



HAL
open science

Extraction automatique d'évaluations multicritère pour l'identification d'objectifs prioritaires et réalistes dans le domaine du tourisme.

Abdelhak Imoussaten, Benjamin Duthil, François Troussel, Jacky Montmain

► To cite this version:

Abdelhak Imoussaten, Benjamin Duthil, François Troussel, Jacky Montmain. Extraction automatique d'évaluations multicritère pour l'identification d'objectifs prioritaires et réalistes dans le domaine du tourisme.. 25èmes Rencontre francophone sur la Logique Floue et ses applications (LFA) 2016, Nov 2016, La Rochelle, France. hal-01933811

HAL Id: hal-01933811

<https://hal.science/hal-01933811v1>

Submitted on 29 Jul 2022

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.

Extraction automatique d'évaluations multicritère pour l'identification d'objectifs prioritaires et réalistes dans le domaine du tourisme

Automated extraction of multicriteria assessments for allocating realistic and prioritized targets in tourism applications

A. Imoussaten¹ B. Duthil² F. Troussel¹ J. Montmain¹

¹ Centre de Recherche LGI2P, École des mines d'Alès

² Laboratoire Informatique, Image et Interaction L3I

¹ Site EERIE, Parc scientifique G. Besse, 30035 Nîmes Cedex 1, France

² Pôle Sciences & Technologie, Avenue Michel Crêpeau, 17042 La Rochelle Cedex 1, France

Résumé :

Cet article s'intéresse à la détermination des performances qu'il convient d'améliorer en priorité dans une évaluation multicritère. Il s'agit d'identifier les dimensions selon lesquelles l'amélioration serait la plus profitable tout en restant réalisable. Un indice de plus-value estime la pertinence d'une amélioration au regard de ces deux aspects. Cette proposition est intégrée dans un système de recommandation qui exploite automatiquement des bases de critiques en langage naturel recueillies sur le web pour estimer les préférences des utilisateurs et les chances de succès des améliorations, puis identifier les paramètres de l'indice de plus-value. Cela permet de faciliter les phases d'évaluation par critère et d'identification du modèle de préférence des utilisateurs, généralement complexes et fastidieuses, plus particulièrement quand les interactions entre les critères sont prises en compte. Cette proposition est appliquée à l'amélioration de services hôteliers dans la région Nouvelle Aquitaine qui souhaite faire un diagnostic et améliorer l'attractivité de ses hôtels.

Mots-clés :

Analyse multicritère, transformé de Möbius, Capacité, Intégrale de Choquet, extraction d'opinions.

Abstract:

This paper proposes a method to diagnose the performances that should be improved first in a multidimensional assessment. This method computes the subset of criteria that maximizes the compromise between challenging and achievable goals. A worth-index is introduced to estimate the relevancy of potential improvements w.r.t these two perspectives. This proposal is integrated into a recommender system that automatically processes a corpus of critics in natural language collected on the web to capture the users' preferences and the chances of success of potential improvements, and then identify the required parameters of the worth-index. This method optimizes the processing of the available amount of data on websites to facilitate assessment of criteria and establish the customers' preference model. These two steps are generally tedious and tricky especially when criteria may interact. An application case is proposed. It concerns the Nouvelle Aquitaine region that aims improving the attractiveness of its tourist hotels.

Keywords:

MCDA, Möbius representation, Capacity, Choquet Integral, Opinion-mining.

1 Introduction

Ces dernières années, de nombreuses entreprises et sites Web ont mis en place des systèmes permettant l'analyse des préférences de leurs utilisateurs afin de mieux répondre aux attentes de ces derniers. À ce jour, les systèmes de recommandation sont présents dans de multiples domaines comme le tourisme/les loisirs, la publicité, le commerce électronique, les films, etc. En raison de l'explosion de la quantité de données diffusées sur l'Internet ces dernières années, rechercher et analyser ces contenus d'opinions pertinents est devenu une tâche difficile pour les utilisateurs qui sont souvent noyés par la masse d'informations. La capacité à synthétiser et à analyser ces contenus est devenue une nécessité pour les décideurs, qu'il s'agisse d'un dirigeant d'hôtel ou bien d'un décideur politique en charge d'un territoire, afin qu'ils soient en mesure de prendre les bonnes décisions au bon moment, et surtout de mesurer l'impact de leurs actions. Ceci explique l'intérêt croissant porté aux systèmes de recommandation tant par les utilisateurs que par les décideurs. La problématique de recommandation a été identifiée comme étant le moyen d'aider les individus à trouver des informa-

