

## IDENTIFICAÇÃO DOS PARÂMETROS DA INTEGRAL DE CHOQUET PARA PROBLEMAS DE CLASSIFICAÇÃO MULTICRITÉRIO

**Renata Pelissari**

Faculdade de Ciências Aplicadas, UNICAMP  
Rua Pedro Zaccaria, 1300, Limeira, SP, 13484-350, Brasil  
renatapelissari@gmail.com, renatapelissari@unicamp.br

**Alvaro José Abackerli**

EMBRAPII - Brazilian Association for Industrial Research and Innovation  
SBN, QD 1, BL I, 70040-913, Brasília, DF, Brazil  
alvaro.abackerli@embrapii.org.br

**Leonardo Tomazeli Duarte**

Faculdade de Ciências Aplicadas, UNICAMP  
Rua Pedro Zaccaria, 1300, Limeira, SP, 13484-350, Brasil  
leonardo.duarte@fca.unicamp.br

### RESUMO

A integral de Choquet é um operador de agregação amplamente utilizado para lidar com critérios dependentes em apoio à decisão multicritério. Um problema prático com relação ao seu uso é a identificação de seus parâmetros, conhecidos como capacidade de Choquet. Apresentamos nessa trabalho um novo método baseado na Análise Multicritério de Aceitabilidade Estocástica (SMAA) para identificação da capacidade de Choquet em problemas de classificação. A fim de fornecer a possibilidade de explorar o espaço de capacidades considerando as preferências iniciais do decisor, introduzimos o conceito de *cenários de categorização viáveis* e novas medidas descritivas.

**PALAVRAS CHAVE.** Problemas de classificação multicritério. Elicitação de preferências. Integral de Choquet. SMAA.

**Tópicos (ADM – Apoio à Decisão Multicritério. ESTMP – Estatística e Modelos Probabilísticos)**

### ABSTRACT

The Choquet integral is an aggregation operator widely used to deal with interacting criteria in multiple criteria decision analysis (MCDA). A practical problem regarding its use is the identification of their parameters, which are known as Choquet capacity. In the context of sorting problems, we propose a novel method based on the the Stochastic Acceptability Multicriteria Analysis (SMAA). In order to enable exploration of the space of capacities by considering the decision makers' initial preferences, we introduce the concept of *feasible categorization scenarios* and new descriptive measures.

**KEYWORDS.** Multiple criteria sorting problem. Preference elicitation. Capacity identification. Choquet integral. SMAA

**Topics (Multiple criteria Decision Analysis. Statistics and Probability)**

## 1. Introdução

Dado um conjunto de alternativas  $A = \{a_1, a_2, \dots, a_m\}$ , o objetivo de um problema de classificação multicritério, interesse desse artigo, é atribuir uma categoria  $K_i$ , obtida de uma coleção de categorias predefinidas e ordenadas  $K_1, K_2, \dots, K_k$ , a cada alternativa de  $A$ . Esse processo de atribuição é alcançado considerando a avaliação das alternativas em relação a um conjunto de  $n$  critérios  $G = \{g_1, g_2, \dots, g_n\}$ . Nesse trabalho, mais especificamente, consideramos problemas nos quais os critérios não são independentes podendo haver interações entre eles.

Um modelo de preferência bastante utilizado em decisão multicritério capaz de lidar com esse tipo de problema é a integral de Choquet [Marichal, 2000; Grabisch e Labreuche, 2010]. Assim como a maioria dos operadores de agregação utilizados em decisão multicritério, a integral de Choquet é um modelo paramétrico e, portanto, sua aplicação requer a identificação de parâmetros conhecidos como capacidade de Choquet [Marichal, 2000; Grabisch e Labreuche, 2010].

Pode não ser fácil para o decisor definir as capacidades de Choquet de forma precisa. Por isso, diferentes abordagens vêm sendo propostas na literatura para a identificação da capacidade de Choquet a partir de informações imprecisas do decisor [Roubens, 2002; Grabisch et al., 2008; Angilella et al., 2010; Hüllermeier e Tehrani, 2013; Tehrani e Hüllermeier, 2013; Havens e Anderson, 2019; Yang et al., 2019; Duarte, 2018; Benabbou et al., 2017]. Uma dessas perspectivas considera a metodologia SMAA (*Stochastic Multicriteria Acceptability Analysis*), como apresentado por Angilella et al. [2015]. A ideia da metodologia SMAA é realizar uma análise reversa do espaço de valores de parâmetros factíveis, apresentando medidas descritivas que representem as preferências do decisor, sem necessariamente solicitá-las. Portanto, para possibilitar a análise reversa, essas medidas descritivas estão no centro da metodologia SMAA. Dependendo do tipo de problema de decisão e do modelo de agregação adotado, diferentes medidas descritivas são propostas Pelissari et al. [2020]. Apesar dos métodos de classificação baseados no SMAA existentes [Tervonen et al., 2009; Lahdelma e Salminen, 2010; Pelissari et al., 2019, 2021], incluindo uma combinação da Integral de Choquet com o SMAA [Arcidiacono et al., 2021], ainda existem algumas limitações relacionadas as medidas descritivas apresentadas nesse trabalhos:

- Nos métodos de classificação baseados no SMAA, a decisão é feita considerando o índice de aceitabilidade de categoria, que indica a probabilidade de uma alternativa ser categorizada em cada categoria. Com isso, a categorização de cada alternativa é analisada individualmente e, conseqüentemente, a categorização de diferentes alternativas pode resultar de diferentes sistemas de preferência, isto é, de diferentes capacidades de Choquet no nosso contexto.
- Enquanto a maioria dos métodos de ordenamento baseados no SMAA apresentam o vetor de pesos centrais – medida descritiva que indica os pesos dos critérios que fazem uma alternativa ocupar o primeiro lugar no ordenamento – não existe medida similar para os métodos de classificação.

