\{Marque;Forme\}) = 0.5 \qquad \mu_{CO}(\{Marque;Matière\}) = 0.8 \qquad \mu_{CO}(\{Matière;Forme\}) = 0.5 \\
 &\mu_{CO}(\{Marque;Forme;Matière\}) = 1
 \end{aligned}$$

Les données proposées par l'utilisateur doivent rester cohérente avec les principes de normalité, de positivité et de monotonie. Il suffit d'ajouter quelques tests de validité simples aux éléments de réponse. Il convient donc de vérifier que les scores donnés aux singletons de critères soient inférieurs aux scores donnés à leur combinaison. On observe ici la mise en évidence d'une sous-additivité entre la matière et la forme, c'est à dire que l'utilisateur actif trouve que certaines formes sont incompatibles avec la matière. Cet effet peut être mesuré (Définition 2.1) par l'indice d'interaction [14]

DEFINITION 2.1– Soit $N = e_1, \dots, e_n$ un ensemble de critères et μ une capacité sur N

$$I^\mu(A) = \sum_{K \subset N \setminus i} \frac{(n-k-|A|)!k!}{n-|A|+1!} \sum (-1)^{|A|-|L|} \mu(K \cup L)$$

où A est une partie non vide de N et avec $k = |K|$

Pour deux critères l'indice d'interaction est :

$$I_{i,j} = \sum_{K \subset N \setminus \{i,j\}} \frac{(n-|K|-2)!|K|!}{(n-1)!} [\mu(K \cup \{i,j\}) - \mu(K \cup \{i\}) - \mu(K \cup \{j\}) + \mu(K)]$$

Dans notre exemple on obtient :

$$I(\{\text{marque, forme}\}) = 1/20 = 0.05$$

$$I(\{\text{marque, matière}\}) = 3/20 = 0.15$$

$$I(\{\text{forme, matière}\}) = -1/20 = -0.05$$

La sur-additivité est mesurée par un chiffre positif tandis que la sous-additivité (redondance) est négative (couleur et matière). L'importance des critères peut être mesuré par l'indice de Shapley défini ci-dessous [14].

DEFINITION 2.2– Soit $N = e_1, \dots, e_n$ un ensemble de critères et μ une capacité sur N

$$\phi^\mu(e_i) = \sum_{K \subset N \setminus i} \frac{(n-k-1)!k!}{n!} [\mu(K \cup \{e_i\}) - \mu(K)]$$

avec $k = |K|$

Sur notre exemple on obtient :

$$\phi_{CO}(\text{marque}) = 23/60 = 0.383$$

$$\phi_{CO}(\text{forme}) = 11/60 = 0.183$$

$$\phi_{CO}(\text{matiere}) = 26/60 = 0.433$$

C'est donc la matière qui est prépondérante pour remplir un objectif de confort. On peut noter que ces indicateurs peuvent aussi être habilement utilisés pour sélectionner les futurs articles proposés à la vente en ligne.

2.2. Analyse des critères des objets

On demande à l'utilisateur actif de mesurer les modalités pour un attribut donné à l'aide d'échelles de notation ou de classements. Les modalités ont été au préalable obtenues par focus groupe ou par analyse des historiques de choix (filtrage passif). La matière présente ici trois modalités évaluée dans l'optique du confort : 100% coton ; coton+élasthane ; polyester (voir Tab. 2.2). On peut construire pour un tableau de scores, pour les modalités de chaque attribut (Tab. 2.3), vu sous l'angle du confort, qui servira de support à l'analyse des objets. Il s'agit d'une simple fonction d'identification par mot-clé et d'association avec des valeurs.

100% coton	coton+élasthane	polyester
0,6	0,8	0,2

Tableau 2.2. Scores des marques

Diesel	Jules	Kiabi	Devred
0,6	0,6	0,4	0,5

Tableau 2.3. Scores des formes

Les tableaux de scores sont propres à chaque utilisateur, réutilisables, car enregistrables dans le compte client, et modifiables à volonté. Il peut être intéressant d'étendre les évaluations des attributs à un filtrage collaboratif. Les valeurs des attributs utilisées pour effectuer les recommandations résultent alors de moyennes des notations des pairs.

2.3. Calcul des scores globaux

Soit un jean défini par un score de 0.2 pour sa marque, de 0.6 pour sa forme et 0.9 pour sa matière (voir ci-dessus Fig. 2). L'évaluation globale du produit à l'aide du jeu d'attributs et de ses coefficients dans l'optique du confort donne :

$$C_{\mu_{CO}}(0.2, 0.6, 0.9) = (0.2-0) * \mu_{CO}(\{C, F, M\}) + (0.6-0.2) * \mu_{CO}(\{F, M\}) + (0.9-0.6) * \mu_{CO}(\{M\})$$

$$C_{\mu_{CO}}(0.2, 0.6, 0.9) = 0.2 * 1 + 0.4 * 0.5 + 0.3 * 0.4 = 0.52$$

Les logiciels intégrant le calcul de l'intégrale de Choquet sont encore très rares. Le plus connu est le paquet "kappalap" du logiciel libre R (Fig. 5) dont nous proposons un exemple de script ci-dessous [14] [11].