tions, ou des éléments, qui sont susceptibles de les intéresser. Généralement, on considère un ensemble d'utilisateurs *Users* et un ensemble d'éléments *Items* (services ou produits) pouvant être recommandés à chacun des utilisateurs. Par ailleurs, une recommandation multicritère contribue à améliorer la qualité d'un SR (Système de Recommandation) car elle permet d'explicitier les caractéristiques pour lesquelles un item a été proposé à l'utilisateur [1] [3] [9] [10].

Dans [3] et [4], une méthode non-supervisée d'extraction d'opinions multicritère est proposée. Elle permet aux utilisateurs de s'affranchir de la contraignante évaluation partielle sur chaque critère des items à recommander : les utilisateurs déposent simplement leurs critiques comme à l'accoutumée, sous la forme de textes en langage naturel exprimant leurs opinions, et le système se charge de les analyser automatiquement. Il les décompose d'abord selon les critères d'évaluation (segmentation thématique) avant de calculer les polarités ou opinions pour chacun des extraits résultant de la segmentation (opinion-mining/sentiment analysis).

L'approche proposée dans cet article consiste à exploiter l'extraction automatique d'opinions pour faciliter l'acquisition d'informations nécessaires à l'application d'outils issus de la décision multicritère. Elle permet l'identification du modèle de préférence des internautes prenant en compte les interactions entre critères exprimées à travers les critiques recueillies. Le but de l'approche est de permettre au décideur d'identifier les axes d'amélioration prioritaires, qui vont lui fournir un gain maximal. L'indice ainsi créé fournit une indication sur la capacité à réaliser cette amélioration. Pour exposer notre approche, nous nous focalisons sur l'analyse de données touristiques issues de TripAdvisor et nous nous limiterons aux critiques déposées sur les hôtels de la région Nouvelle Aquitaine. De plus, l'évaluation des hôtels sera limitée à six critères d'intérêts.

La section 2 résume l'approche *Synopsis*

de détection d'opinions que nous avons développée dans [3]. Des critiques d'internautes sont donc extraites et les scores d'opinions sont calculés relativement au groupe de critères sélectionné. La section 3 présente les éléments mathématiques issus de l'analyse multicritère utiles à l'approche proposée. La section 4 détaille comment déterminer le sous-ensemble de critères sur lesquels il est le plus avantageux de s'améliorer. La section 5 présente un cas d'application sur les hôtels de la région Nouvelle Aquitaine qui sont référencés sur *TripAdvisor*.

2 Extraction d'opinions

Sur la base de méthodes statistiques, notre approche construit un dictionnaire de descripteurs d'opinion pour une thématique donnée. Ce dictionnaire est ensuite utilisé pour calculer automatiquement la polarité des extraits de texte relatifs aux critères. Deux étapes sont distinguées dans cette évaluation multicritère :

- dans un premier temps, les portions de texte se rapportant à l'un des critères d'évaluation sont extraites avec notre approche *Synopsis* décrite dans [5]. Il s'agit donc de segmenter le texte par thème (ici, les critères) ;
- ensuite, pour chaque critère, la polarité des segments qui ont été extraits par *Synopsis* est calculée. Il s'agit là de l'étape d'opinion-mining à proprement parler [4].

2.1 Segmentation thématique par critère

L'approche *Synopsis* est utilisée pour identifier les extraits de texte qui traitent des critères considérés et se compose de 3 étapes : Construction automatique du corpus d'apprentissage utilisé pour apprendre les mots caractéristiques de chaque critère ; Apprentissage automatique de ces mots caractéristiques, appelés descripteurs, et construction d'un lexique associé au critère ; Segmentation du texte en utilisant le lexique associé au critère.

