A formal justification of a simple aggregation function based on criteria and rank weights

Christophe Labreuche[1]

Abstract. In this paper, we have studied the problem of defining an aggregation function that is both easy to understand/explain and powerful enough to capture subtle decision strategies. The weighted sum and the ordered-weighted average are two simple and easy to understand aggregation functions, but with limited modelling power. They represent two orthogonal decision strategies: the importance of criteria for the first one and the rank orders for the second one. The question we address in this paper is the following one: is it possible to combine the expressive power of these two aggregation functions, while having a resulting aggregation function that remains very simple to understand? The solution is considered in the family of Choquet integrals. Our approach consists in formalizing this problem as a convex optimization problem, where the constraints are semantics assigned on the criteria weights and rank weights, and the cost function measures the specificity (entropy) of the solution. We provide the analytical solution to this problem and interpret its expression.

1 Introduction

There is a vast diversity of aggregation functions aiming to cover specific needs in different fields such as Multi-Criteria Decision Aiding (MCDA), Information Fusion, Classification, Social Choice among many others. In MCDA, aggregation functions represent the preferences of a decision maker concerning alternatives defined on several criteria. Usually, the identification of the right decision model follows from two stages: one first determines the suitable class of aggregation functions (e.g. the family of weighted sums), and then the parameters of the model (e.g. the weights in the weighted sum) are derived from learning data. The choice of the class of aggregation functions to be used, is governed by two conflicting principles. Firstly, the decision maker may express subtle decision strategies. Rich decision models are then required to represent these strategies. An example of such strategies is the presence of a veto criterion, or intolerance among criteria. These behaviors characterize the presence of interaction among criteria. The Choquet integral is a very versatile class of aggregation functions able to capture interacting criteria [4, 6]. According to the first principle, the richer the model the better. The second principle states that the decision model should be understandable to the user. This is complied with simple models such as the Weighted Sum (WS), or the Ordered Weighted Average (OWA) which recommendation can be easily grasped by a decision maker. We note that these two aggregation functions are particular cases of the Choquet integral. According to the second principle, the simpler the model the better.

It is very complex to find aggregation functions which are good regarding these two principles. For instance, the WS is the most widely used model in MCDA, due to its simplicity. It is only based on criteria weights composed on n values, representing the importance of each criterion, where n is the number of criteria. The recommendations obtained by such a model can be explained in an intuitive way [9, 2, 12]. However, its modelling power is very limited as it cannot represent any kind of interaction among criteria. The OWA operator is another example of function that can also be easily explained to a user. Its parameters are composed of rank weights with n components. The main practical interest of these weights is that they can be interpreted as a quantifier [36, 33]. It can model interacting criteria but all criteria are treated in a symmetric way. The WS and OWA operators satisfy the second principle but not the first one. On the opposite side, the Choquet integral w.r.t. a general capacity offers a tremendous versatility at the price of an exponential number of parameters – namely 2^n. The interpretation of a capacity is not easy in practice and its elicitation requires a significant amount of time.

There exist several intermediate models between the simple WS or OWA, and a general Choquet integral. The two-additive capacity can model the importance of criteria and interaction among only pairs of criteria, which yields $\frac{n(n+1)}{2}$ parameters. One needs then elaborate elicitation techniques due to this relative large number of parameters [22, 21, 5].

Another class of intermediate models combine the benefits of the WS and OWA operators, by using as parameters two weight vectors: weights p on criteria and weights w on ranks. There exist several proposals based on these two weight vectors: the *Weighted OWA* (WOWA) operator [31], the *Hybrid Weighted Averaging* (HWA) operator [32], the *Semi-Uninorm OWA* (SUOWA) operator [15], and the Ordered Weighted Averaging Weighted Average (OWAWA) operator [23]. HWA has a simple expression but fails to fulfil basic important properties in MCDA, such as idempotency. For this reason, HWA is not suitable in the context of MCDA. OWAWA is a simple linear combination of WS and OWA, and its interpretation is not so intuitive. On the other hand, WOWA and SUOWA operators have a quite complex expression, which is not intuitive for a decision maker. Moreover, the two weights vectors p and w are

[1] Thales Research & Technology 1 avenue Augustin Fresnel 91767 Palaiseau Cedex - France, email: christophe.labreuche@thalesgroup.com

combined in a way that is difficult to understand. At the end, one cannot readily see what is the contribution of p and w in the formula. Hence these proposals fail to satisfy the second principle on understandability.

In a previous paper [13], we proposed a new approach to define aggregation functions based on two weight vectors, that is understandable to the user. A key requirement is to make sure that the two weight vectors p and w have a clear semantics in the aggregation function. The existing approaches first define an aggregation formula parametrized by w and p, and then analyze their properties. We proceed in the opposite way. The idea is to consider a general capacity and impose some linear constraints to ensure clear interpretation of p and w.

Given the semantics of the two weights p and w represented as linear constraints, the aim of this paper is to identify a unique capacity. To this end, we adopt an elicitation approach. We indeed search for the least specific capacity fulfilling these constraints. This amounts to maximizing the entropy. When the monotonicity conditions are not saturated, we are able to show that there is a unique solution to this problem and to find the analytical expression of this solution. It is simply the weighted sum plus the OWA operator minus the average. In order to obtain an interpretable model, we rewrite this expression by removing the minus term. The so-obtained expression is simple and intuitive and can be used as an aggregation model to be learned from a training dataset. This paper provides thus a formal justification of this model.

2 Background

The set of criteria is denoted by $N = \{1, \ldots, n\}$. We are interested in an aggregation function $H : \mathbb{R}^N \to \mathbb{R}$.

2.1 Background on the Choquet integral

2.1.1 Capacities and Choquet integral

In order to generalize the weighted sum model, one may assign weights not only to single criteria but also to each subset of criteria.

Definition 1. *A* game *on $N = \{1, \ldots, n\}$ is a set function $\mu : 2^N \to [0, 1]$ such that $\mu(\emptyset) = 0$.*

Let $\mathcal{G}_N$ be the set of games on N.

Definition 2. *A* capacity *(also called* fuzzy measure*) on N is a game such that [3, 29]*

- *boundary condition: $\mu(N) = 1$,*
- *monotonicity: $\forall A \subseteq B \subseteq N$, $\mu(A) \leq \mu(B)$.*

Let $\mathcal{M}_N$ be the set of all capacities on N.