Motivados por essas limitações, nosso objetivo é combinar a integral de Choquet à metodologia SMAA, propondo o método de classificação denominado SMAA-S-Choquet. O método proposto baseia-se no conceito de *cenários de categorização viáveis*, o que significa que, ao invés de analisar a probabilidade de categorização de cada alternativa individualmente, analisamos a probabilidade de cada possível categorização envolvendo atribuições para todas as alternativas. Tal procedimento requer a definição de uma nova medida descritiva que será chamada de *Índice de Aceitabilidade do Cenário* (*Scenario Acceptability index* – SAI). Além disso, propomos o vetor de *Capacidade Central do Cenário* (*Scenario Central Capacity* – SCC), que ajuda a identificar as capacidades que tornam possível um determinado cenário.

Este artigo está organizado da seguinte forma. Na Seção 2, fornecemos informações básicas sobre o uso da integral de Choquet e da metodologia SMAA. Nossa proposta, o método SMAA-S-Choquet com as novas medidas descritivas são apresentados na Seção 3. Na Seção 4, apresentamos um exemplo numérico. Conclusões, algumas limitações do método proposto e possíveis trabalhos futuros são discutidos na Seção 5.

## 2. Preliminares

Nesta seção introduzimos a Integral de Choquet e apresentamos a metodologia SMAA.

### 2.1. Integral de Choquet

A integral de Choquet é baseada no conceito de capacidade (medida fuzzy) que atribui um peso a cada subconjunto de critérios. Mais precisamente, denotando por  $2^G$  o conjunto de potência de  $G$ , a função  $\mu : 2^G \rightarrow [0, 1]$  é chamada de capacidade em  $2^G$  se as seguintes condições forem satisfeitas:

- (a) Condições de contorno:  $\mu(\emptyset) = 0$  and  $\mu(G) = 1$ ,
- (b) Condições de monotonicidade:  $\forall S \subseteq T \subseteq G, \mu(S) \leq \mu(T)$ .

Em um problema de decisão em que há interação entre critérios e sua importância é representada por uma capacidade  $\mu$ , a soma ponderada pode ser estendida através da integral de Choquet que é uma função  $CI : A \rightarrow R_0^+$  dada por

$$CI(a) = \sum_{j=1}^n [g_{(j)}(a) - g_{(j-1)}(a)] \mu(N_j), \quad (1)$$

em que  $g_{(0)}(a) \leq g_{(1)}(a) \leq \dots \leq g_{(n)}(a)$ ,  $g_{(0)}(a) = 0$ ,  $N_j = \{(j), \dots, (n)\}$ , para  $j = 1, \dots, n$ .

Com o objetivo de reduzir o número de parâmetros a serem elicitados, adotamos a representação da capacidade 2-aditiva introduzida por Grabisch [1997b]. A propriedade 2-aditiva implica que pode haver interação entre quaisquer dois critérios, mas não entre três ou mais, e permite reduzir o número de parâmetros de  $2^n - 2$  para  $n(n+1)/2$ , o que representa uma redução significativa à medida que  $n$  cresce.

Conforme apresentado em Grabisch [1997a] e Grabisch [2000], uma representação útil da integral de Choquet pode ser dada pelo índice de interação generalizado. Para o caso da capacidade 2-aditiva, o índice de interação generalizado entre dois critérios  $\{g_j, g_s\}$  é representado por  $I_{js}$  e, quando o índice de interação generalizado é calculado para um único critério, digamos  $g_j$ , dá-se origem ao índice de importância de Shapley, denotado por  $I_j$ . Ainda no contexto da capacidade 2-aditiva, as capacidades são completamente determinadas por Shapley e pelos índices de interação.

O índice de interação generalizado admite uma interpretação direta e prática. O índice de Shapley  $I_j$  quantifica a importância média de um único critério  $g_j$  para o processo de decisão, semelhante aos pesos dos critérios nos métodos de decisão multicritério tradicionais. Além disso,  $I_{ij} > 0$  modela uma sinergia entre  $g_j$  e  $g_s$ ,  $I_{js} < 0$  representa uma redundância entre  $g_j$  e  $g_s$ , e  $I_{js} = 0$  significa que os critérios  $g_j$  e  $g_s$  não interagem. As condições de contorno e monotonicidade são garantidas quando as seguintes restrições são mantidas:

$$\sum_{j=1}^n I_j = 1 \quad \text{and} \quad I_j - \frac{1}{2} \sum_{s \neq j} |I_{js}| \geq 0, \forall j. \quad (2)$$

O modelo de preferência integral de Choquet 2-aditivo com base no índice de interação generalizado é então expresso por

$$CI(a) = \sum_{I_{js} > 0} \min\{g_j(a), g_s(a)\} I_{js} + \sum_{I_{js} < 0} \max\{g_j(a), g_s(a)\} |I_{js}| + \sum_{j=1}^n g_j(a) (I_j - \frac{1}{2} \sum_{s \neq j} |I_{js}|). \quad (3)$$

Em problemas de ordenamento, a alternativa com o valor  $CI(a)$  mais alto ocupa o primeiro lugar, seguida pela alternativa com o segundo  $CI(a)$  mais alto e assim por diante. Para a problemática de classificação aqui abordada, considera-se que as categorias são definidas por limiares que representam as fronteiras entre categorias consecutivas. Seja então  $R = \{r_1, \dots, r_{k+1}\}$  o conjunto de limiares que caracterizam as  $k$  categorias, no qual  $r_1$  e  $r_{k+1}$  representam os melhores e piores limiares, respectivamente. Como as categorias são completamente ordenadas, cada limiar é preferido ao seu sucessivo, isto é,  $r_1 \succ r_2 \succ \dots \succ r_k \succ r_{k+1}$ . Então, o valor  $CI(a)$  deve ser comparado aos limiares das categorias da seguinte forma (Benabbou et al. [2017], Greco et al. [2015]):

$$a \in A \text{ é classificada na categoria } K_h, \text{ se, e somente se, } CI(a) \in [r_h, r_{h+1}[. \quad (4)$$

## 2.2. Metodologia SMAA

A metodologia SMAA (*Stochastic Multicriteria Acceptability Analysis*) foi inicialmente proposta por Lahdelma et al. [1998] para problemas de decisão multicritérios nos quais os pesos dos critérios não estão disponíveis ou não são explícitos. Desde então, muitas variantes do SMAA foram desenvolvidas para diferentes tipos de problemas, dando origem a uma família de métodos capazes de elicitar pesos de critérios e modelar dados incertos [Pelissari et al., 2020].