C'est le jean 2 qu'il faut recommander puis le jean 1. De manière pratique, puisque les recommandations sont hiérarchisées, il est possible de paramétrer le nombre de recommandations et de s'en tenir, par exemple au top-10 ou aux recommandations dépassant un certain score global (méthode avec seuil). L'étude des indices d'importances peut permettre d'identifier les segments d'utilisateurs privilégiant les mêmes objectifs et/ou les mêmes critères, et, par conséquent, d'associer la procédure à une forme de recommandation collaborative. Pour bien comprendre l'intérêt de l'intégrale de Choquet dans les systèmes de recommandation objet, nous nous sommes limités, dans la présentation méthodologique, à un seul bénéfice recherché (ou objectif). Cependant, les futurs consommateurs font souvent leur choix en ayant plusieurs objectifs à l'esprit. Pour un jean, ils peuvent rechercher à la fois le confort et l'esthétisme, c'est à dire avoir plusieurs objectifs. L'intégrale de Choquet permet de prendre en compte ces situations et d'établir un modèle de préférence multi-objectif [4]. Pour cela il suffit d'ajouter un niveau hiérarchique en recherchant les scores au niveau des bénéfices. Toutefois cela implique une interactivité qui risque de

Regular	Slim	Taper	Bootcut
0,4	0,7	0,4	0,3

Tableau 2.4. Scores des modalités pour le critère "matière"

Article	Matière	Marque	Forme
Jean 1	coton (0,6)	Jules (0,6)	Regular (0,4)
Jean 2	coton+elasthane (0,8)	Diesel (0,6)	Slim (0,7)
Jean 3	coton (0,6)	Kiabi (0,4)	Regular (0,4)
Jean 4	Polyester (0,2)	Kiabi (0,4)	Regular (0,4)

Tableau 2.5. Tableau de scores par article

Article	Score global
1	0,5
2	0,71
3	0,46
4	0,3

Tableau 2.6. Scores globaux par article

```

— script simple —
library(kappalab)
mu<-capacity(c(0,0.3,0.2,0.4,0.5,0.8,0.5,1))
Shapley.value(mu)
x<-c(0.2,0.6,0.9)
Choquet.integral(mu,x)
+ Shapley.value(mu)
  1          2          3
  0.3833333  0.1833333  0.4333333
+ Choquet.integral(mu,x)
[1]0.52
——— Script de calcul pour les 4 jeans de la gamme avec R-cran———
mu<-capacity(c(0,0.3,0.2,0.4,0.5,0.8,0.5,1))
jeans<-matrix(c(0.6,0.8,0.6,0.2,0.6,0.6,0.4,0.4,0.4,0.7,0.4,0.4),ncol=3)
for (i in seq(1,4))print(Choquet.integral(mu,jeans[i,]))

```

Figure 5. Exemple de script sous R avec le package Kappalab

ne plus être soutenable dans un système de recommandation tant l’effort à fournir par l’utilisateur actif sera important.

3. Validation expérimentale

L’idée étant séduisante, nous avons testé la méthode sur 94 étudiants (49 garçons et 45 filles). Les objectifs étaient les suivants :

- Valider le système de recueil de l’information
- Valider la faisabilité de traitement : mise en évidence des capacités, des indices d’interaction et d’importance, calcul des scores agrégés.

- Valider la cohérence des recommandations avec les motivations de l'utilisateur.
- Mesurer le nombre d'items de premier niveau (bénéfices recherchés) soutenables compte tenu de l'élévation du niveau d'interactivité.
- Explorer la faisabilité d'un filtrage collaboratif pour parfaire les recommandations.

Nous avons utilisé des questions à échelles à 9 niveaux (échelle MacBeth). Les tests de validité ont permis de garantir la monotonie qui semblait évidente à la plupart des interviewés. Les bénéfices et les attributs ont été sélectionnés par un focus-groupe. Les modalités des attributs sont tirés des caractéristiques des produits. Concernant l'évaluation du nombre de bénéfices au départ, les interviewés pouvaient librement configurer leur choix sur la base de 1 à 4 objectifs (ou bénéfices). Une large majorité s'est appuyée sur deux bénéfices (62/94 voir Tab. 3.7). Il s'est avéré que les interviewés trouvaient les questions fastidieuses dans une configuration à trois bénéfices. Le nombre et le choix d'attributs associés à chaque