Several interesting particular cases of a capacity can be described.

Definition 3. *A capacity is said to be* additive *if $\mu(A \cap B) = \mu(A) + \mu(B)$ for every pair (A, B) of disjoint coalitions. A capacity is said to be* symmetric *if $\mu(A)$ depends only on the cardinality of A.*

The *Choquet integral* of $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ w.r.t. a capacity μ is defined by the following expression [3] :

$$C_\mu(a_1, \ldots, a_n) = \sum_{i=1}^{n} a_{\sigma(i)} \times \Big[\mu\left(\{\sigma(i), \cdots, \sigma(n)\}\right) - \mu\left(\{\sigma(i+1), \cdots, \sigma(n)\}\right)\Big], \qquad (1)$$

where σ is a permutation on N such that $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$, and $\mu\left(\{\sigma(i+1), \cdots, \sigma(n)\}\right) := 0$ when $i = n$. The Choquet integral has been proved to be able to model both the importance of criteria and the interaction between criteria.

The weighted sum is given by

$$\mathrm{WS}_p(a) = \sum_{i \in N} p_i \, a_i$$

where $p = (p_1, \ldots, p_n)$ is the weight of criteria. It corresponds to a Choquet integral with respect to an additive capacity: $\mu_{\mathrm{WS}_p}(S) = \sum_{i \in S} p_i$.

The *Ordered Weighted Average* (OWA) is defined by for any $a \in \mathbb{R}^N$ [34, 36]

$$\mathrm{OWA}_w(a) = \sum_{j=1}^{n} w_j \, a_{\sigma(n-j+1)}$$

where w_j are the weights allotted to the j^{th} largest value of vector a. It corresponds to a Choquet integral with respect to a symmetric capacity: $\mu_{\mathrm{OWA}_w}(S) = \sum_{j=1}^{|S|} w_j$.

2.1.2 Shapley value

It is not easy to interpret in a synthetic way a capacity, as it contains 2^n parameters. The most importance index helping to interpret a capacity is the concept of mean importance, defined as the Shapley value of the capacity [28]

$$\phi_i(\mu) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} \left(\mu(S \cup \{i\}) - \mu(S)\right)$$

2.1.3 Index of tolerance and intolerance

Capacities – more specifically belief functions which are infinitely monotone capacities – are used in decision under uncertainty to represent the attitude of the decision maker towards risk or uncertainty. A very classical strategy depicted by Ellsberg's paradox is risk aversion. The counterpart of risk aversion in MCDA is the concept of *intolerance*. Roughly speaking, a decision maker is intolerant if one needs to be good on most of criteria to have a good overall evaluation [19]. Marichal introduced tolerance and intolerance indices to formalize this idea [20].

Vetos and favors are situations of extreme intolerance and tolerance respectively. Criterion i is said to be a veto (resp. a favor) if $C_\mu(a) = 0$ (resp. $C_\mu(a) = 1$) whenever $a_i = 0$ (resp. $a_i = 1$) [4]. In other word, a bad evaluation on a veto criterion cannot be compensated by the other criteria. The concept of a veto can be generalized to subsets of cardinality k: A Choquet integral C_μ (or equivalently its underlying capacity μ) is *at* *most k-intolerant* if $C_\mu(a) = 0$ whenever $a_{\sigma(k)} = 0$ [19]. It is *k-intolerant* if, in addition, $C_\mu(a) \neq 0$ for some $a_{\sigma(k-1)} = 0$. One can prove that C_μ is at most k-intolerant if and only if $\mu(A) = 0$, $\forall A \subseteq N$ with $|A| \leq n-k$ [19]. The index of k-intolerance is the mean value of $C_\mu(a)$ over all a such that $a_{\sigma(k)} = 0$ [20]:

$$\mathrm{intol}_k(\mu) = \frac{n-k+1}{(n-k)\binom{n}{k}} \sum_{\substack{K\subseteq N\\|K|=k}} E(C_\mu(0_K, Z_{-K}))$$

where E denotes expectation, assuming that the random inputs $Z_1, \dots, Z_n$ are independent and uniformly distributed. This gives:

$$\mathrm{intol}_k(\mu) = 1 - \frac{1}{n-k} \sum_{t=0}^{n-k} \frac{1}{\binom{n}{t}} \sum_{\substack{T\subseteq N\\|T|=t}} \mu(T).$$

Likewise, a Choquet integral C_μ is *at most k-tolerant* if $C_\mu(a) = 1$ whenever $a_{\sigma(n-k+1)} = 1$ [19]. The index of k-tolerance is the mean value of $C_\mu(a)$ over all a such that $a_{\sigma(n-k+1)} = 1$ [20]:

$$\mathrm{tol}_k(\mu) = \frac{n-k+1}{(n-k)\binom{n}{k}} \sum_{\substack{K\subseteq N\\|K|=k}} E(C_\mu(1_K, Z_{-k})) - \frac{1}{n-k},$$

This gives:

$$\mathrm{tol}_k(\mu) = \frac{1}{n-k} \sum_{t=k}^{n} \frac{1}{\binom{n}{t}} \sum_{\substack{T\subseteq N\\|T|=t}} \mu(T) - \frac{1}{n-k}.$$

2.2 Existing proposals based on two weight vectors

We describe in this section the existing proposals based on two weight vectors: the criteria weights p and the rank weights w.

2.2.1 Wished properties

We start by giving some important properties that the aggregation function $H : \mathbb{R}^N \to \mathbb{R}$ shall satisfy. The following properties are taken from [7]:

- Continuity: H is continuous
- Idempotency: $H(\alpha, \dots, \alpha) = \alpha$ for every α
- Monotonicity: H is monotone
- Compensation: $\min_{i\in N} a_i \leq H(a) \leq \max_{i\in N} a_i$. Note that this property follows directly from Idempotency and Monotonicity

Consider now an aggregation function $H_{p,w}$ based on the two weights p and w. The following property is wished:

- Generalization of WS_p and OWA_w [15]: One says that $H_{p,w}$ generalizes both the WS and OWA, if $H_{p,\eta} = \mathrm{WS}_p$ and $H_{\eta,w} = \mathrm{OWA}_w$, where η denotes the uniform vector $(\frac{1}{n}, \dots, \frac{1}{n})$.