No SMAA, diferentes tipos de informações incertas são representadas por distribuições de probabilidade e o cálculo é implementado por meio de uma simulação de Monte Carlo. Inicialmente, os valores das variáveis aleatórias são amostrados das suas distribuições de probabilidade. Em seguida, um procedimento de agregação é aplicado para obter um ordenamento das alternativas ou uma classificação das alternativas em categorias pré-definidas – dependendo da problemática considerada. Por fim, o SMAA calcula medidas descritivas que dão suporte à decisão final. Dependendo da variante SMAA e da configuração do problema, diferentes medidas descritivas são calculadas, como a probabilidade de uma alternativa ser a mais preferida, a probabilidade de uma alternativa dominar outra ou a probabilidade de uma alternativa ser categorizada em uma determinada categoria (veja Pelissari et al. [2020] para um resumo das medidas descritivas dos diferentes métodos SMAA).

No caso de problemas de classificação ordinal, nosso interesse aqui, três métodos baseados no SMAA foram propostos. Tervonen et al. [2009] propôs o método SMAA–TRI, que combina o SMAA com o método ELECTRE–TRI, e perfis de referência centrais são usados na definição das categorias. Lahdelma e Salminen [2010] propôs o SMAA–OC, cujo procedimento de agregação é baseado em funções utilidade, e limiares são usados na definição das categorias. Pelissari et al. [2020] propôs o SMAA–Fuzzy–FlowSort combinando SMAA com o método Fuzzy–FlowSort. Nesse método, as categorias são definidas por limiares e dados qualitativos utilizados na matriz de decisão e parâmetros do modelo são modelados por números *fuzzy*. Recentemente, Pelissari et al. [2021] estenderam o SMAA–Fuzzy–FlowSort para cenários com critérios hierárquicos, propondo o SMAA–FFS–H.

A única medida descritiva fornecida por todos esses métodos de classificação baseados no SMAA é o índice de aceitabilidade da categoria que indica a probabilidade da alternativa  $a_i$  ser

classificada na categoria  $K_h$ . A estimativa do índice de aceitabilidade da categoria por simulação é dada por

$$K_i^h = \text{hit}_i^h / N \quad (5)$$

em que  $\text{hit}_i^h$  é número de vezes que a categoria  $K_h$  foi atribuída à alternativa  $a_i$  e  $N$  é o número de iterações considerado no processo de simulação.

### 3. O método SMAA-S-Choquet e novas medidas descritivas

Nesta seção apresentamos o método de classificação baseado no SMAA e na integral de Choquet proposto e introduzimos as novas medidas descritivas.

#### 3.1. Novas medidas descritivas

A maioria dos métodos de ordenamento SMAA calcula, entre outras medidas descritivas, o vetor de peso central. Este índice descreve os pesos dos critérios que fazem uma alternativa ocupar o primeiro lugar no ordenamento. Ao apresentar o vetor de peso central ao decisor, a análise reversa pode ser conduzida: em vez de definir os pesos dos critérios e construir uma solução para o problema, o decisor pode aprender quais pesos levam uma alternativa a ser a preferida. Não existe nenhuma medida semelhante para métodos de classificação baseados no SMAA.

De fato, a única medida descritiva fornecida pelos métodos de classificação SMAA é o índice de aceitabilidade da categoria, que informa a probabilidade de uma alternativa ser atribuída a uma categoria e é estimado conforme definido na Eq. 5. Com base neste índice, uma decisão final pode ser alcançada pelo decisor: geralmente ele decide alocar uma alternativa na categoria com o maior índice de aceitabilidade, como discutido em Tervonen et al. [2009], Lahdelma e Salminen [2010] e Pelissari et al. [2019]. Por exemplo, vamos considerar um problema de tomada de decisão com duas alternativas e duas categorias em que os índices de aceitabilidade obtidos são:  $K_1^1 = 0,9$  e  $K_1^2 = 0,1$ , e  $K_2^1 = 0,38$  e  $K_2^2 = 0,62$ . Nesse caso, é natural que o decisor conclua que as categorias  $K_1$  e  $K_2$  sejam atribuídas às alternativas  $a_1$  e  $a_2$ , respectivamente. Também é razoável supor que o decisor deseja classificar o conjunto de todas as alternativas com base no mesmo vetor de pesos, ou seja, se a alternativa  $a_1$  é alocada na categoria  $K_1$  considerando que o critério 1 é mais importante que o critério 2, espera-se que a categorização de  $a_2$  também seja baseada no fato de que o critério 1 é mais importante do que o critério 2. No entanto, não há garantia de que isso ocorra ao se usar o índice de aceitabilidade da categoria. Portanto, é possível que a classificação final de duas ou mais alternativas resulte de diferentes sistemas de preferências. Além disso, ao utilizar apenas o índice de aceitabilidade da categoria para tomar uma decisão final, o decisor não possui nenhuma informação complementar que indique o sistema de preferência que resultou na respectiva decisão.

