SHAP-Tree-MCDA: Explanation of Hierarchical Multi-Criteria Decision Aiding models, based on the Shapley value

Christophe Labreuche[1]

[1] Thales Research & Technology, Palaiseau, France
christophe.labreuche@thalesgroup.com

1 Purpose

One of the major challenges of Artificial Intelligence (AI) is to explain its predictions and make it transparent for the user. The explanations can take very different forms depending on the area. We are interested explaining a multi-criteria decision model. Depending on the operational use-case, several explanation approaches can be envisaged. A first situation occurs when the decision maker (DM) can spend some time to study the stakes and consequences of each alternative on the points of view, analyze their pros and cons, and come up with a recommendation that he shall defend to several stakeholders. In this case, the explanation shall be complete and takes the form of a kind of proof. There exists other situations in which the DM has to make a quick decision and is under stress and time pressure. The DM is not seeking for complete explanations, as it is the case in the following example taken from [2].

Example 1 (Example 1 in [2]). The DM is a Tactical Operator of an aircraft aiming at Maritime Patrol. It consists in monitoring a maritime area and in particular looking for illegal activity. The DM is helped by an automated system that evaluates in real time a Priority Level (PL) associated to each ship in this area. The higher the PL the more suspicious a ship and the more urgent it is to intercept it. The PL is used to raise the attention of the DM on some specific ships. The computation of the PL depends on several criteria: 1. Incoherence between Automatic Identification System (AIS) data and radar detection; 2. Suspicion of drug smuggling on the ship; 3. Suspicion of human smuggling on the ship; 4. Current speed (since fast boats are often used to avoid being easily intercepted); 5. Maximum speed since the first detection of the ship (it represents the urgency for the potential interception); 6. Proximity of the ship to the shore (since smuggling ships often aim at reaching the shore as fast as possible). ■

In the previous example, as in most real-applications, the criteria are not considered in a flat way but are organized as a tree. The criteria are indeed organized hierarchically with several nested aggregation functions. The hierarchical structure shall represent the natural decomposition of the decision reasoning into points of view and sub-points of view. In the previous example, the six criteria are organized as in Fig. 1. The tree of the DM contains four aggregation nodes: 7. Suspicion of illegal activity; 8. Kinematics; 9. Capability to escape interception; 10. Overall PL.

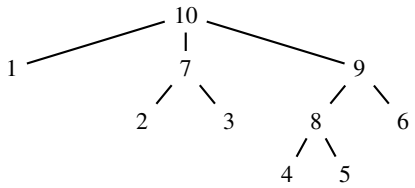

Fig. 1. Hierarchy of criteria for Ex. 2.

In [2], a new explanation approach for hierarchical MCDM models has been introduced. The current software is an implementation of this algorithm.

Section 2 proposes a short description of the preference model and explaination approach.

2 Reminders: short description of the Preference Model and explaination approach

This section proposes a summary of the preference model and the explaination approach. We refeer the reader to [2] for further developments on the approach.

2.1 MCDA Model

We are given a set of criteria $N = \{1, \ldots, n\}$, each criterion $i \in N$ being associated with an attribute X_i, either discrete or continuous. The alternatives are characterized by a value on each attribute and are thus associated to an element in $X = X_1 \times \cdots \times X_n$. We assume that the preferences of the DM over the alternatives are represented by a utility model $U : X \to \mathbb{R}$.

The hierarchy of criteria is represented by a rooted tree T, defined by the set of nodes M_T (i.e. the set of criteria and aggregation nodes), and the children $\mathrm{Ch}_T(l)$ of node l (i.e. the nodes that are aggregated at each node l) [1]. We also denote by $N_T \subseteq M_T$ the set of leaves of tree T (i.e. the criteria), by $s_T \in M_T$ the root of tree T (i.e. the top aggregation node), by $\mathrm{Desc}_T(l)$ the set of descendants of l, and by $\mathrm{Leaf}_T(l)$ the leaves at or below $l \in M_T$. A hierarchical model on criteria N is such that $N_T = N$.