2.2.2 WOWA

Torra introduced an aggregation function based the criteria weights p and the rank weights w. It is called *Weighted OWA* (WOWA) operator [31]. It depends also on a quantifier $Q : [0,1] \to [0,1]$, which a non-decreasing function such that $Q(0) = 0$ and $Q(1) = 1$. The quantifier Q is defined from rank weights w in the following way [35, 36]

$$Q\left(\frac{k}{n}\right) - Q\left(\frac{k-1}{n}\right) = w_k, \quad \text{for } k = 1, \dots, n.$$

The WOWA has the following expression:

$$\mathrm{WOWA}^Q_{p,w}(a) = \sum_{j=1}^{n} q_j \, a_{\tau(n-j+1)}$$

where

$$q_j = Q\left(\sum_{k=1}^{j} p_{\tau(n-k+1)}\right) - Q\left(\sum_{k=1}^{j-1} p_{\tau(n-k+1)}\right)$$

A WOWA is particular case of a Choquet integral for the following capacity

$$\mu_{\mathrm{WOWA}^Q_{p,w}}(S) = Q\left(\sum_{i\in S} p_i\right).$$

2.2.3 SUOWA

Another aggregation function constructed from the two vectors p and w has recently been proposed by Llamazares [15, 16]. Its definition is based a semi-uninorm and is thus called *Semi-Uninorm OWA* operator (SUOWA in short). A *semi-uninorm* is a mapping $U : [0,1]^2 \to [0,1]$ if it is monotonic and possesses a neutral element $e \in [0,1]$ (such that $U(e,x) = U(x,e) = x$ for every x).

Let us define a set function v by

$$v(S) = |S| \, U\left(\frac{\mu_{\mathrm{WS}_p}(S)}{|S|}, \frac{\mu_{\mathrm{OWA}_w}(S)}{|S|}\right) \tag{2}$$

The factor $\frac{1}{|S|}$ comes from the fact that when $p = \eta$ (resp. $w = \eta$), $\frac{\mu_{\mathrm{WS}_p}(S)}{|S|}$ (resp. $\frac{\mu_{\mathrm{OWA}_w}(S)}{|S|}$) is independent of S and is always equal to $\frac{1}{n}$. Hence property "generalization of WS_p and OWA_w" is satisfied, if the neutral element is $e = \frac{1}{n}$.

The drawback with this expression is that it may be non-monotone. Hence Llamazares considers the monotonic cover of v:

$$\mu_{\mathrm{SUOWA}^U_{p,w}}(S) = \max_{T\subseteq S} v(T).$$

The monotonic cover can be computed recursively [16]

$$\mu_{\mathrm{SUOWA}^U_{p,w}}(S) = \max\left(v(S), \max_{i\in S} \; \mu_{\mathrm{SUOWA}^U_{p,w}}(S \setminus \{i\})\right).$$

Then the SUOWA is the Choquet integral with respect to $\mu_{\mathrm{SUOWA}^U_{p,w}}$.

Set-function v is bounded by 1 iff $U \in \widetilde{\mathcal{U}}^{\frac{1}{n}}$, where $\widetilde{\mathcal{U}}^{\frac{1}{n}}$ is the set of semi-uninorms with neutral element $e = \frac{1}{n}$ such that $U(\frac{1}{k}, \frac{1}{k}) \leq \frac{1}{k}$ for all $k \in \mathbb{N}$ [16, Proposition 5]. In the context of MCDA, we will generally consider idempotent semi-uninorms. They are necessarily elements in $\widetilde{\mathcal{U}}^{\frac{1}{n}}$. There exists in the literature many examples of semi-uninorms with neutral element $e = \frac{1}{n}$ [17]. Let us mention the following one:

$$U_{T_L}(x, y) = \begin{cases} \max(x, y) & \text{if } (x, y) \in [\frac{1}{n}, 1]^2, \\ \max(x + y - \frac{1}{n}, 0) & \text{otherwise} \end{cases}$$

With this semi-uninorm, if $\min_{i\in N} p_i + \min_{1\leq j\leq n} w_j \geq \frac{1}{n}$, and for every $1 \leq k \leq n$, either $\frac{1}{k}\sum_{1\leq j\leq k} w_j \leq \frac{1}{n}$ or for all $S \subseteq N$ with $|S| = k$, $\frac{1}{k}\sum_{i\in S} p_i \leq \frac{1}{n}$, then

$$\mu_{\mathrm{SUOWA}^{U_{T_L}}_{p,w}}(S) = \mu_{\mathrm{WS}_p}(S) + \mu_{\mathrm{OWA}_w}(S) - \frac{|S|}{n} \qquad (3)$$

2.2.4 Other proposals

Other proposals based on two weights have been proposed in the literature. The hybrid weighted averaging (HWA) amounts to applying the OWA operator to the vector of the scores multiplied by the corresponding criteria weights [32]. A very close concept has been introduced by B. Roy, under the name of Ordered Average with Two Weights (OA2W) – called MO2P (Moyenne Ordonnée Doublement Pondérée) in French [25, 26]. The Ordered Weighted Averaging Weighted Average (OWAWA) operator is simply a convex combination of a weighted sum and an OWA operator [23].

2.3 Entropy of a capacity

The entropy is a well-known concept defined in information theory. The Shannon entropy of a probability distribution with probabilities $p_1, \ldots, p_n$ over n possible outcomes is defined by [27]

$$H_{\mathrm{Sh}}(p) := \sum_{i=1}^{n} h(p_i) \ ,$$

where $h(u) = -u \ \ln u$ with the convention $0 \ln 0 := 0$. In the case of partial knowledge on the probabilities, the probability distribution that is chosen is the one that maximizes the entropy. A nice property of this maximization is that one gets always a unique solution thanks to the convexity of the entropy functional.