Para contornar esta limitação, apresentamos o conceito de *cenário de categorização viável* (ou simplesmente “cenário de categorização”) como a junção das categorizações de todas as alternativas obtidas em cada iteração da simulação. Assim, ao invés de analisar a classificação individual de cada alternativa, analisamos a combinação das classificações de todas as alternativas, de modo que cada combinação, referida como “cenário de categorização”, seja baseada em um único sistema de preferência. Mais especificamente, o “cenário de categorização” é um vetor de tamanho  $m$  cujos elementos representam a categorização das alternativas. Por exemplo, em um problema com três alternativas  $a_1$ ,  $a_2$  e  $a_3$  alocadas nas categorias  $K_2$ ,  $K_2$  e  $K_3$ , respectivamente, um possível cenário de categorização é dado pelo vetor  $(K_2, K_2, K_3)$ . Diferentes “cenários de categorização” podem ser obtidos em diferentes iterações da simulação.

Em termos matemáticos, definimos o cenário de categorização  $\mathbf{v}_t$  por:

$$\mathbf{v}_t = \{(K^t(a_1), K^t(a_2), \dots, K^t(a_m))\},$$

em que  $K^t(a_i) = K_h$  é a categoria na qual a alternativa  $a_i$  é alocada. Para simplificação e para melhor visualização, cada elemento desse vetor também pode ser denotado simplesmente pelo índice que representa a categoria,  $K^t(a_i) = h$ .

Nossa segunda proposta é levar em consideração a probabilidade de ocorrência de cada cenário viável para decidir pelas classificações finais. Para tanto, introduzimos uma nova medida descritiva, o “Índice de Aceitabilidade do Cenário” (*Scenario Acceptability Index* – SAI). Enquanto o índice de aceitabilidade da categoria representa a probabilidade de uma alternativa ser alocada em uma categoria, o SAI representa a probabilidade de ocorrência de um cenário. Portanto, as estimativas para o SAI são calculadas pela seguinte expressão

$$SAI_t = \frac{B_t}{N}. \quad (6)$$

em que  $N$  é o número total de iterações e  $B_t$  é o número de vezes que o cenário  $\mathbf{v}_t$  ocorre.

O SAI de um cenário impossível é igual a 0, o SAI de qualquer cenário viável está entre 0 e 1 e a soma de todos os SAI é igual a 1. Quanto maior o valor  $SAI_t$ , maior a probabilidade de ocorrência do cenário  $\mathbf{v}_t$ . Portanto, sugere-se que o SAI seja utilizado na priorização dos cenários e, conseqüentemente, na decisão de em quais categorias as alternativas podem ser alocadas.

O vetor de capacidades de Choquet que torna um cenário possível também é uma informação relevante a ser usada pelo decisor para priorizar cenários. Para tanto, propomos outra medida descritiva, o vetor de Capacidade Central do Cenário (*Scenario Central Capacity* – SCC), que se baseia no vetor de capacidade central introduzido por Angilella e Mazzù [2015]. Enquanto o último representa a média das preferências que tornam uma alternativa preferida em um problema de ordenação, o vetor de Capacidade Central do Cenário, denotado por  $\mathbf{I}_t^c$ , representa as preferências médias do decisor (índices de Shapley e de interação) que levam a ocorrência do cenário  $\mathbf{v}_t$ . O SCC é considerado como o centróide esperado do espaço de capacidades favoráveis e é definido por

$$\mathbf{I}_t^c = \frac{\sum_{x=1}^{B_t} \mathbf{I}_t^x}{B_t}, \quad (7)$$

em que  $\mathbf{I}_t^x$  representa as preferências do decisor que fazem o cenário  $\mathbf{v}_t$  ocorrer. O vetor  $\mathbf{I}_t^c$  é indexado considerando uma ordem lexicográfica dos índices de Shapley e de interação, assim como o vetor  $\mathbf{I}_t^x$ . É importante notar que os elementos do vetor  $\mathbf{I}_t^c$  também são capacidades, como demonstrado na Proposição 3.1.

**Proposição 3.1.** *O centróide  $\mathbf{I}_t^c$  é uma capacidade. Portanto, em termos dos índices de Shapley e de interação, os elementos de  $\mathbf{I}_t^c$  seguem as condições 2.*

Demonstração: sem perda de generalidade, vamos supor que  $B_t = 2$ ,  $\mathbf{I}_t^1 = \{I_1^{(1)}, I_2^{(1)}, \dots, I_n^{(1)}, I_{12}^{(1)}\}$  e  $\mathbf{I}_t^2 = \{I_1^{(2)}, I_2^{(2)}, \dots, I_n^{(2)}, I_{12}^{(2)}\}$ . Por definição,  $\mathbf{I}_t^1$  e  $\mathbf{I}_t^2$  seguem as condições 2. Assim,

$$I_1^{(1)} + I_2^{(1)} + \dots + I_n^{(1)} = 1, \quad (8)$$

$$I_1^{(2)} + I_2^{(2)} + \dots + I_n^{(2)} = 1, \quad (9)$$

$$I_1^{(1)} - \frac{1}{2}|I_{12}^{(1)}| \geq 0, \quad (10)$$

$$I_2^{(1)} - \frac{1}{2}|I_{12}^{(1)}| \geq 0, \quad (11)$$

$$I_1^{(2)} - \frac{1}{2}|I_{12}^{(2)}| \geq 0, \quad (12)$$

$$I_2^{(2)} - \frac{1}{2}|I_{12}^{(2)}| \geq 0. \quad (13)$$

Pela equação 7, os elementos do centróide  $I_t^c$  são dados por

$$I_t^c = \left( \frac{I_1^{(1)} + I_1^{(2)}}{2}, \frac{I_2^{(1)} + I_2^{(2)}}{2}, \dots, \frac{I_n^{(1)} + I_n^{(2)}}{2}, \frac{I_{12}^{(1)} + I_{12}^{(2)}}{2} \right). \quad (14)$$