Pour construire le lexique, l'approche utilise un ensemble de mots appelés "mots germes". Ces mots germes permettent d'une part de caractériser sémantiquement le critère considéré, et d'autre part d'amorcer l'apprentissage des descripteurs caractéristiques d'un critère [5]. La construction du corpus d'apprentissage est automatique. Très schématiquement, plus un mot est fréquemment repéré dans le voisinage d'un mot germe du critère (comptage sur fenêtre glissante), plus on considère comme "forte" l'appartenance (sémantique) de ce mot au champ lexical du critère.

Synopsis construit son corpus d'apprentissage à partir de documents du web. L'appartenance d'un descripteur au champ lexical du critère résulte du calcul d'une fonction de discrimination; un seuil sur cette fonction permet de définir l'étiquette "forte" sur l'appartenance au critère. Plus ce seuil est haut, plus la phase d'apprentissage des descripteurs du dictionnaire de *Synopsis* est sélective. Les seuils retenus sont calculés automatiquement à partir de l'analyse des paliers de la fonction d'appartenance au critère [3].

2.2 Analyse de sentiments

L'approche d'extraction d'opinions (ou d'analyse de sentiments) est une adaptation de l'approche *Synopsis* à l'extraction d'opinions pour les phases de construction automatique du corpus d'apprentissage et d'apprentissage des descripteurs. Notons que les descripteurs d'opinion appris sont généralement propres au domaine d'analyse (ici l'hôtellerie). Les mots germes deviennent des mots germes d'opinion [4].

Au final, nous calculons pour un texte de critique donné un score d'opinion pour un critère donné (pour plus détails voir [3] et [4]). Soit CR l'ensemble des critiques étudiées et $N = \{1, \dots, n\}$ l'ensemble d'indices des critères sur lesquels sera basée la recommandation. L'approche proposée permet d'extraire de chaque

critique $j \in CR$ un vecteur de scores $x^j = (x_1^j, \dots, x_n^j)$ ou $x_i^j \in [0, 1]$ est le score d'opinion extrait de la critique j pour le $i^{\text{ème}}$ critère.

3 Rappels sur la théorie multi-attribut et l'intégrale de Choquet

3.1 La théorie multi attribut (MAUT : Multi Attribute Utility Theory)

Soit $N = \{1, \dots, n\}$ un ensemble d'attributs où le $i^{\text{ème}}$ attribut prend ses valeurs dans l'ensemble X_i . Une alternative est représentée par un élément $x = (x_1, \dots, x_n)$ du produit cartésien $X = X_1 \times \dots \times X_n$. L'objectif de la MAUT est de construire une relation de préférence \succsim sur X pour exprimer les préférences du décideur. Pour cela, on considère que \succsim est représentable par une fonction d'utilité $U : X \rightarrow \mathbb{R}$ telle que :

$$x \succsim y \Leftrightarrow U(x) \geq U(y) \quad \forall x, y \in X \quad (1)$$

La fonction U de l'équation (1) peut prendre plusieurs formes, la plus utilisée est le modèle additif :

$$\forall x \in X, \quad U(x) = \sum_{i=1}^n \omega_i u_i(x_i) \quad (2)$$

avec $\sum_{i=1}^n \omega_i = 1$, $\omega_i \geq 0$ pour tout $i \in N$ et $u_i : X_i \rightarrow \mathbb{R}$ représente les préférences du décideur sur l'attribut i , pour tout $i \in N$.

La forme additive modélise une compensation entre les attributs et requiert l'indépendance des attributs [11]. Cette forme additive est la plus répandue grâce à sa simplicité et son interprétation intuitive. Il s'agit, en général, d'une hypothèse trop simplificatrice car en réalité les attributs interagissent. Pour résoudre ce problème, un modèle plus général pour U dans l'équation (1) a été proposé. Il s'agit du modèle décomposable de Krantz et al. [12] où U s'écrit de la façon suivante :

$$\forall x \in X, \quad U(x) = F(u_1(x_1), \dots, u_n(x_n)) \quad (3)$$

avec pour tout $i \in N$ $u_i : X_i \rightarrow \mathbb{R}$ représente les préférences du décideur sur l'attribut i et F :

$\mathbb{R}^n \rightarrow \mathbb{R}$ est un opérateur d'agrégation. Pour faciliter les notations, $(u_1(x_1), \dots, u_n(x_n))$ sera noté $u(x)$ dans la suite. Le modèle de l'équation (3) nécessite de résoudre deux problèmes : (1) exprimer les préférences du décideur sur chaque attribut ; et (2) agréger ses préférences. L'existence des utilités u_i (une relation de préférence sur l'attribut i) nécessite la faible séparabilité de la relation \succsim [12].