The concept of entropy has been extended to capacities. Let $\mathcal{Q}(N)$ denote the set of maximal chains of the Hasse diagram $\left(2^N, \subseteq\right)$. Recall that a maximal chain in $\left(2^N, \subseteq\right)$ is a sequence $\emptyset, \{\tau(1)\}, \{\tau(1), \tau(2)\}, \ldots, \{\tau(1), \ldots, \tau(n)\}$ where τ is a permutation of N. On each maximal chain, one can define the following probability distribution: $p^{\tau}_{\tau(i)} = \mu(\{\tau(i), \ldots, \tau(n)\}) - \mu(\{\tau(i+1), \ldots, \tau(n)\})$. The entropy of μ is then defined as the mean value of the entropy of each maximal chain [18] :

$$\hat{H}_{\mathrm{Sh}}(\mu) := \frac{1}{n!}\sum_{\tau} H_{\mathrm{Sh}}(p^{\tau})$$
$$= \sum_{i=1}^{n} \sum_{A\subseteq N\setminus\{i\}} \gamma^n_{|A|} h\left(\delta_i \mu(A)\right),$$

where $\gamma^n_p = \frac{(n-p-1)!p!}{n!}$ and $\delta_i\mu(A) := \mu(A \cup \{i\}) - \mu(A)$. The problem of identifying a capacity can be solved by maximizing the entropy of the capacity given some preference information.

Other entropy can be used. The Rényi entropy of a probability distribution is defined by $\frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} {p_i}^{\alpha}\right)$, where $\alpha > 0$ is a parameter different from 1. For $\alpha = 2$, we obtain $-\log\left(\sum_{i=1}^{n} {p_i}^{2}\right)$. We can simplify this expression to obtain a quadratic function, taking $h(u) = -u^2$ [10] :

$$\hat{H}^2_{\mathrm{Re}}(\mu) := -\sum_{i=1}^{n} \sum_{A\subseteq N\setminus\{i\}} \gamma^n_{|A|} \left(\delta_i\mu(A)\right)^2 .$$

Finally Kojadinovic's variance is obtained by taking $h(u) = \left(u - \frac{1}{n}\right)^2$:

$$\hat{V}(\mu) := \sum_{i=1}^{n} \sum_{A\subseteq N\setminus\{i\}} \gamma^n_{|A|} \left(\delta_i\mu(A) - \frac{1}{n}\right)^2 .$$

2.4 Estimation of a capacity under incomplete information

One of the main issue in preference learning is the identification of the capacity given some preference information. In most of cases, the preference information takes the form of pairwise comparisons between alternatives (ranking problem), and assignment of alternatives to a class (sorting problem)

When the amount of input information is large, one can adopt statistical machine learning techniques, such as extensions of logistic regression. A noticeable example is the Choquistic regression [30].

When the input data is scarce, preference information is usually represented as (possibly hard) constraints. One then looks for a capacity fulfilling these constraints. We are now under incomplete information in the sense that the preference information is far from uniquely specifying the capacity. Several strategies can be adopted to determine a unique capacity. The first one consists in separating as much as possible the training examples provided by the decision maker. This is the approach considered in SVM. In [24, 5], when the decision maker says that he prefers an alternative a to another alternative b, he will not be completely happy if the two options turn out to have very close evaluations. One then seeks to maximize the difference of score between these two alternatives. The preference information are translated into constraints. These latter?s define a polyhedron. Maximizing the difference of score amounts to being as far as possible to the boundary of the constraints, that is the facets of the polyhedron. This approach amounts thus to looking for the midpoint in the polyhedron. The *most representative* parameter vector is a generalization of this approach [8].

Another venue is to find the capacity that is the least specific, i.e. that maximizes the entropy [18, 10]. The objective here is to have the simplest possible capacity fulfilling the preference information. In particular, if there is an additive capacity satisfying the preference information, then one shall find this. This is the least commitment principle.

Depending on the application, one of these approaches shall be chosen.

3 Definition of a capacity derived from two weights

We are given two weight vectors: the criteria weights $p = (p_1, \ldots, p_n)$, and the rank weights $w = (w_1, \ldots, w_n)$. As we have seen, a major problem with the existing proposals combining these two weights vectors is about the interpretation of the weights p and w.

The existing approaches first define an aggregation formula parametrized by w and p, and then analyze their properties. We proceed in the opposite way. We start by providing some properties that we would like to have for an aggregation function based on the two weight vectors p and w. These properties will be considered as constraints on the capacity μ. One obtains thus a set of admissible capacities, among which one may select a unique one, maximizing the entropy or minimizing the variance.

The aim of this section is to provide a clear interpretation of p and w, and derive constraints on the capacity from these interpretations. Intuitively, the two vectors p and w are orthogonal in their interpretation, as p relates the importance of criteria and w relates to the type of interaction among criteria. This section is based on [13].

3.1 Constraints on p

The weight vector p represents the relative importance of criteria in the weighted sum WS_p. Coefficient is then the importance of criterion i. To give a more precise interpretation of p_i, we can use the MACBETH approach [1]. In this approach, the criteria weights cannot be defined independently of the criteria scales. One thus needs to define two *levels* on each criterion. We can use the two standard 0 and 1 levels – where 0 (resp. 1) means that criterion i is not met at all (resp. completely satisfactory). The importance p_i of criterion i is then defined as the added-value on the overall score going from the lower level 0 to the upper level 1 on criterion i, the value on the other criteria being fixed – e.g. to 0:

$$p_i = \mathrm{WS}_p(1_i, 0_{-i}) - \mathrm{WS}_p(0, \dots, 0). \tag{4}$$

We note that the value on criteria $N \setminus \{i\}$ can be changed to any other value, with no consequence on the result. This is due to the fact that all criteria are independent in the weighted sum.

Let us now consider a capacity μ. Then parameter p_i shall be naturally interpreted as the mean importance of criteria i in μ. There exists several generalizations of (4) to capacities. The leading concept to interpret the mean importance of criteria is the Shapley value. As a result, the Shapley value of μ shall be equal to p:

$$\forall i \in N \qquad \phi_i(\mu) = p_i. \tag{5}$$

The choice of the Shapley value comes from Cooperative Game Theory. Relation (5) can also be justified taking a variational standpoint. Indeed, the importance index of a criterion should reflect to which extent criterion i influences the overall score C_μ. This influence is naturally measured by the amount of variation this criterion produces on C_μ. As this influence is in general dependent on the value of the other attributes $N \setminus \{i\}$, one shall take the average value over these attributes. It has been shown that the mean value over $[0,1]^n$ of the partial derivative of the Choquet integral w.r.t. variable i is equal to $\phi_i(\mu)$ [7][sect. 10.3]. This provides another justification of (5).