Primeiro demonstramos que os elementos de  $I_t^c$  atendem as condições de contorno. Realmente, considerando equações 8 e 9, temos que

$$\begin{aligned} \frac{I_1^{(1)} + I_1^{(2)}}{2} + \frac{I_2^{(1)} + I_2^{(2)}}{2} + \dots + \frac{I_n^{(1)} + I_n^{(2)}}{2} = \\ \frac{I_1^{(1)} + I_2^{(1)} + \dots + I_n^{(1)}}{2} + \frac{I_1^{(2)} + I_2^{(2)} + \dots + I_n^{(2)}}{2} = \frac{1}{2} + \frac{1}{2} = 1. \end{aligned}$$

Agora, vamos mostrar que os índices de interação de  $I_t^c$  atendem à condição de monotonicidade. Como  $I_{12}^{(1)}$  e  $I_{12}^{(2)}$  são sempre positivos ou negativos (ou seja, eles sempre têm o mesmo sinal), o módulo da soma é igual à soma dos módulos. Assim,  $|I_{12}^{(1)} + I_{12}^{(2)}| = I_{12}^{(1)} + I_{12}^{(2)}$ . Também considerando as condições 10 and 12, temos

$$\begin{aligned} \frac{I_1^{(1)} + I_1^{(2)}}{2} - \frac{1}{2} \left| \frac{I_{12}^{(1)} + I_{12}^{(2)}}{2} \right| &= \frac{I_1^{(1)} + I_1^{(2)}}{2} - \frac{1}{2} \left( \frac{|I_{12}^{(1)}|}{2} + \frac{|I_{12}^{(2)}|}{2} \right) = \\ \frac{1}{2} \left( I_1^{(1)} - \frac{1}{2} |I_{12}^{(1)}| \right) + \frac{1}{2} \left( I_1^{(2)} - \frac{1}{2} |I_{12}^{(2)}| \right) &\geq 0. \end{aligned}$$

Da mesma forma, com base nas condições 11 e 13, demonstramos que

$$\frac{I_2^{(1)} + I_2^{(2)}}{2} - \frac{1}{2} \left| \frac{I_{12}^{(1)} + I_{12}^{(2)}}{2} \right| \geq 0.$$

□

### 3.2. Aplicando o SMAA a integral de Choquet para classificação

Nesta seção, apresentamos o método SMAA-S-Choquet, que é baseado no modelo de preferência integral Choquet 2-aditivo e expresso em termos dos índices de Shapley e de interação.

O método SMAA-S-Choquet engloba quatro etapas: (1) definição dos dados de entrada, (2) processo de simulação, (3) cálculo das medidas descritivas e (4) etapa de exploração. Antes de discutir essas etapas em detalhes, devemos, no entanto, apresentar a Etapa 0, anterior à definição dos dados de entrada, muito semelhante em todos os problemas de classificação multicritério, nos quais o problema de decisão é estruturado. A etapa de estruturação do problema compreende: (0.1) estabelecimento do objetivo do problema decisório; (0.2) identificação dos atores envolvidos no processo de decisão (como o decisor e o analista) e definição de suas responsabilidades; (0.3) listagem das alternativas de decisão; (0.4) listagem dos critérios de avaliação; e (0.5) definição das categorias às quais as alternativas serão atribuídas, incluindo o número de categorias e seus significados.

Uma vez que o problema de decisão está configurado, entramos na Fase 1, em que os seguintes dados de entrada devem ser definidos: matriz de decisão, conjunto de limiares das categorias e algumas informações de preferência sobre as capacidades de Choquet (parametrizadas no

domínio do índice de interação generalizado). Em termos matemáticos, cada elemento da matriz de decisão denotado por  $g_j(a_i)$  representa a avaliação da alternativa  $a_i$  em relação ao critério  $g_j$ . As avaliações das alternativas podem ser definidas por dados intervalares ou estocásticos e, portanto, assumimos avaliações das alternativas como variáveis aleatórias, denotadas por  $\xi_{ij}$ , com função de densidade de probabilidade  $f_X(\xi)$  no espaço

$$X = \{\xi_{ij} \in R^m \times R^n : \xi_{ij}, i = 1, \dots, m, j = 1, \dots, n\}.$$

Da mesma forma, os limiares das categorias podem ser definidos por dados intervalares ou estocásticos. Nesses casos, os limiares das categorias são representados por variáveis aleatórias  $\psi$  com função densidade de probabilidade  $f_Y(\psi)$  no espaço  $Y$ . Os limiares são definidos de forma a serem distribuídos independentemente e, portanto, as distribuições  $f_Y(\psi)$  não devem se sobrepor. Consequentemente, o espaço  $Y$  é definido por:

$$Y = \{\psi \subseteq R^{(k+1)} \times R^n : \psi_{hj} - \psi_{lj} < 0, \forall h < l, h = 1, \dots, k, j = 1, \dots, n\}.$$

Com relação às capacidades de Choquet, embora os valores exatos das capacidades não sejam exigidos do decisor, algumas informações de preferência ainda são. Devemos supor que o decisor é capaz de fornecer alguma informação de preferência indireta em termos do índice de interação, indicando quais pares de critérios são sinérgicos ou redundantes, e em termos do índice de Shapley, indicando a importância relativa dos critérios. Essas informações de preferência resultam nas seguintes restrições:

- Restrições relacionadas às informações de preferência do decisor em relação ao índice de interação:

$$E^{DMP-I} \{ I_{js} > 0, \text{ se os critérios } j \text{ e } s \text{ são sinérgicos}, I_{js} < 0, \text{ se os critérios } j \text{ e } s \text{ são redundantes.}$$

- Restrições relacionadas às informações de preferência do decisor em relação ao índice de Shapley:

$$E^{DMP-S} \{ I_j \geq I_s, \text{ se } j \succeq s, I_j > I_s, \text{ se } j \succ s, I_j = I_s, \text{ se } j \sim s.$$

- Além das restrições relacionadas às preferências do decisor, devem ser consideradas as restrições que garantam a monotonicidade e as condições de contorno introduzidas na Eq. 2.