Nb de Bénéfices	1	2	3
Nb de Personnes	9	62	23

Tableau 3.7. Nombre de bénéfices par individu

bénéfice était librement déterminé par les interviewés Nous avons remarqué que la « Beauté », représentant l'aspect esthétique, apparaît être le bénéfice le plus recherché (73,40% des individus). A l'opposé, le bénéfice « Image recherchée » est peu sollicité (25,53% des individus). Les répondants se sont basés en moyenne sur 2,15 bénéfices pour fonder leur choix. Chaque répondant a réalisé 4,48 évaluations d'attributs. Chaque bénéfice a donc été évalué en moyenne par 2,08 attributs. Le bénéfice « image recherchée » est le plus souvent associé à l'attribut style (75,00% des cas). La « qualité-prix » est, bien sûr, souvent, associée au prix (84,48% des cas). Les deux attributs principaux de la « beauté » sont la couleur et la forme (respectivement 76,81% et 78,26% des cas). L'attribut principal du « confort » est la forme (78,43% des cas). Les sondés devaient faire un choix parmi dix articles présentés. Le but étant de vérifier la cohérence entre les préférences exprimées et modélisées dans la capacité et les choix. Le résultat n'est pas probant au niveau individuel, car les objets présentés aux choix ne comportaient pas toujours les caractéristiques déclarées comme importante par tel ou tel sondé. Certaines marques n'étaient par exemple pas présentes. Ce calcul d'importance moyenne montre à nouveau que la « beauté » est le bénéfice le

Bénéfices	Image renvoyée	Qualité-Prix	Beauté	Confort
Imp. moyenne	0,1	0,26	0,39	0,25

Tableau 3.8. Importance moyenne des bénéfices

plus utilisé (Tab. 3.8) et que « l'image recherchée » n'est pas un bénéfice prépondérant. Il est possible d'opérer une clusterisation des sondés dans l'optique d'un filtrage collaboratif. Nous avons pu identifier, dans notre échantillon, quatre catégories de sondés (Tab. 3.9). La catégorie A rassemble les « esthètes » réceptifs aux aspects esthétiques dans les dimensions à la fois personnelles et sociales. La catégorie B rassemble les « pragmatiques » sensibles à la qualité-prix et au confort. La catégorie C est représentée par les « sensuels » qui privilégient la beauté dans sa dimension esthétique personnelle et le confort qui renvoie au plaisir du toucher. La catégorie D est constituée d'individus « prudents » sensibles à la beauté mais pas à n'importe quel prix. On peut également étudier les interactions entre les différents bénéfices ou attributs pour chaque individu. Pour cela on réalise un tableau résumant les effectifs des interactions négatives et positives entre bénéfices, calculées au niveau individuel (Tab. 3.10). Lorsqu'il y a autant de

	Importance moyenne des critères et effectifs par clusters (A, B, C, D)								
Bénéfices	ϕ_{Amoy}	N_A	ϕ_{Bmoy}	N_B	ϕ_{Cmoy}	N_C	ϕ_{Dmoy}	N_D	Total
Image	0,418	20	0,032	4	0	0	0	0	24
Qualité-Prix	0,188	10	0,365	26	0	0	0,439	22	58
Beauté	0,335	15	0,095	10	0,686	22	0,561	22	69
Confort	0,059	5	0,508	30	0,314	16	0	0	51
Taille		20		30		22		22	

Tableau 3.9. Clustering des critères

	I({IR,QP})	I({IR,BE})	I({IR,CO})	I({QP,BE})	I({QP,CO})	I({BE,CO})
Nb. +	6	2	2	6	5	6
Nb. -	7	12	6	26	21	20

Tableau 3.10. Interactions entre critères

valeurs positives que négatives, on peut supposer qu'il n'y a pas de consensus au niveau des synergies entre les critères. A l'opposé, lorsque les valeurs négatives dominent largement, on peut supposer qu'il existe a priori de la synergie négative entre les critères. Les objets peuvent également être segmentés en fonction des scores selon la méthode des k moyennes². La classe A contient les valeurs les plus élevées d'utilité. On a cherché à vérifier que les jeans réellement choisis étaient associés à des niveaux d'utilités théoriques élevés (Classe A, éventuellement B). Le tableau ci-dessous donne le nombre de jeans choisis dans chaque classe. Lorsque les utilités calculées se répartissent en deux classes, sur 63 jeans choisis, 42 appartenaient à la classe d'utilité la plus élevée. La table 3.12 donne la répartition des effectifs selon trois classes d'utilités théoriques. On remarque (Tab 3.11 et 3.12), que les classes associées aux scores globaux les plus élevés ont des effectifs plus élevés que les classes associées aux scores plus faibles.