3.2 Constraints on w

The rank weight vector w that we wish to interpret, is the parameter vector of the OWA operator OWA_w. In an OWA operator, all criteria are symmetric and have thus the same importance. Hence w describe only the way criteria are interacting together.

The $*$-intolerant indices depict interaction among criteria, and are not dependent on a particular pair or subset of criteria – unlike the interaction indices. Hence the k-intolerant index might be suitable to interpret the weights w. In order to check this, let compute the k-intolerant index on the OWA operator OWA_w:

$$\begin{aligned}
\mathrm{intol}_k(\mu_{\mathrm{OWA}_w}) &= 1 - \frac{1}{n-k} \sum_{t=0}^{n-k} \frac{1}{\binom{n}{t}} \sum_{\substack{T \subseteq N \\ |T|=t}} \mu_{\mathrm{OWA}_w}(T) \\
&= 1 - \frac{1}{n-k} \sum_{t=0}^{n-k} \sum_{j=1}^{t} w_j = 1 - \frac{1}{n-k} \sum_{t=1}^{n-k} t\, w_t.
\end{aligned}$$

There is a clear relation between the k-intolerant index and the value of the $n-k$ first values in the weight vector w. However, this relation is not so trivial, and it will not be convenient as a constraint.

An essential property of the weights w is the following [14]: $\mathrm{OWA}_w(1_S, 0_{N\setminus S}) = w_1 + w_2 + \cdots + w_s$ for any $S \subseteq N$ with $|S| = s$. This property generalizes relation (4) which is at the root of the MACBETH approach. We note that $\mathrm{OWA}_w(1_S, 0_{N\setminus S})$ is equal to the OWA capacity μ_{OWA_w} at S. In order to generalize this property to a non-symmetric capacity, we just need to replace $\mathrm{OWA}_w(1_S, 0_{N\setminus S})$ by the average value of $\mu(S)$ over all subset of cardinality s. This corresponds to the term $\frac{1}{\binom{n}{t}} \sum_{\substack{T \subseteq N \\ |T|=t}} \mu(T) =: A_t(\mu)$ in the expression of intol_k. There is a simple linear relation between the $*$-intolerant indices and the A_* indices:

$$\begin{aligned}
&\mathrm{intol}_1(\mu) = 1 - \frac{1}{n-1} \sum_{t=1}^{n-1} A_t(\mu) \\
&\dots \\
&\mathrm{intol}_k(\mu) = 1 - \frac{1}{n-k} \sum_{t=1}^{n-k} A_t(\mu) \\
&\dots \\
&\mathrm{intol}_{n-1}(\mu) = 1 - A_1(\mu)
\end{aligned}$$

Enforcing constraints on the $*$-intolerant indices or on the A_* indices is completely equivalent. Due to the simple relations $A_t(\mu_{\mathrm{OWA}_w}) = \sum_{j=1}^{t} w_j$, we choose to interpret the weights w from the A indices. Note that $A_t(\mu)$ is the average value of an alternative being very good on t criteria and very bad on the remaining ones.

Hence we obtain the following constraints for the interpretation of the w order weights:

$$\forall s \in \{1, \dots, n-1\} \qquad \frac{1}{\binom{n}{s}} \sum_{S \subseteq N \,:\, |S|=s} \mu(S) = \sum_{k=1}^{s} w_k. \tag{6}$$

Note that the equation with $s = n$ has been removed as it is trivially satisfied for every capacity.

4 Formal definition of a simple aggregation function based on criteria and rank weights

This section proposes to define a simple aggregation function based on criteria and rank weights as an optimization problem. The optimization problem is described in Section 4.1. Then the analytical solution to this problem is found in Section 4.3. Finally an interpretation of this solution is proposed in Section 4.4.

4.1 Set of admissible capacities

In sum, the capacities that provide a clear interpretation of the two weight vectors p and w are the ones that fulfil both (5) and (6).

We note that neither WOWA nor SUOWA satisfy these properties in the general case.

We note that the n relations in (5) are not independent, as $\sum_{k=1}^{n} p_k = 1$. It is sufficient to consider only $n-1$ of them. It is easier to have only independent inequalities. W.l.o.g. we remove the last one and we set $\overline{N} = \{1, \ldots, n-1\}$. Given the previous interpretation of p and w, we consider all capacities fulfilling (5) and (6):

$$\mathcal{M}_N(p,w) := \Big\{ \mu \in \mathcal{M}_N \ : \\ \forall i \in \overline{N} \quad \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} \left(\mu(S \cup \{i\}) - \mu(S)\right) = p_i, \\ \forall s \in \{1, \ldots, n-1\} \quad \frac{1}{\binom{n}{s}} \sum_{S \subseteq N \,:\, |S|=s} \mu(S) = \sum_{k=1}^{s} w_k \Big\}.$$

Any capacity in this set is admissible.

4.2 Finding the less specific admissible capacity

The question that now arises is which capacity within $\mathcal{M}_N(p,w)$ shall be chosen. Following Section 2.4, two approaches can be adopted.

The first one is to separate as much as possible the preference information or consider the most representative parameter vector. We note that $\mathcal{M}_N(p,w)$ is the set of capacities fulfilling a number of equality constraints. Hence, it is not possible to further separate the constraint. The only remaining possiility is to separate the monotonicity conditions in $\mathcal{M}_N(p,w)$. This means that we would like to maximize the difference $\mu(S \cup \{i\}) - \mu(S)$ for every $i \in N$ and every $S \subseteq N \setminus \{i\}$. There is objectively no reason to do so. As a reminder, the initial motivation of [24, 5] is that when the decision maker expresses that an alternative is preferred to another one, he expects some significant difference between these two differences. But in the context of $\mathcal{M}_N(p,w)$, the decision maker does not expect any separation of the monotonicity conditions. We then look at the second approach to select an element in $\mathcal{M}_N(p,w)$.

When the available preference information is very sparse, a safe way to use this information is to consider the least specific model that fulfils this information. This has led to the *least commitment principle* aiming at maximizing the entropy. In the context of the Choquet integral, Kojadinovic *et al* have proposed either to maximize the entropy [11] or to minimize the variance [10].