Chamaremos de modelo compatível uma capacidade cuja representação do índice de interação generalizado satisfaça o conjunto de restrições  $E = E^{DMP-I} \cup E^{DMP-S} \cup E^{MB}$ . A falta total de informações de preferência só é possível em relação ao índice Shapley. Nesse caso,  $E$  se reduz a  $E = E^{DMP-I} \cup E^{MB}$ . No SMAA, o conjunto de restrições é usado para restringir o espaço de capacidades viáveis e, no processo de simulação (que discutiremos a seguir), as capacidades são geradas aleatoriamente dentro do espaço  $E$ .

A Etapa 2 do SMAA-S-Choquet, referente ao processo de simulação, começa com a amostragem das avaliações dos critérios e dos limiares de suas respectivas distribuições de probabilidade. As avaliações dos critérios amostrados devem estar entre o pior e o melhor limiares amostrados. Para elicitação da capacidade, aplicamos um processo de simulação baseado no método Hit-And-Run (HAR) [Tervonen et al., 2013]. Utilizando os dados amostrados e as capacidades de Choquet elicítadas, a integral de Choquet é aplicada com base na Eq. 4. O resultado obtido em cada iteração é a alocação das alternativas nas categorias pré-definidas.

As seguintes estatísticas devem ser coletadas no processo de simulação:

- $B_t$  : o número de vezes que o cenário  $\mathbf{v}_t$  ocorre.
- $\mathbf{I}_t^x$  : Índices de Shapley e de interação associados ao cenário  $\mathbf{v}_t$ , para cada vez  $x = 1, \dots, B_t$  que o cenário  $\mathbf{v}_t$  ocorre. Este vetor é indexado considerando uma ordem lexicográfica dos índices de Shapley e de interação.

O processo de simulação deve ser repetido um número suficiente de vezes para extrair informação estatística robusta – em trabalhos anteriores, 10.000 iterações são propostas [Tervonen e Lahdelma, 2007]. Após a última iteração, medidas descritivas são calculadas (Etapa 3). As seguintes medidas descritivas são calculadas no método proposto:

- Índice de aceitabilidade da categoria, aplicando a Eq. 5;
- Índice de aceitabilidade do cenário (SAI), aplicando a Eq. 6;
- Vetor de capacidade central do cenário (SCC), aplicando a Eq. 7.

A partir das duas medidas descritivas aqui propostas, SAI e vetor SCC, o decisor consegue fazer sua categorização final (Etapa 4). Cenários com valores de SAI mais altos são mais prováveis de ocorrer e mais robustos, ou seja, menos sensíveis a mudanças nos valores de capacidade. Portanto, sugere-se escolher entre os cenários com os valores SAI mais elevados, que chamamos de cenários prioritários. Por sua vez, o vetor SCC permite uma análise reversa das capacidades no sentido de que pode ser utilizado para distinguir cenários. Portanto, o decisor deve analisar os vetores SCC para escolher dentre um dos cenários prioritários.

Vale ressaltar que o índice de aceitabilidade da categoria ainda pode ser útil, especialmente para questões de análise de robustez (como discutido em Tervonen et al. [2009]). Portanto, pode ser analisado em conjunto com o SAI e o vetor SCC.

#### 4. Exemplo numérico: uma aplicação do método a dados sintéticos

Nesta seção, o método proposto é aplicado a um exemplo sintético inspirado nos estudos conduzidos por Grabisch e Labreuche [2005] e Angilella e Mazzù [2015].

O reitor de uma escola técnica deseja categorizar cinco alunos em 3 categorias, alunos aceitos ( $K_1$ ), em lista de espera ( $K_2$ ) e rejeitados ( $K_3$ ), com base em suas notas em Matemática e Estatística. Um aluno bom em Matemática e em Estatística é, claro, bem apreciado, mas como os alunos bons em Matemática em geral também são bons em Estatística, o reitor não quer supervalorizar os alunos que têm boas notas em ambas. Portanto, podemos dizer que existe uma interação negativa (redundância) entre Matemática e Estatística.

Para aplicar o SMAA-S-Choquet, informações de preferência indireta em termos do índice de interação são necessárias. Para representar a redundância entre Matemática e Estatística, o reitor definiu  $I_{12} < 0$ . Por outro lado, definir informações de preferência com relação ao índice Shapely não é obrigatório, mas essas informações podem ser incluídas no modelo se o decisor desejar, e o reitor define que a importância de ambas as disciplinas é a mesma ( $I_1 = I_2$ ).

Os alunos foram avaliados com base em 10 testes aplicados em cada disciplina, conforme apresentado na Tabela 1 (as notas em Estatística foram omitidas por questão de espaço). Assumimos aqui que a avaliação de cada aluno em relação a cada disciplina é uma variável aleatória seguindo distribuições de probabilidade estimada da amostra formada pelas notas obtidas nos 10 testes.

Aplicando o teste de normalidade Shapiro-Walk nos valores apresentados na Tabela 1, verificamos que as amostras de todas as notas dos alunos são compatíveis com uma distribuição normal, uma vez que os valores- $P$  são maiores que 0,05 (ver Tabela 2 para notas em Matemática).

Os parâmetros da distribuição normal, média e desvio padrão, são estimados também a partir da amostra formada pelos 10 testes. Os parâmetros de cada distribuição normal também são apresentados na Tabela 1.