The preference model is composed of an aggregation function H_l at each node $l \in M_T \setminus N_T$ and a partial utility function u_i for each criterion $i \in N_T$ (criteria). For $x \in X$, we can compute $U(x)$ recursively from a function v_i^U defined at each node $i \in M_T$:

- $v_i^U(x) = u_i(x_i)$ for every leaf $i \in N_T$,
- $v_l^U(x) = H_l\big((v_k^U(x))_{k \in \mathrm{Ch}_T(l)}\big)$ for every aggregation node $l \in M_T \setminus N_T$,
- $U(x) = v_{s_T}^U(x)$ is the overall utility.

Example 2 (Ex. 2 cont.). We have

$$v_i^U(x) = u_i(x_i) \text{ for } i \in \{1,2,3,4,5,6\}$$
$$v_7^U(x) = H_7(v_2^U(x), v_3^U(x)) \ , \ v_8^U(x) = H_8(v_4^U(x), v_5^U(x))$$
$$v_9^U(x) = H_9(v_6^U(x), v_8^U(x)) \ , \ U(x) = v_{10}^U(x) = H_{10}(v_1^U(x), v_7^U(x), v_9^U(x)) \ \blacksquare$$

2.2 Shapley Value

In Cooperative Game Theory, a *game* on N is a set function $v : 2^N \to \mathbb{R}$ such that $v(\varnothing) = 0$, N is the set of players, and $v(S)$ (for $S \subseteq N$) is the amount of wealth produced by S when they cooperate. It is a non-normalized capacity. The Shapley value is a fair share of the global wealth $v(N)$ produced by all players together, among themselves [3]:

$$\phi_i^{\mathrm{Sh}}(N, v) := \sum_{S \subseteq N \setminus i} \frac{(n - |S| - 1)!|S|!}{n!} \big[v(S \cup \{i\}) - v(S)\big]. \tag{1}$$

It can also be written as an average over the permutaiton on N:

$$\phi_i^{\mathrm{Sh}}(N, v) := \frac{1}{2^n} \sum_{\pi \in \Pi(N)} \big[v(S_\pi(i)) - v(S_\pi(i) \setminus \{i\})\big], \tag{2}$$

where $S_\pi(\pi(k)) := \{\pi(1), \ldots, \pi(k)\}$ and $\Pi(N)$ is the set of permutations on N.

2.3 Influence Index

Consider two alternatives x and y in X. One wishes to explain the reasons of the difference of preference between x and y. The explanation proposed in [2] takes the form of an index measuring the degree to which each node in M_T contributes to the difference of preference between x and y. An influence index denoted by $I_i(x, y; U, T)$ is computed for each node $i \in M_T$ for utility model U on the hierarchy T of criteria. The influence index is some kind of Shaley value applied to the game $v(S) = U(y_S, x_{N \setminus S})$ for all $S \subseteq N$, where $(y_S, x_{N \setminus S})$ denotes an alternative taking the values of y in S and the values of x in $N \setminus S$. As for the Shapley value, it is defined from permutations on N. Its expression is defined by [2]:

$$I_i(x, y, T, U) = \begin{cases} \frac{1}{|\Pi(T)|} \sum_{\pi \in \Pi(T)} \delta_\pi^{x,y,T,U}(i) \text{ if } i \in N_T \\ \sum_{k \in \mathrm{Leaf}_T(i)} I_k^{\mathrm{EOw}}(x, y, T, U) \text{ else.} \end{cases} \tag{3}$$

where $\delta_\pi^{x,y,T,U}(i) := U(y_{S_\pi(i)}, x_{N \setminus S_\pi(i)}) - U(y_{S_\pi(i) \setminus \{i\}}, x_{(N \setminus S_\pi(i)) \cup \{i\}})$. In (3), the set of admissible orderings $\Pi(T)$ is defined as the set of orderings of elements of N for which all elements of a subtree of T are consecutive. More precisely, $\pi \in \Pi(T)$ iff, for every $l \in M_T \setminus N$, indices $\pi^{-1}(\mathrm{Leaf}_T(l))$ are consecutive.