In the context of the double weights p and w, we aim thus at finding the element in $\mathcal{M}_N(p,w)$ which maximizes some entropy. We rewrite here as the minimization of a quantity $\sum_{i\in N} \sum_{S \subseteq N\setminus\{i\}} \frac{|S|!(n-|S|-1)!}{n!} E(\mu(S \cup \{i\}) - \mu(S))$, where E is the opposite of some entropy.

This yields the following convex optimization problem under linear constraints:

$$\begin{array}{ll} \text{Minimize} & \displaystyle\sum_{i\in N} \sum_{S \subseteq N\setminus\{i\}} \frac{|S|!(n-|S|-1)!}{n!} E(\mu(S \cup \{i\}) - \mu(S)) \\ \text{under} & \left| \begin{array}{l} \forall i \in N \ \forall S \subseteq N \setminus \{i\} \quad \mu(S \cup \{i\}) \geq \mu(S) \\ \forall i \in \overline{N} \\ \quad \sum_{S \subseteq N\setminus\{i\}} \frac{|S|!(n-|S|-1)!}{n!} \left(\mu(S \cup \{i\}) - \mu(S)\right) = p_i \\ \forall s \in \{1, \ldots, n-1\} \\ \quad \frac{1}{\binom{n}{s}} \sum_{S \subseteq N \,:\, |S|=s} \mu(S) = \sum_{k=1}^{s} w_k \end{array} \right. \end{array} \tag{7}$$

where

$$\begin{aligned} E(u) &= u \ln(u) \quad \text{for the Shannon entropy} \\ E(u) &= u^2 \quad \text{for the Renyi entropy with } \alpha = 2 \\ E(u) &= \left(u - \frac{1}{n}\right)^2 \quad \text{for the variance} \end{aligned}$$

This problem is a convex optimization problem under linear constraints.

4.3 Main result

Optimization problem (7) can be very complex due to the numerous constraints. We wish in this section to provide the analytical solution. In the general case, the analytical solution cannot be done and one must use an optimization algorithm. In order to be able to provide an analytical solution, we assume that the monotonicity conditions are not saturated, i.e. that the capacity is strictly monotone. This yields to the following problem.

$$\begin{array}{ll} \text{Minimize} & \displaystyle\sum_{i\in N} \sum_{S \subseteq N\setminus\{i\}} \frac{|S|!(n-|S|-1)!}{n!} E(\mu(S \cup \{i\}) - \mu(S)) \\ \text{under} & \left| \begin{array}{l} \forall i \in N \ \forall S \subseteq N \setminus \{i\} \quad \mu(S \cup \{i\}) > \mu(S) \\ \forall i \in \overline{N} \\ \quad \sum_{S \subseteq N\setminus\{i\}} \frac{|S|!(n-|S|-1)!}{n!} \left(\mu(S \cup \{i\}) - \mu(S)\right) = p_i \\ \forall s \in \{1, \ldots, n-1\} \\ \quad \frac{1}{\binom{n}{s}} \sum_{S \subseteq N \,:\, |S|=s} \mu(S) = \sum_{k=1}^{s} w_k \end{array} \right. \end{array} \tag{8}$$

The following result gives the expression of the solution to this problem when E is taken as the variance.

Proposition 1. *Consider the optimization problem (7), in which the variance is used for E. Assume furthermore that the inequality constraints are not saturated in (7) (we obtain thus (8)). Then there is a unique solution to this problem, which corresponds to the following aggregation function:*

$$\mathrm{WS}_p + \mathrm{OWA}_w - \mathrm{AM}, \tag{9}$$

where $\mathrm{AM}(a) = \frac{\sum_{i\in N} a_i}{n}$.

The proof of this result is omitted due to space limitation.

Aggregation function (9) is a Choquet integral w.r.t. to the following capacity:

$$\mu(S) = \mu_{\mathrm{WS}_p}(S) + \mu_{\mathrm{OWA}_w}(S) - \mu_{\mathrm{AM}}(S) \tag{10}$$

$$= \sum_{i\in S} p_i + \sum_{k=1}^{|S|} w_k - \frac{|S|}{n} \tag{11}$$

for all $S \subseteq N$. We note that this corresponds to (3) and is thus a particular case of a SUOWA operator.

4.4 Interpretation of the solution

We denote by $H_{p,w}$ the aggregation function (9) obtained in the main result of this paper.

Parameters p and w satisfy the normalization conditions:

$$\forall i \in N \quad p_i \geq 0 \text{ and } \sum_{k\in N} p_k = 1, \tag{12}$$

$$\forall i \in N \quad w_i \geq 0 \text{ and } \sum_{k\in N} w_k = 1. \tag{13}$$

Let us now check whether (9) satisfies the conditions of Section 2.2.1. Let us start with the monotonicity conditions. We have that $\mu(S\cup\{i\}) = H(1_{S\cup\{i\}}, 0_{-S\cup\{i\}}) = \sum_{j\in S\cup\{i\}} p_j + \sum_{k=1}^{s+1} w_k - \frac{s+1}{n}$ and $\mu(S) = H(1_S, 0_{-S}) = \sum_{j\in S} p_j + \sum_{k=1}^{s} w_k - \frac{s}{n}$. Thus $\mu(S\cup\{i\}) - \mu(S) = H(1_{S\cup\{i\}}, 0_{-S\cup\{i\}}) - H(1_S, 0_{-S}) = p_i + w_{s+1} - \frac{1}{n}$. As a result, the monotonicity conditions are fulfilled with (9) iff

$$\min_{i\in N} p_i + \min_{k\in N} w_k \geq \frac{1}{n}. \tag{14}$$

Moreover, (9) is clearly continuous, idempotent and compensatory. Lastly (9) generalizes WS_p and OWA_w as $H_{p,\eta} = \mathrm{WS}_p$ and $H_{\eta,w} = \mathrm{OWA}_w$. Hence all conditions of Section 2.2.1 are fulfilled under condition (14).

Let us now discuss on the interpretation of (9). Even though (9) is composed of three aggregation functions WS_p, OWA_w and AM, which are individually very simple to understand, the overall expression is not so simple as it contains a negative term, which is not so simple to explain. The reason is that it violates monotonicity.