Tabela 1: As notas dos alunos em Matemática e Estatística.

| Estudantes | Matemática - 10 tests |       |       |       |       |       |       |       |       |          | Estatística |     |       |     |
|------------|-----------------------|-------|-------|-------|-------|-------|-------|-------|-------|----------|-------------|-----|-------|-----|
|            | $T_1$                 | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ | $T_7$ | $T_8$ | $T_9$ | $T_{10}$ | Média       | dp. | Média | dp. |
| $a_1$      | 6.7                   | 8     | 6.5   | 6.6   | 3.9   | 4.3   | 1.2   | 3.8   | 5.5   | 5.4      | 5.2         | 1.9 | 4.2   | 0.8 |
| $a_2$      | 8.3                   | 9.1   | 9.1   | 8.0   | 8.3   | 9.3   | 9.3   | 9.4   | 9.1   | 9.8      | 8.9         | 0.6 | 9.3   | 0.2 |
| $a_3$      | 3.1                   | 4.5   | 4.0   | 4.4   | 4.5   | 5.2   | 3.8   | 4.0   | 3.7   | 4.3      | 4.1         | 0.6 | 6.3   | 0.7 |
| $a_4$      | 6.9                   | 8.5   | 6.1   | 9.5   | 9.5   | 8.5   | 8.3   | 9.6   | 8.4   | 7.7      | 8.3         | 1.4 | 6.9   | 0.6 |
| $a_5$      | 9.1                   | 6.8   | 6.3   | 9.8   | 8.2   | 9.8   | 8.4   | 8.2   | 8.3   | 7.9      | 8.3         | 1.1 | 4.7   | 0.8 |

Tabela 2: Teste de normalidade Shapiro-Wilk aplicado as amostras das notas dos alunos.

|         | $a_1$   | $a_2$   | $a_3$  | $a_4$   | $a_5$   |
|---------|---------|---------|--------|---------|---------|
| W       | 0.94914 | 0.89507 | 0.9706 | 0.90892 | 0.92977 |
| valor-P | 0.6584  | 0.1932  | 0.8963 | 0.2737  | 0.4456  |

Ao definir limiares das categorias, o reitor demonstrou imprecisão em relação a alguns deles. Representaremos essa imprecisão por dados intervalares modelados por uma distribuição uniforme. O reitor também assume que os limiares das categorias são os mesmos para ambas as disciplinas  $R = \{10, [7, 5; 8], [4, 5; 5], 0\}$ . A aplicação do método SMAA-S-Choquet a esse problema resultou em 40 cenários de categorização. A Tabela 3 exhibe os cenários de prioridade e seus respectivos índices de aceitabilidade ( $SAI_t$ ) e vetores de capacidade central  $I_t^c$ .

Tabela 3: Medidas descritivas do SMAA-S-Choquet para os cinco cenários prioritários.

| Cenários                | $SAI_t$ (%) | Vetor SCC |       |          | Índice de aceitabilidade da categoria |       |       |                     |
|-------------------------|-------------|-----------|-------|----------|---------------------------------------|-------|-------|---------------------|
|                         |             | $I_1$     | $I_2$ | $I_{12}$ | $K_1$                                 | $K_2$ | $K_3$ | Categorização Final |
| $v_1 = (2, 1, 2, 1, 2)$ | 18          | 0.5       | 0.5   | -0.46    | 3                                     | 55    | 42    | $K_2$               |
| $v_2 = (2, 1, 2, 2, 2)$ | 16          | 0.5       | 0.5   | -0.39    | 100                                   | 0     | 0     | $K_1$               |
| $v_3 = (3, 1, 2, 2, 2)$ | 13          | 0.5       | 0.5   | -0.34    | 0                                     | 94    | 6     | $K_2$               |
| $v_4 = (3, 1, 2, 1, 2)$ | 13          | 0.5       | 0.5   | -0.42    | 54                                    | 46    | 0     | $K_1$               |
| $v_5 = (2, 1, 2, 1, 1)$ | 11          | 0.5       | 0.5   | -0.62    | 34                                    | 66    | 0     | $K_2$               |

Analisando a Tabela 3, é possível escolher tanto o cenário  $v_1$  quanto o  $v_2$  já que ambos apresentam os maiores valores de SAI ( $SAI_1 = 18\%$  e  $SAI_2 = 16\%$ ). Para decidir entre esses cenários dois, é importante identificar o cenário cujas capacidades melhor representam as preferências do decisor. Como o reitor definiu que  $I_1 = I_2$  e obviamente o processo HAR resultou em  $I_1 = I_2 = 0.5$  para todos os cenários, a análise para decidir entre os cenários  $v_1$  e  $v_2$  se restringe a análise do índice  $I_{12}$ . Pode-se notar que quando  $|I_{12}|$  aumenta, existe uma tendência dos estudantes serem categorizados em melhores categorias. Realmente, quanto maior o valor absoluto de  $I_{12}$ , mais redundantes são consideradas as duas disciplinas. Assim, pode-se assumir que seria mais natural ao reitor escolher o cenário  $v_2$  pois, embora considere o fator de redundância no problema, é importante que os alunos ainda tenham notas minimamente boas em ambas as disciplinas (o que é representado por um índice de interação comparativamente menor).

Na Tabela 3 apresentamos o índice de aceitabilidade da categoria. Se a decisão fosse feita considerando apenas esse índice, cada estudante seria alocado na categoria com o maior índice –

como é normalmente recomendado nos métodos de classificação SMAA [Tervonen et al., 2009; Lahdelma e Salminen, 2010; Pelissari et al., 2019]). Nesse caso, a categorização final seria a mesma sugerida pelo cenário  $v_1$ , que não é o cenário escolhido pelo reitor quando o SAI e o vetor SCC são utilizados, como discutido anteriormente. Voltando a analisar o índice de aceitabilidade da categoria, vemos que com 54% dos valores de parâmetro viáveis, o aluno  $a_4$  é alocado na categoria  $K_1$  (como no cenário 1), e uma parte bastante grande dos valores viáveis (46%) aloca o aluno na categoria  $K_2$  (como no cenário 2). Portanto, alocar o aluno  $a_4$  na categoria  $K_2$  como sugerido pelo cenário 2 faz sentido quando analisamos o índice de aceitabilidade da categoria.