3.2 L'intégrale de Choquet

Concernant l'opérateur F , dans le cas quantitatif, l'intégrale de Choquet est un candidat intéressant [6]. En effet, c'est un opérateur d'agrégation qui généralise plusieurs opérateurs classiques : la somme pondérée du modèle additif, la somme pondérée ordonnée (OWA), le *min*, le *max*, le *min* et *max* pondérés, etc [8]. En plus, l'opérateur de l'intégrale de Choquet permet de modéliser les interactions entre les attributs à travers une capacité [14] [6].

Dans la suite, nous allons faire un bref rappel sur les notions de capacité et de l'intégrale de Choquet.

Une *capacité* μ sur N est une fonction d'ensemble de 2^N dans $[0, 1]$ vérifiant :

1. $\mu(\emptyset) = 0$,
2. $\mu(N) = 1$ (normalisée),
3. monotone : $\forall K, T \subseteq N$,
 $K \subseteq T \implies \mu(K) \leq \mu(T)$.

La quantité $\mu(I)$ représente l'importance de la combinaison de critères $I \subseteq N$.

Dans cette partie, nous considérons que les attributs sont évalués sur la même échelle $[0, 1]$, ainsi nous confondons $u_i(x_i)$ avec $x_i \in [0, 1]$ sans perte de généralité.

Dans l'aide à la décision multicritère, l'évaluation par l'*intégrale de Choquet* d'une alternative $x = (x_1, \dots, x_n) \in [0, 1]^n$ par rapport à la capacité μ , notée $C_\mu(x)$, est définie par :

$$C_\mu(x) = \sum_{i=1, n} (x_{\sigma(i)} - x_{\sigma(i-1)})\mu(A_\sigma(i)) \quad (4)$$

où $0 = x_{\sigma(0)} \leq x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)}$ (σ est une permutation sur N), $A_\sigma(i) = \{\sigma(i), \sigma(i+1), \dots, \sigma(n)\}$.

La transformée de Möbius est une notion fondamentale dans la théorie de la capacité [2]. Dans le cadre de cet article, elle permet de faire le lien entre les notions de capacité et de fonction de croyance.

La transformée de Möbius d'une capacité μ sur N est une fonction $m : 2^N \rightarrow [0, 1]$ définie par :

$$m(T) = \sum_{K \subseteq N: K \subseteq T} (-1)^{|T \setminus K|} \mu(K), \quad \forall T \subseteq N \quad (5)$$

Quand m est donnée, il est possible de retrouver la capacité μ à partir de la formule suivante :

$$\mu(T) = \sum_{K \subseteq N: K \subseteq T} m(K), \quad \forall T \subseteq N \quad (6)$$

Il existe un lien entre la notion de capacité et la théorie des fonctions de croyance de Dempster et Shafer. En effet, la capacité μ définie dans l'équation (6) est une fonction de croyance ssi sa transformée de Möbius m est une fonction de masse [16] :

- $m(\emptyset) = 0$;
- $\sum_{T \subseteq N} m(T) = 1$;
- $m(T) \geq 0, \forall T \subseteq N$.

L'*intégrale de Choquet* de $x = (x_1, \dots, x_n)$ par rapport à la capacité μ , peut être formulée en fonction des masses $m(I)$ en utilisant la transformée de Möbius associée à μ :

$$C_\mu(x) = \sum_{I \subseteq N} m(I) \min_{i \in I} x_i \quad (7)$$

3.3 L'indice de plus value

La MAUT permet de représenter les préférences du décideur sans se soucier des contraintes matérielles de réalisation des alternatives. En général, les préférences du décideur sont utilisées pour évaluer, classer, comparer des alternatives. Elles peuvent être aussi employées pour explorer les champs d'amélioration possibles à partir d'un état donné afin de guider

le décideur. Dans ce sens, des travaux récents [7] [13] [15] proposent un indice qui permet de déterminer l'ensemble des critères sur lesquels il est le plus avantageux d'améliorer la solution actuelle étant donné un opérateur d'agrégation (ici, l'intégrale de Choquet).