We wish to rewrite (9) with only non-negative terms. We define parameters $v^p = (v^p_1, \dots, v^p_n)$ and $v^w = (v^w_1, \dots, v^w_n)$ such that

$$\forall i \in N \quad v^p_i \geq 0 \text{ and } v^w_i \geq 0, \tag{15}$$

$$\sum_{k\in N} v^p_k + \sum_{k\in N} v^w_k = 1. \tag{16}$$

We define

$$A_{v^p,v^w}(a) = \sum_{i\in N} v^p_i\, a_i + \sum_{i\in N} v^w_i\, a_{\tau(n_i+1)}.$$

The next result shows that models $H_{p,\eta}$ and A_{v^p,v^w} are equivalent.

Proposition 2. *For every p, w satisfying (12), (13) and (14), there exists v^p and v^w satisfying (15) and (16) such that $H_{p,\eta} \equiv A_{v^p,v^w}$. Conversely for every v^p and v^w satisfying (15) and (16), there exists p, w satisfying (12), (13) and (14) such that $H_{p,\eta} \equiv A_{v^p,v^w}$.*

Proof : Consider p, w satisfying (12), (13) and (14). Let $\underline{p} := \min_{i\in N} p_i$, $\underline{w} := \min_{i\in N} w_i$ and $\Delta := \underline{p} + \underline{w} - \frac{1}{n}$. We have $\Delta \geq 0$ by (14).

We set $v^p_i := p_i - \underline{p} + \frac{\Delta}{2}$ and $v^w_i := w_i - \underline{w} + \frac{\Delta}{2}$. Then

$$\begin{aligned} A_{v^p,v^w}(a) &= \sum_{i\in N} v^p_i\, a_i + \sum_{i\in N} v^w_i\, a_{\tau(n-i+1)} \\ &= \sum_{i\in N} p_i\, a_i + \sum_{i\in N} w_i\, a_{\tau(n-i+1)} + (\Delta - \underline{p} - \underline{w}) \sum_{i\in N} a_i \\ &= H_{p,\eta}(a). \end{aligned}$$

On the other hand, as $\Delta \geq 0$, relation (15) is fulfilled. Finally, $\sum_{k\in N} v^p_k + \sum_{k\in N} v^w_k = \sum_{k\in N} p_k + \sum_{k\in N} w_k + n(\Delta - \underline{p} - \underline{w}) = 1$. Hence (16) is satisfied.

Conversely, let v^p and v^w satisfying (15) and (16). We set $p_i = v^p_i + \frac{1}{n}\sum_{k\in N} v^w_k$ and $w_i = v^w_i + \frac{1}{n}\sum_{k\in N} v^p_k$. Then

$$\begin{aligned} H_{p,\eta}(a) &= \sum_{i\in N} p_i\, a_i + \sum_{i\in N} w_i\, a_{\tau(n-i+1)} - \frac{1}{n}\sum_{i\in N} a_i \\ &= \sum_{i\in N} v^p_i\, a_i + \sum_{i\in N} v^w_i\, a_{\tau(n-i+1)} \\ &+ \left(\frac{1}{n}\sum_{i\in N} v^p_i + \frac{1}{n}\sum_{i\in N} v^w_i - \frac{1}{n} \right) \sum_{i\in N} a_i \\ &= A_{v^p,v^w}(a). \end{aligned}$$

On the other hand, $p_i \geq 0$ and $w_i \geq 0$. Moreover, $\sum_{i\in N} p_i = \sum_{i\in N} v^p_i + \sum_{k\in N} v^w_k = 1$ and $\sum_{i\in N} w_i = \sum_{i\in N} v^w_i + \sum_{k\in N} v^p_k = 1$. Finally,

$$\begin{aligned} p_i + w_k - \frac{1}{n} &= v^c_i + v^w_k + \frac{1}{n}\sum_{j\in N} v^p_j + \frac{1}{n}\sum_{j\in N} v^w_j - \frac{1}{n} \\ &= v^c_i + v^w_k \geq 0 \end{aligned}$$

Hence p, w satisfy (12), (13) and (14). ■

Aggregation function A_{v^p,v^w} is very simple to interpret and understand.

Example 1. *For $p = (0.1, 0.1, 0.1, 0.1, 0.6)$ and $w = (0.1, 0.1, 0.1, 0.1, 0.6)$, monotonicity condition (14) is fulfilled. Moreover, $H_{p,w}$ can be rewritten as $H_{p,w}(a) = 0.5a_5 + 0.5a_{\tau(1)}$.*

Following p and w, the emphasis is put on criterion 5 and also in the best satisfied criteria. This is recovered by the previous expression. No other term than a_5 and $a_{\tau(1)}$ appear in H as the monotonicity condition are saturated.

If values 0.6 are decreased in p and w, other terms will appear.

5 Conclusion

In this paper, we have studied the problem of defining an aggregation function that is both easy to understand/explain and powerful enough to capture subtle decision strategies. Our starting point is to generalize both the weighted sum and the OWA, characterized by their parameters – namely the criteria weights and the rank weights. We consider the family of Choquet integrals and we aim at looking for a simple capacity define from criteria and rank weights. We look at this problem as an optimization problem. The constraints are the monotonicity conditions on the capacity and conditions on the weights. More precisely, the Shapley value of the capacity shall be equal to the criteria weights, and the rank weights shall be derived from the intolerance indices. The optimization functional is some kind of entropy as we look for the simplest and thus less specific capacity. We are especially interested in the case where the monotonicity constraints are not saturated and the objective function of the variance of the capacity. We are thus able to show that the so-obtained optimization problem has a unique solution. We can derive the analytical expression of the unique solution: it is equal to the weighted sum of the criteria weights plus the OWA of the rank weights minus the average. Lastly we can rewrite this aggregation functions to make it simple to understand, getting rid of the minus term. We simply obtain a weighted sum plus an OWA operator, where the sum of all their coefficients sum-up to one. This paper can be seen as a formal justification of this latter aggregation model.

The model obtained in the main result is very simple to understand. The price to pay is that it is much less versatile than a general Choquet integral. Its representation power is anyhow sufficient when the decision maker only provides importance and rank order weights.

REFERENCES

[1] C. A. Bana e Costa, J.M. De Corte, and J.-C. Vansnick, 'MACBETH', *International Journal of Information Technology and Decision Making*, **11**, 359–387, (2012).