2.4 Influence Index of the Restricted Tree

The complexity of computing I_i is equal to $|\Pi(T)|$, which is far too large. It has been shown in [2] that one can reduce this complexity by taking profit of some symmetries among permutations in $\Pi(T)$. The symmetries can be seen considering subtrees of T.

We consider a subtree T' of T having the same root as T, taking a subset of nodes of T and having the same edges than T between nodes that are kept.

Definition of $U_{T'}$: Given $((u_i)_{i\in N_T}, (H_i)_{i\in M_T\setminus N_T})$ and a subtree T' of T, we can define $((u'_i)_{i\in N_{T'}}, (H'_i)_{i\in M_{T'}\setminus N_{T'}})$ by $u'_i = u_i$ for $i \in N_{T'} \cap N_T$, $u'_i(x_i) = x_i$ for $i \in N_{T'} \setminus N_T$ and $H'_i = H_i$ for $i \in M_{T'} \setminus N_{T'}$. The overall utility on the subtree is denoted by $U_{T'}$. We set $X_i = \mathbb{R}$ for every $i \in M_T \setminus N_T$. Then for $x \in X$, $U(x) = U_{T'}(x^{T'})$ where $x^{T'} \in X_{T'}$ is defined by $x_i^{T'} = x_i$ if $i \in N_{T'} \cap N_T$ and $x_i^{T'} = v_i^U(x)$ otherwise.

Definition of $T_{[j]}$: A particular subtree is when a node $j \in M_T$ of T becomes a leaf, and thus all descendants of j are encapsulated and represented by j. We define the restricted tree $T_{[j]}$ by $M_{T_{[j]}} := (M_T \setminus \mathrm{Desc}_T(j)) \cup \{j\}$, $N_{T_{[j]}} := (N_T \setminus \mathrm{Leaf}_T(j)) \cup \{j\}$, $s_{T_{[j]}} := s_T$, and $\mathrm{Ch}_{T_{[j]}}(l) = \mathrm{Ch}_T(l)$ for all $l \in M_{T_{[j]}} \setminus N_{T_{[j]}}$.

Definition of $T_{[J]}$: For $J = \{j_1, \ldots, j_p\}$, we set $T_{[J]} := \Big(\big((T)_{[j_1]}\big)_{[j_2]} \cdots\Big)_{[j_p]}$.

Let us thus consider I_i for some fixed $i \in N$. The path from s_T to i in T consists of the nodes $r_0 = s_T, r_1, \ldots, r_t = i$. Let $J = \bigcup_{l=1}^{t-1} \mathrm{Ch}_T(r_{l-1}) \setminus \{r_l\}$. Then we have [2]

$$I_i(x, y, T, U) = I_i(x^{T_{[J]}}, y^{T_{[J]}}, T_{[J]}, U_{T_{[J]}}). \tag{4}$$

The influence index can be equivalently be computed on the restricted tree $T_{[J]}$.

3 Software description

3.1 Distribution

The software package is composed of the following elements

File name	Description
HierarchicalExplanation.exe	Executable software
Explanation_DataForOneTest.xml	Input file containing the model and the alternatives to be explained
verboseTrace.txt	Output file containing the results of the explanation

The executable file is runnable under Windows. It takes as input "Explanation_DataForOneTest.xml" and generates as ouput file "verboseTrace.txt". Taking the notation of Section 2, file "Explanation_DataForOneTest.xml" contains the two options x and y, and a hierarchical description of model U (implicitly containing tree structure T), and file "verboseTrace.txt" contains the values of $I_i(x, y, T, U)$ for all leaves of the tree T.