L'indice en question, appelé *indice de plus value*, sera noté $\omega_\mu(x^0, I)$ pour un opérateur d'agrégation $F = C_\mu$, un ensemble de critères $I \subseteq N$ et une solution de départ $x^0 \in [0, 1]^n$. Il est défini par :

$$\omega_\mu(x^0, I) = \int_0^1 [C_\mu(\varphi_{\tau, I}(x^0)) - C_\mu(x^0)] d\tau \quad (8)$$

où $\varphi_{\tau, I}(x^0) = z \in [0, 1]^n$ est le vecteur des améliorations sur I tels que $z_i = (1 - \tau)x_i^0 + \tau$, $\forall i \in I$ et $z_i = x_i^0$, $\forall i \in N \setminus I$. Le choix d'intégrer le long de l'hyperbissectrice permet de garantir une amélioration homogène des critères de I .

4 Reformulation de l'indice de plus value

4.1 L'indice de plus value pour des gains qui ne sont pas équiprobables

La formule (8) exprime la moyenne du gain de toutes les améliorations possibles sur les critères de I entre $\varphi_{0, I}(x^0) = x^0$ et $\varphi_{1, I}(x^0)$ (satisfaction parfaite sur I et même satisfaction que x^0 sur $N \setminus I$). Dans [7] [13] il est supposé que ces améliorations se produisent de façon équiprobable. Nous proposons ici une nouvelle version de l'indice de plus value où d'une part, une valeur de probabilité est associée à chaque amélioration x à partir de x^0 , et d'autre part les améliorations x ne sont pas nécessairement considérées homogènes (on a simplement $x_i \geq x_i^0$, $\forall i \in I$) :

$$\omega_\mu(x^0, I) = \int_{[x^{0\uparrow}]_I} p_I(x_I, x_{-I}^0) [C_\mu(x_I, x_{-I}^0) - C_\mu(x^0)] dx_I \quad (9)$$

où $[x^{0\uparrow}]_I$ désigne l'ensemble des solutions qui

améliorent x^0 sur I , $p_I(x)$ exprime la probabilité d'obtenir la solution x , $y = (x_I, x_{-I}^0)$ signifie que $y_i = x_i$ si $i \in I$ et $y_i = x_i^0$ sinon et $C_\mu(x)$, $x \in [0, 1]^n$ est définie par l'équation (7).

4.2 Construction de p_I et de μ pour l'analyse de sentiment

Les informations nécessaires pour appliquer la nouvelle formule de l'indice de plus value de l'équation (9) sont la capacité μ et les probabilités p_I .

L'approche proposée ici consiste d'une part, à construire μ en tant que fonction de croyance calculée à partir d'une fonction de masse m . La fonction de masse m est obtenue à partir de statistiques sur les critères utilisés par les internautes. En effet, le site internet utilise un ensemble N de critères pour évaluer les services ou produits, mais les internautes ne choisissent dans la pratique qu'un sous-ensemble de ceux-ci pour exprimer leur évaluation du site. Ainsi, l'importance d'un sous-ensemble de critères I , $\mu(I)$, est mesurée par le nombre de fois où il a été utilisé dans l'évaluation (en totalité ($m(I)$) ou en partie ($m(J)$, $J \subset I$)) : $\mu(I) = \sum_{J \subseteq I} m(J)$.