[2] G. Carenini and J.D. Moore, 'Generating and evaluating evaluative arguments', *Artificial Intelligence*, **170**, 925–952, (2006).

[3] G. Choquet, 'Theory of capacities', *Annales de l'Institut Fourier*, **5**, 131–295, (1953).

[4] M. Grabisch, 'The application of fuzzy integrals in multicriteria decision making', *European J. of Operational Research*, **89**, 445–456, (1996).

[5] M. Grabisch, I. Kojadinovic, and P. Meyer, 'Using the Kappalab R package for Choquet integral based multi-attribute utility theory', in *Proc. Int. Conf. on Information Processing and Management of Unicertainty (IPMU'06)*, pp. 1702–1709, Paris, France, (July 2006).

[6] M. Grabisch and Ch. Labreuche, 'A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid', *Annals of Operation Research*, **175**, 247–286, (2010).

[7] M. Grabisch, J.L. Marichal, R. Mesiar, and E. Pap, *Aggregation functions*, Cambridge University Press, 2009.

[8] M. Kadzinski, S. Greco, and R. Slowinski, 'Selection of a representative value function in robust multiple criteria ranking and choice', *European Journal of Operational Research*, **217**, 541–553, (2012).

[9] D.A. Klein, *Decision analytic intelligent systems: automated explanation and knowledge acquisition*, Lawrence Erlbaum Associates, 1994.

[10] I. Kojadinovic, 'Minimum variance capacity identification', *Quaterly J. of Operations Research (4OR)*, **12**, 23–36, (2006).

[11] I. Kojadinovic, J.-L. Marichal, and M. Roubens, 'An axiomatic approach to the definition of the entropy of a discrete Choquet capacity', *Information Sciences*, **172**, 131–153, (2005).

[12] Ch. Labreuche, 'A general framework for explaining the results of a multi-attribute preference model', *Artificial Intelligence*, **175**, 1410–1448, (2011).

[13] Ch. Labreuche, 'On capacities characterized by two weight vectors', in *16th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, Eindhoven, The Netherland, (2016).

[14] Ch. Labreuche, B. Mayag, and B. Duqueroie, 'Extension of the MACBETH approach to elicit an OWA operator', *EURO J. on Decision Processes*, **3**, 65–105, (2015).

[15] B. Llamazares, 'An analysis of some functions that generalize weighted means and OWA operators', *Int. J. of Intelligent Systems*, **28**, 380–393, (2013).

[16] B. Llamazares, 'Constructing Choquet integral-based operators that generalize weighted means and OWA operators', *Information Fusion*, **23**, 131–138, (2015).

[17] Bonifacio Llamazares, 'SUOWA operators: Constructing semiuninorms and analyzing specific cases', *Fuzzy Sets and Systems*.

[18] J.-L. Marichal, 'Aggregation of interacting criteria by means of the discrete Choquet integral', in *Aggregation operators: new trends and applications*, eds., T. Calvo, G. Mayor, and R. Mesiar, volume 97 of *Studies in Fuzziness and Soft Computing*, 224–244, Physica Verlag, (2002).

[19] J.-L. Marichal, 'Tolerant or intolerant character of interacting criteria in aggregation by the Choquet integral', *European J. of Operational Research*, **155**(3), 771–791, (2004).

[20] J.-L. Marichal, 'k-intolerant capacities and Choquet integrals', *European J. of Operational Research*, **177**(3), 1453–1468, (2007).

[21] B. Mayag, M. Grabisch, and Ch. Labreuche, 'An interactive algorithm to deal with inconsistencies in the representation of cardinal information', in *Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pp. 148–157, Dortmund, Germany, (june 28-July 02 2010).

[22] B. Mayag, M. Grabisch, and Ch. Labreuche, 'A representation of preferences by the Choquet integral with respect to a 2-additive capacity', *Theory and Decision*, **71**, 297–324, (2011).

[23] J.M. Merigo, 'A unified model between the weighted average and the induced OWA operator', *Expert Systems and Applications*, **38**, 11560–11572, (2011).

[24] P. Meyer and M. Roubens, 'Choice, ranking and sorting in fuzzy multiple criteria decision aid', in *Multiple Criteria Decision Analysis*, eds., J. Figueira, S. Greco, and M. Ehrgott, 471–503, Kluwer Academic Publishers, (2005).

[25] B. Roy, 'Les logiques compensatoires et les autres', Technical report, Note de Recherche LAMSADE No 16, University of Paris Dauphine, (1996).

[26] B. Roy, 'Double pondération pour calculer une moyenne: pourquoi et comment?', *RAIRO Operations Research*, **41**(2), 125–139, (2007).

[27] C.E. Shannon, 'A mathematical theory of communication', *Bell System Tech. J.*, **27**, 379–423, (1948).

[28] L. S. Shapley, 'A value for n-person games', in *Contributions to the Theory of Games, Vol. II*, eds., H. W. Kuhn and A. W. Tucker, number 28 in Annals of Mathematics Studies, 307–317, Princeton University Press, (1953).

[29] M. Sugeno, 'Fuzzy measures and fuzzy integrals', *Trans. S.I.C.E.*, **8**(2), (1972).

[30] A. Fallah Tehrani, W. Cheng, K. Dembczynski, and E. Hüllermeier, 'Learning monotone nonlinear models using the Choquet integral', *Machine Learning*, **89**, 183–211, (2012).

[31] V. Torra, 'The weighted OWA operator', *Int. J. of Intelligent Systems*, **12**, 153–166, (1997).

[32] Z.S. Xu and Q.L. Dai, 'An overview of operators for aggregating information', *Int. J. of Intelligent Systems*, **18**, 953–969, (2003).

[33] R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer, Norwell, MA, 1997.

[34] R. R. Yager, 'On ordered weighted averaging aggregation operators in multicriteria decision making', *IEEE Trans. Systems, Man & Cybern.*, **18**, 183–190, (1988).

[35] R. R. Yager, 'Families of OWA operators', *Fuzzy Sets and Systems*, **55**, 255–271, (1993).

[36] R. R. Yager, 'Quantifier guided aggregation using OWA operators', *Int. J. of Intelligent Systems*, **11**, 49–73, (1996).