3.2 Description of the input file "Explanation_DataForOneTest.xml"

Here is a sample of file "Explanation_DataForOneTest.xml":

Line #	Text
1	<com.thalesgroup.trt.fr.Explanation.Test.DataForOneTest id=”1”>
2	<metricDataForX id=”2”>
3	<entry>
4	<string>univ27</string>
5	<string>0.29412342856868834</string>
6	</entry>
7	<entry>
8	<string>univ24</string>
9	<string>0.1802996199839244</string>
10	</entry>
. . .	
119	</metricDataForX>
120	<metricDataForY id=”3”>
121	<entry>
122	<string>univ27</string>
123	<string>0.7144922588306435</string>
124	</entry>
. . .	
237	</metricDataForY>
238	<runtimeInNewExpFramework id=”4”>
239	<rootNode class=”com.thalesgroup.trt.fr.myriad.runtime.template.MacroCriteria.SiposChoquetAggregation” id=”5”>
240	<longName>0</longName>
241	<name>0</name>
242	<observationMinMax id=”6”>
243	<min>0.0</min>
244	<max>1.0</max>
245	<isUtilityComputableByAttributeValues>false</isUtilityComputableByAttributeValues>
246	<nodesWrongValue id=”7”/>
247	<nodesMissingValues id=”8”/>
248	<nodesCorrectValues id=”9”/>
249	<childrendObservation id=”10”/>
250	</observationMinMax>
251	<capacity class=”com.thalesgroup.trt.fr.myriad.runtime.template.MacroCriteria.TwoAdditiveMeasure” id=”11”>
252	<numCriteria>3</numCriteria>
253	<measure id=”12”>
254	<double>0.13925623631922388</double>
255	<double>0.5662866576288852</double>
256	<double>0.29445710605189096</double>
257	<double>0.0</double>
258	<double>0.27851247263844775</double>
259	<double>0.0</double>
260	</measure>
. . .	

Line #	Text
2010	</com.thalesgroup.trt.fr.myriad.runtime.template.MacroCriteria.SiposChoquetAggregation>
2011	</inputMacroCriteria>
2012	</rootNode>
2013	<model class="com.thalesgroup.trt.fr.myriad.runtime.RunTimeInstantiatorFromMC" id="505">
2014	<root class="com.thalesgroup.trt.fr.myriad.runtime.template.MacroCriteria.SiposChoquetAggregation" reference="5"/>
2015	</model>
2016	</runtimeInNewExpFramework>
2017	</com.thalesgroup.trt.fr.Explanation.Test.DataForOneTest>

Lines 2-119 describes options x. Each "entry" gives the value of x on a metric. The first field in an entry is the name of the metric, and the second field is the corresponding value. For example, x take value 0.29412342856868834 on metric "univ27". Likewise, option y is described in lines 120-237.

The rest of the file contains a description of the hierarchical multi-criteria model.

3.3 Description of the output file "verboseTrace.txt"

File "verboseTrace.txt" contains the outputs of the explanation software. The first part of the file is a summary of the hierarchical multi-criteria model.

The last part of the file takes the following form:

Results of the Influence Degrees for each node:
{39=4.085481883668726E-4, 38=4.663005546100074E-5, 37=6.439050656465541E-5, 36=2.2178465028959937E-5, 33=-2.263734107221553E-5, 32=3.0279619518436574E-5, 31=4.073095348436561E-5, 30=2.834315722161699E-4, 17=0.001887646941396652, 15=0.006944443224301233, 44=0.014316407421110771, 13=0.0165599046335034, 43=0.057750849721029104, 12=0.001207997270707862, 11=0.005663645400661312, 41=0.052803320258659144, 10=7.180825387992351E-4, 40=0.0028497935860724408, 8=-0.0013875382879824124, 7=0.006712070133478401, 28=3.594233974002705E-5, 27=2.3105040124367384E-5, 24=2.7351453679845306E-4, 2=0.22337569944877284, 1=0.027480745070370412, 22=8.18761406053539E-4, 21=9.184339146618062E-4, 20=0.0014993201907387275}
Computation Time (s) = 19.687

The list under bracket contains the values of the influence degrees for all criteria. One of these terms – e.g. "39=4.085481883668726E-4" – means that criteria with label "39" has an influence index of "4.085481883668726E-4".

4 Proprietary Licence: Software License Agreement for "SHAP-Tree-MCDA"

IMPORTANT: READ CAREFULLY BEFORE, INSTALLING, COPYING, OR OTHERWISE ACCESSING OR USING THIS SOFTWARE.