D'autre part, les valeurs de probabilité empirique, sont calculées à partir des scores extraits des critiques pour chaque critère. Soit CR_I l'ensemble des critiques fournissant des évaluations sur le sous-ensemble de critère I . On note x_I^j le vecteur $(x_i^j, i \in I)$ des scores obtenus pour chaque critère extrait de la critique $j \in CR_I$. Les valeurs de probabilités marginales $p_I(x)$ sont calculées comme la probabilité d'occurrence du score x chez les internautes ayant utilisé le sous-ensemble I de critères pour évaluer les services ou produits. En effet, nous considérons que les utilisateurs n'ayant évalués que I ne se préoccupent pas de ce qui se passe sur le reste. Ainsi, ils sont considérés satisfaits sur $N \setminus I$. Ainsi, $p_I(x_I, x_{N \setminus I}) = p_I(x_I)$ où :

$$p_I(x) = \frac{|\{j \in CR_I : x_I^j = x\}|}{|CR_I|} \quad (10)$$

Lorsque $p_I(x)$ est grand, cela signifie qu'un nombre significatif d'internautes a évalué le site à x . Autrement dit, plus $p_I(x)$ est grand, plus x semble facilement atteignable. Autrement dit, une amélioration sur I apparaît d'autant plus réalisable, que des évaluations satisfaisantes sur I ont déjà pu être observées.

Finalement, avec les équations (6) et (10), on peut calculer l'indice de plus value de chaque sous-ensemble de critères avec la formule de l'équation (9) établi sur l'ensemble $[x^{0\uparrow}]_I$ des solutions améliorant x^0 sur I .

5 Cas d'application

Ce cas d'application utilise des critiques sur des hôtels de la région Nouvelle Aquitaine qui sont issues du site spécialisé *TripAdvisor*. Afin d'améliorer l'attractivité des hôtels sur ce territoire, les décideurs politiques doivent savoir quels sont les critères qu'il faut considérer en priorité pour réaliser une démarche d'amélioration continue sur leur territoire et ainsi proposer éventuellement de nouvelles infrastructures afin d'améliorer l'opinion des gens. Plus formellement, il s'agit de déterminer le sous-ensemble de critères sur lesquels l'amélioration recherchée apparaît la plus avantageuse et accessible en maximisant l'indice de plus-value proposé dans l'équation (9).

Lors de cette étude **33162** critiques écrites en langage naturel ont été considérées et évaluées sur six critères : *Literie*, *Emplacement*, *Rapport qualité-prix*, *Propreté*, *Service* et *Chambre*. Le tableau 1 montre la fréquence d'apparition de chaque critère dans les évaluations, nous pouvons remarquer que le critère *Service* est le plus souvent retenu par les internautes.

Le tableau 2 présente la fréquence des sous-ensembles de critères considérés par les internautes, autrement dit chaque groupe de personnes considère que le sous-ensemble qu'il évalue est le plus important pour sa satisfaction. Le sous-ensemble $\{1, 2, 3, 5, 6\}$ a été le plus évalué par les internautes avec

Tableau 1 – Indices des critères et leurs fréquences d'apparition dans les évaluations

critère	indice
Literie	1 :(20074)
Emplacement	2 :(20837)
Rapport qualité prix	3 :(22597)
Propreté	4 :(1714)
Service	5 :(27282)
Chambres	6 :(25979)

une fréquence de 13915. Cinq autres sous-ensembles dépassent un millier d'évaluations : $\{6\}$, $\{2, 5, 6\}$, $\{3, 5, 6\}$, $\{2, 3, 5, 6\}$ et $\{1, 5, 6\}$. Ainsi, on peut constater que *Chambres* seul est un critère très important et le critère *Propreté* n'apparaît dans aucun des sous-ensembles ayant été utilisé plus de 1000 fois.

Décider de s'améliorer sur un sous-ensemble de critères nécessite de faire un compromis entre le nombre d'internautes considérant ce sous-ensemble comme important (nombre d'internautes à satisfaire), d'une part, et le gain que peut procurer ce sous-ensemble d'autre part.