Portanto, fica claro que o índice de aceitabilidade da categoria não é suficiente para embasar a decisão final, uma vez que também são necessárias informações sobre as capacidades.

## 5. Conclusão

Neste artigo, um novo método de identificação da capacidade de Choquet foi proposto para problemas de classificação multicritério, integrando a metodologia SMAA à formulação de classificação da integral de Choquet. A originalidade do método proposto baseia-se principalmente na introdução do conceito de cenários de categorização e na proposição de duas novas medidas descritivas, o SAI e o vetor SCC. Uma perspectiva de estudo futuro inclui a extensão do conceito de cenário de categorização e das duas medidas descritivas propostas para outros métodos de classificação SMAA. Dessa forma, a análise reversa do espaço de pesos dos critérios viáveis, uma das principais características do SMAA, seria possível em métodos de classificação SMAA como SMAA-OC, SMAA-Tri e SMAA-Fuzzy-FlowSort.

## Referências

- Angilella, S., Greco, S., e Matarazzo, B. (2010). Non-additive robust ordinal regression: A multiple criteria decision model based on the choquet integral. *European Journal of Operational Research*, 201(1):277–288.
- Angilella, S. e Mazzù, S. (2015). The financing of innovative SMEs: A multicriteria credit rating model. *European Journal of Operational Research*, 244(2):540–554.
- Angilella, S., Corrente, S., e Greco, S. (2015). Stochastic multiobjective acceptability analysis for the Choquet integral preference model and the scale construction problem. *European Journal of Operational Research*, 240(1):172–182.
- Arcidiacono, S. G., Corrente, S., e Greco, S. (2021). Robust stochastic sorting with interacting criteria hierarchically structured. *European Journal of Operational Research*, 292(2):735–754.
- Benabbou, N., Perny, P., e Viappiani, P. (2017). Incremental elicitation of choquet capacities for multicriteria choice, ranking and sorting problems. *Artificial Intelligence*, 246:152 – 180.
- Duarte, L. (2018). A novel multicriteria decision aiding method based on unsupervised aggregation via the choquet integral. *IEEE Transactions on Engineering Management*, 65(2):293–302.
- Grabisch, M. (2000). A graphical interpretation of the choquet integral. *IEEE Transactions on Fuzzy Systems*, 8(5):627–631.
- Grabisch, M., Kojadinovic, I., e Meyer, P. (2008). A review of methods for capacity identification in choquet integral based multi-attribute utility theory: Applications of the kappalab r package. *European Journal of Operational Research*, 186(2):766–785.
- Grabisch, M. e Labreuche, C. (2005). *Fuzzy measures and integrals in MCDA*, p. 563–604. Springer, Berlin.

- Grabisch, M. e Labreuche, C. (2010). A decade of application of the choquet and sugeno integrals in multi-criteria decision aid. *Annals of Operational Research*, 175(1):247–286.
- Grabisch, M. (1997a). Alternative representations of discrete fuzzy measures for decision making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 05(05):587–607.
- Grabisch, M. (1997b).  $k$ -order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92(2):167–189.
- Greco, S., Mousseau, V., e Słowiński, R. (2015). Multiple criteria sorting with a set of additive value functions. *European Journal of Operational Research*, 207(3):1455–1470.
- Havens, T. e Anderson, D. (2019). Machine learning of choquet integral regression with respect to a bounded capacity (or non-monotonic fuzzy measure). volume 2019-June.
- Hüllermeier, E. e Tehrani, A. (2013). Efficient learning of classifiers based on the 2-additive choquet integral. *Studies in Computational Intelligence*, 445:17–29.
- Lahdelma, R. e Salminen, P. (2010). A method for ordinal classification in multicriteria decision making. *International Conference on Artificial Intelligence and Applications*, p. 420–425.
- Lahdelma, R., Hokkanen, J., e Salminen, P. (1998). SMAA - Stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, 106(1):137–143.
- Marichal, J.-L. (2000). An axiomatic approach of the discrete choquet integral as a tool to aggregate interacting criteria. *IEEE Transactions on Fuzzy Systems*, 8(6):800–807.
- Pelissari, R., Oliveira, M. C., Amor, S. B., Kandakoglu, A., e Hellenso, A. L. (2020). SMAA methods and their applications: a literature review and future research directions. *Annals of Operations Research*, 293: 433–493.
- Pelissari, R., José Abackerli, A., Ben Amor, S., Célia Oliveira, M., e Infante, K. M. (2021). Multiple criteria hierarchy process for sorting problems under uncertainty applied to the evaluation of the operational maturity of research institutions. *Omega*, 103:102381. ISSN 0305-0483.
- Pelissari, R., Oliveira, M. C., Amor, S. B., e Abackerli, A. J. (2019). A new flowsort-based method to deal with information imperfections in sorting decision-making problems. *European Journal of Operational Research*, 276(1):235 – 246.
- Roubens, M. (2002). *Ordinal Multiattribute Sorting and Ordering in the Presence of Interacting Points of View*, p. 229–246. Springer US, Boston, MA.
- Tehrani, A. e Hüllermeier, E. (2013). Ordinal choquistic regression. volume 32, p. 802–809.
- Tervonen, T., Valkenhoef, G., Baştürk, N., e Postmus, D. (2013). Hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research*, 224(3):552–559.
- Tervonen, T., Figueira, J. R. J., Lahdelma, R., Dias, J. A. J., e Salminen, P. (2009). A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research*, 192(1):236–242.
- Tervonen, T. e Lahdelma, R. (2007). Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research*, 178(2):500–513.
- Yang, R., Wang, Y., Hui, B., e Qiu, L. (2019). Data-driven hierarchical classifiers based on choquet integral. *Journal of Intelligent and Fuzzy Systems*, 37(1):1247–1258.