Niveaux	A	B
Effectifs	42	21

Tableau 3.11. Nombre de jeans choisis par classe (2 classes)

Niveaux	A	B	C
Effectifs	28	24	11

Tableau 3.12. Nombre de jeans choisis par classe (3 classes)

4. Conclusion

Récemment suggérée comme algorithme de recommandation, nous proposons ici une application pratique de l'intégrale de Choquet. Ses capacités de prédiction (de recommandation) semblent correctes puisque les produits choisis correspondent à des scores globaux élevés. Nous montrons les limites et les avantages de son utilisation dans le cadre d'un configurateur de recommandation. La possibilité de prendre en compte simplement la multiplicité des objectifs poursuivies pour un achat et les interactions

2. Méthode barycentrique

entre critères est intéressante. Le principe d'une capacité, les indices d'interactions et d'importances permettent de comprendre comment les critères s'articulent pour aboutir à un choix. Ces éléments peuvent servir également de support pour la sélection d'une offre globale répondant aux critères primordiaux observés. En d'autres termes ils permettent d'orienter la gamme. La clusterisation des modèles individuels permet d'étendre l'approche à un filtrage collaboratif. Une approche hybride est donc possible. Les principales limites de la méthode résident dans l'intensité de l'interactivité avec l'utilisateur. Il apparaît difficile de dépasser deux objectifs et trois critères par objectif. Le nombre de comparaisons augmente en effet trop fortement.

Bibliographie

- Francesca Abastante, Salvatore Corrente, Salvatore Greco, Alessio Ishizaka, and Isabella Lami. Using a new parsimonious AHP methodology combined with the Choquet integral : An application for evaluating social housing initiatives. *arXiv :1704.08119 [cs, math]*, April 2017. arXiv : 1704.08119.
- Carlos A. Bana E Costa, Jean-Marie De Corte, and Jean-Claude Vansnick. Macbeth. *International Journal of Information Technology & Decision Making*, 11(2) :359–387, March 2012.
- Denis Bouyssou. *Concepts et méthodes pour l'aide à la décision. : Vol 3, analyse multicritère*. Hermes Science Publications, May 2006.
- Juergen Branke, Salvatore Corrente, Salvatore Greco, Roman Słowiński, and Piotr Zielniewicz. Using Choquet integral as preference model in interactive evolutionary multiobjective optimization. *European Journal of Operational Research*, 250(3) :884–901, May 2016.
- Damien Chaney. L'apport des cartes cognitives à l'analyse des représentations mentales. (French). *Recherche et Applications en Marketing (AFM c/o ESCP-EAP)*, 25(2) :93–115, June 2010.
- Gustave Choquet. Theory of capacities. *Annales de l'institut Fourier*, 5 :131–295, 1954.
- Salvatore Corrente, Salvatore Greco, and Alessio Ishizaka. Combining analytical hierarchy process and Choquet integral within non-additive robust ordinal regression. *Omega*, 61 :2–18, June 2016.
- D. Denneberg. *Non-Additive Measure and Integral*. Springer Science & Business Media, March 2013.
- Soumana Fomba, Pascale ZARATÉ, and Marc Kilgour. A Recommender system based on MultiCriteria Aggregation. In *2nd International Conference on Decision Support Systems Technology (ICDSST 2016) EWG-DSS - EURO Working Group on Decision Support Systems*, pages pp. 1–7, Plymouth, United Kingdom, May 2016.
- Songjie Gong. A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering. *Journal of Software*, 5(7), July 2010.
- M. Grabisch. L'utilisation de l'intégrale de Choquet en aide multicritère à la décision. *Newsletter of the European Working Group " Multicriteria Aid for Decisions*, 3(14) :5–10, 2006.
- Gang Li, Rob Law, Huy Quan Vu, and Jia Rong. Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet Integral. *Tourism Management*, 36 :321–330, June 2013.
- Nira Liberman, Yaacov Trope, and Elena Stephan. Psychological distance. *Social psychology : Handbook of basic principles*, 2 :353–383, 2007.
- Michel Grabisch, Ivan Kojadinovic, and Patrick Meyer. Using the Kappalab R package for capacity identification in Choquet integral based MAUT, 2006.
- Bilel Moulahi, Lynda Tamine, and Sadok Ben Yahia. L'intégrale de Choquet discrète pour l'agrégation de pertinence multidimensionnelle. In *Conférence francophone en Recherche d'Information et Applications - CORIA 2013*, pages pp. 1–16, Neuchâtel, Switzerland, April 2013.
- S. Nadi, M. Samiei, H. R. Salari, and N. Karami. A NEW MULTI-CRITERIA EVALUATION MODEL BASED ON THE COMBINATION OF NON-ADDITIVE FUZZY AHP, CHOQUET INTEGRAL AND SUGENO λ -MEASURE. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W4 :423–428, September 2017.
- Osmud Rahman. Understanding Consumers' Perceptions and Buying Behaviours : Implications for Denim Jeans Design. *Journal of Textile and Apparel, Technology and Management*, 7(1), April 2011. Cited by 0002.