On considère ici que la solution de départ est le vecteur x^0 où chaque x_i^0 est la moyenne arithmétique des évaluations obtenues sur le $i^{\text{ème}}$ critère pour tous les internautes. Le tableau 3 résume la plus value apportée par chaque sous-ensemble de critères. Nous pouvons remarquer que sur le corpus considéré, la plus-value la plus importante est atteinte lorsqu'on améliore les critères 1, 2, 3, 5 et 6 simultanément (0.129728). Cependant, ce sous-ensemble de critères est très grand, il est généralement impossible de mettre en place une démarche d'amélioration sur un ensemble de critères aussi large car le décideur est tributaire des contraintes liées à son établissement (ex : la taille des chambres, sa situation géographique), il cherche donc à trouver une solution adaptée aux différentes marges de manœuvre dont il dispose et respectant les préférences des internautes. Par exemple, si le budget du décideur ne

Tableau 2 – Fréquences des sous-ensembles de critères (éléments focaux)

I	$freq(I)$	I	$freq(I)$
{6}	1809	{5}	189
{4}	112	{4,6}	58
{4,5,6}	29	{3}	210
{3,5}	606	{3,5,6}	1208
{3,4,6}	7	{3,4,5}	37
{2}	37	{2,6}	17
{2,5,6}	1045	{2,4}	4
{2,4,5,6}	49	{2,3}	9
{2,3,5}	556	{2,3,5,6}	2645
{2,3,4,5}	41	{2,3,4,5,6}	118
{1,6}	191	{1,5}	420
{1,4}	31	{1,4,6}	23
{1,4,5,6}	53	{1,3}	39
{1,3,5}	790	{1,3,5,6}	982
{1,3,4,6}	10	{1,3,4,5}	68
{1,2}	4	{1,2,6}	10
{1,2,5,6}	599	{1,2,4,5}	19
{1,2,3,5}	143	{1,2,3,5,6}	13915
{1,2,3,4,5,6}	800	{5,6}	801
{4,5}	6	{3,6}	194
{3,4}	14	{3,4,5,6}	52
{2,5}	370	{2,4,5}	22
{2,3,6}	5	{2,3,4,6}	3
{1}	199	{1,5,6}	1196
{1,4,5}	22	{1,3,6}	47
{1,3,4}	12	{1,3,4}	12
{1,3,4,5,6}	75	{1,2,5}	377
{1,2,4,5,6}	38	{1,2,3,4,5}	11

lui permet d'améliorer que k critères ($k \in N$), il est alors plus intéressant pour lui de choisir le sous-ensemble d'au plus k critères maximisant son gain. Pour $k = 3$, le tableau 3 montre que le sous-ensemble {3, 5, 6} avec un gain de 0.079161 semble être le plus profitable.

De façon générale, le décideur peut déterminer les sous-ensembles de critères qui sont les moins contraignants pour lui et de choisir parmi eux celui apportant un gain maximal.

6 Conclusion

Cet article propose une méthode pour déterminer les performances qu'il convient d'améliorer en priorité dans une évaluation multicritère. Il s'agit d'identifier les dimensions selon lesquelles l'amélioration serait la plus

Tableau 3 – Plus-value de chaque sous-ensemble de critères

I	$\omega_\mu(x^0, I)$	I	$\omega_\mu(x^0, I)$
{2}	0.000007	{1,2}	0.000061
{5}	0.000119	{1,6}	0.000412
{3}	0.001135	{2,5}	0.001354
{6}	0.001672	{1,5}	0.001919
{3,6}	0.002245	{3,4}	0.002422
{1,2,5}	0.002571	{1,2,4,5}	0.003270
{4,6}	0.003763	{1,2,3,4,5}	0.003778
{2,4,5}	0.003788	{4}	0.003835
{1,3,6}	0.003848	{1,4,5}	0.004683
{1,2,3,5}	0.005158	{1,2,5,6}	0.006147
{2,3}	0.007005	{1,2,4,5,6}	0.007496
{2,3,4,6}	0.009022	{2,6}	0.010921
{3,5}	0.011081	{1,3,4,5,6}	0.011090
{1,2,6}	0.011214	{2,3,5}	0.011792
{1,3,5}	0.012578	{2,3,4,5}	0.013024
{3,4,5}	0.013441	{1,3,4,5}	0.014624
{1,3,5,6}	0.016795	{1,4,5,6}	0.017351
{1,5,6}	0.020818	{2,5,6}	0.021741
{2,4,5,6}	0.022331	{4,5,6}	0.026349
{5,6}	0.029423	{3,4,5,6}	0.074522
{3,5,6}	0.079161	{2,3,4,5,6}	0.087297
{2,3,5,6}	0.087308	{1,2,3,4,5,6}	0.120251
{1,2,3,5,6}	0.129728		

profitable tout en restant réalisable. Elle étend la notion d'indice de plus-value lorsque les réalisations ne sont pas équiprobables.