BY OPENING THE EXECUTABLE FILE, YOU (THE LICENSEE) ACKNOWLEDGE YOUR ACCEPTANCE OF THE TERMS AND CONDITIONS OF THE FOLLOWING LICENSE AGREEMENT.

IF YOU DO NOT AGREE TO ALL OF THE TERMS AND CONDITIONS STATED BELOW, PLEASE DO NOT USE THIS SOFTWARE.

4.1 Grant of License:

THALES hereby grants to Licensee during the term of the Agreement a non-transferable and non-exclusive licence, to use the Software for internal and research (non commercial)uses only. The Licensee shall not:

- directly or indirectly, modify, adapt, translate, reverse engineer, reverse assemble, decompile or disassemble the Software, in whole or in part or otherwise attempt to derive the source code for the Software in whole or in part, unless authorized by law
- distribute, rent, sublicense, lease, resell or assign the Software,
- use the Software for any other purpose than the one described above. In case Licensee wishes to use the Software for such other purpose, a specific separate agreement shall be concluded between the Licensee and THALES.

The grant of licence to the Licensee on the Software under this agreement is made free of charge.

4.2 Warranty:

The Software is provided on an “as is” basis, without warranties or conditions of any kind, including without limitation, any warranties on its non-infringement, merchantability, secured, innovative or relevant nature, fitness for a particular purpose or compatibility with any equipment or software

The Licensee shall notify Thales if the best delays in the event he identifies a bug in the Software

4.3 Responsability:

THALES SHALL NOT BE LIABLE FOR ANY SPECIAL, CONSEQUENTIAL OR INCIDENTAL DAMAGES OF ANY KIND, INCLUDING BUT NOT LIMITED TO LOST DATA OR LOST PROFITS, IN CONNECTION WITH OR ARISING FROM THE USE OF THE SOFTWARE, OR CAUSED BY A DEFECT, FAILURE OR MALFUNCTION, WHETHER A CLAIM OR SUCH DAMAGE IS BASED UPON WARRANTY, CONTRACT, NEGLIGENCE OR OTHERWISE EVEN IF THALES HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH LOSS. UNDER NO CIRCUMSTANCES SHALL THALES OR ITS LICENSORS BE LIABLE FOR AN AGGREGATED AMOUNT GREATER THAN PAYMENTS MADE TO THALES BY LICENSEE PURSUANT TO THIS LICENSE AGREEMENT FOR THE SOFTWARE THAT CAUSED THE DAMAGES.

IN NO EVENT SHALL THALES BE LIABLE FOR DAMAGES ARISING FROM A MISUSE OR UNATTENDED USE OF THE SOFTWARE.

4.4 Intellectual Property:

THALES and/or certain third parties are and shall remain the exclusive owners of all copyrights and intellectual property rights in and to the Software, and Licensee shall have no rights in the Software except as specifically set forth in this License Agreement.

4.5 Duration / Termination:

This license agreement shall remain in force during 1 year form the date of the first installation.

THALES shall have the right to terminate this License Agreement immediately in the event that Licensee breaches any term or condition hereof and fails to remedy such breach within thirty (30) days after receipt of notice of such breach. Within seven (7) days of any termination of this License Agreement, Licensee (i) shall discontinue the use of, and return to THALES, all copies of the Software and Documentation, (ii) delete such Software programs from any computer, and (iii) notify THALES in writing of such deletion.

4.6 Applicable law and disputes:

The present License Agreement is governed by and construed in accordance with the laws of France.

All disputes between the parties in connection with or arising out of the existence, validity, construction, performance and termination of this License Agreement (or any terms hereof) which the parties are unable to resolve between themselves shall be finally settled by the Tribunal de Commerce de Paris (France).

References

1. R. Diestel. *Graph Theory*. Springer-Verlag, New York, 2005.
2. C. Labreuche and S. Fossier. Explaining multi-criteria decision aiding models with an extended shapley value. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018)*, pages 331–339, Stockholm, Sweden, July 2018.
3. L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.