Elle est basée sur un modèle multicritère. Le couplage d'une approche multicritère avec une approche non-supervisée d'analyse d'opinion permet de s'affranchir de la plupart des coûts relatifs à l'acquisition des données qui, en utilisant la méthode *Synopsis*, peuvent être extraits (sans intervention humaine), de textes en langage naturel. Un deuxième avantage de cette approche apparaît après la mise en œuvre des améliorations, car là encore, l'approche couplée à *Synopsis* permet un suivi automatique de l'évolution des critiques afin de vérifier l'impact de ces améliorations sur le public concerné, ce qui permet éventuellement de corriger dynamiquement des mesures inefficaces ou non-adaptées. Cette approche non-supervisée n'est pas spécifique au domaine du tourisme et elle peut être utilisée dans tous les domaines pour lesquels on dispose d'opinions textuelles ou

non sur les critères d'évaluation. Elle est aussi adaptée pour tenir compte de la hiérarchie du décideur : au niveau d'un hôtel (gérant), d'une chaîne d'hôtels (PDG), d'un territoire (responsable politique)...

Références

- [1] G. Adomavicius and Y. Kwon. New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems*, pages 48–55, 2007.
- [2] A. Chateauneuf and JY. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of möbius inversion. *Mathematical social sciences*, 17(3) :263–283, 1989.
- [3] B. Duthil. *From knowledge extraction to recommender system*. Theses, Ecole Nationale Supérieure des Mines d'Alès, December 2012.
- [4] B. Duthil, F. Troussset, G. Dray, J. Montmain, and Pascal P. Opinion extraction applied to criteria. In Stephen W. Liddle, Klaus-Dieter Schewe, A Min Tjoa, and Xiaofang Zhou, editors, *DEXA (2)*, volume 7447 of *Lecture Notes in Computer Science*, pages 489–496. Springer, 2012.
- [5] B. Duthil, F. Troussset, M. Roche, G. Dray, M. Plantié, J. Montmain, and P. Poncelet. Towards an automatic characterization of criteria. In *DEXA (1)*, pages 457–465, 2011.
- [6] M. Grabisch. The application of fuzzy integrals in multicriteria decision making. *European journal of operational research*, 89(3) :445–456, 1996.
- [7] M. Grabisch and Ch. Labreuche. How to improve acts : an alternative representation of the importance of criteria in mcdm. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(02) :145–157, 2001.
- [8] M. Grabisch and P. Perny. Agrégation multicritère. *Utilisations de la logique floue, Hermes*, 1999.
- [9] A. Imoussaten, B. Duthil, F. Troussset, and J. Montmain. Possibilistic interpretation of a sentiment analysis for a web recommender system. *LFA'2013, Reims*, 2013.
- [10] A. Imoussaten, B. Duthil, F. Troussset, and J. Montmain. A highly automated recommender system based on a possibilistic interpretation of a sentiment analysis. In *IPMU*, pages 536–545. Springer, 2014.
- [11] R. L. Keeney and H. Raiffa. *Decision analysis with multiple conflicting objectives*. Wiley & Sons, New York, 1976.
- [12] D.H. Krantz, Duncan R. Luce, P. Suppes, and A. Tversky. *Foundations of measurement (additive and polynomial representations)*, vol. 1, 1971.
- [13] Ch. Labreuche. Determination of the criteria to be improved first in order to improve as much as possible the overall evaluation. In *Int. Conf. IPMU 2004*, pages 609–616, Perugia, Italy, 2004.
- [14] JL. Marichal. Agrégation de critères interactifs au moyen de l'intégrale de choquet discrète. *LFA'99, Valenciennes, France*, pages 130–140, 1999.
- [15] J. Montmain, Ch. Labreuche, A. Imoussaten, and F. Troussset. Multi-criteria improvement of complex systems. *Information Sciences*, 291 :61–84, 2015.
- [16] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.