Expressing redundancy and complementarity of information sources with fuzzy measures

Vicenç Torra
Institut d'Investigació en Intel·ligència Artificial - CSIC
Campus UAB s/n, 08193 Bellaterra (Catalunya, Spain)
vtorra@iiia.csic.es

Abstract
In this paper we describe the suitability of hierarchically $S$-decomposable fuzzy measures for combining information. In particular, they are adequate when the set of information sources include some redundant sources and also some complementary ones. We show how these two types of sources can be combined in a single aggregation operator. Besides of that, we show how the measures allow modularity and simplicity when modelling the information sources.

Keywords: aggregation operators, fuzzy measures, Choquet integral, redundancy, complementarity

1. Introduction
At present there exist several operators defined to aggregate (or fuse) numerical information. Among the best known we have the weighted mean (characterized by Áczel in [1]), the OWA (Ordered Weighted Averaging) operator [13, 14] and the Choquet integral (see [2, 4] for reference books). When a model for an aggregation process is intended using the weighted mean or the OWA a weighting vector is required while in the Choquet integral a fuzzy measure [12] is required.

Although the use of a fuzzy measure in the Choquet integral makes it extremely flexible and useful when applied, the expressiveness of this measure is also a drawback on the construction of the model. Note that to define a fuzzy measure over a set of $n$ elements, we need $2^n$ values.

To ease in the process of model definition, several approaches have been considered in the literature. On one hand, simplified fuzzy measures have been considered (Sugeno $\lambda$-measure [9], $S$-decomposable ones, k-order models [3], hierarchically decomposable ones [11]). On the other hand, mechanisms to learn the model from examples have been developed [6].

In this work we consider hierarchically decomposable fuzzy measures (HDFM) and we show their suitability for modelling a set of information sources. We show that these measures constitute a model general enough to cope with information sources with a complex underlying structure. In particular, we are specially concerned with the need to consider in a single operator both redundant and complementary information sources. We show that the HDFM can cope with these sources, and we give an example of this case.

The structure of the paper is as follows. In Section 2, we review the hierarchically decomposable fuzzy measures and the Choquet integral. In Section 3, we review the concepts of redundant and complementary information. The next section shows how HDFM can cope with these two types of information. We introduce also two measures (maxiness and drasticness) that correspond to measures of redundancy and complementarity. In Section 5 we finish with the conclusions.

2. Preliminaries
In this section, we define some concepts that are needed after on in this work. We begin defining some basic operators and then fuzzy measures considering decomposable and hierarchical ones.

The section finishes with the definition of Choquet integral. In all these definitions we assume that $X=\{x_1, ..., x_N\}$ is the set of elements (criteria, alternatives, information sources, ...) to consider and that $f(x_i)$ corresponds to the evaluation of the element $x_i$ in $X$.

Definition 2.1. A $t$-conorm $S$ is a binary operation on the unit interval that satisfies at least the following axioms (for all $a$, $b$, $c$ and $d$ in $[0,1]$):
1) $S(a,0)=a$; (2) $b=d$ imply $S(a,b)=S(a,d)$; (3) $S(a,b)=S(b,a)$; $S(a, S(b, d)) = S(S(a, b), d)$.

Two examples of $t$-conorms are the following:
1) $S(a,b) = \max(a, b)$ (maximum $t$-conorm)
2) $S(a,b) = \begin{cases} a & \text{when } b = 0 \\ b & \text{when } a = 0 \\ 1 & \text{otherwise} \end{cases}$
(Drastic $t$-conorm)

Definition 2.2 [12]. A function $\mu: \mathcal{P}(X) \to [0,1]$ is a fuzzy measure if and only if it satisfies the
following axioms: (i) \( \mu(\emptyset) = 0 \); (ii) monotonicity: \( B_1 \subseteq B_2 \subseteq X \) implies \( \mu(B_1) = \mu(B_2) \).

If \( \mu \) is a fuzzy measure such that for all \( A, B \in \wp(X) \) with \( A \cap B = \emptyset \) we have that \( \mu(A \cup B) = S(\mu(A), \mu(B)) \) for a t-conorm \( S \) we say that \( \mu \) is decomposable. It is often considered that fuzzy measures satisfy \( \mu(X) = 1 \). Such measures are called normalized fuzzy measures. This is the case with decomposable fuzzy measures. A different kind of fuzzy measures are hierarchically S-decomposable ones. The general idea underlying a hierarchical fuzzy measure is that the elements over which we define it are organized as a dendogram with n-ary nodes. This is, they follow an structure similar to the one in Figure 1.

**Definition 2.3** [11]. \( H \) is a hierarchy of elements \( X \) if and only if the following conditions are satisfied:

(i) For all \( a \) in \( X \), \( \{a\} \in H \)

(ii) \( |\{r \mid r \notin h \} \cup h \} = 1 \)

(iii) for all \( n \) ? root, \( |\{h_i \mid n \in h_i \text{ and } h_i \in H \} \} = 1 \)

(iv) for all \( h \) in \( H \),

if \(|h| = 1\),

then there exists \( a \in X \) such that \( h = \{a\} \)

(v) for all \( h \in H \) with \(|h| \geq 1\), it is satisfied that for all \( h_i \in h \), \( h_i \in H \).

These conditions corresponds to have that (i) all the elements in \( X \) belong to the hierarchy (the elements \( a \in X \) that \( \{a\} \in H \) are the leaves); (ii) there is only one element in the hierarchy that is not included in another one (this is the root); (iii) all nodes of the hierarchy (except the root) are included in another one; (iv) the only singletons we have are the leaves; (v) all the nodes in the hierarchy (except the leaves) are defined by means of nodes already existing in the hierarchy.

**Example 2.1.** Using the example in Figure 1, the hierarchy \( H \) is defined as:

\[
H = \{\{x_1\}, \{x_2\}, ..., \{x_{16}\}, \alpha = (\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}), \beta = (\{x_5\}, \{x_6\}, \{x_7\}), \gamma, \delta, \varepsilon, \zeta = (\{\alpha, \beta, \gamma\} \in \{\{\delta, \varepsilon\}\} \}
\]

Note that if we express \( \theta \) with detail we have that:

\[
\theta = \{\{\zeta, \eta\} = \{(\alpha, \beta, \gamma), (\delta, \varepsilon)\} = \{(\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}), \gamma, (\delta, \varepsilon)\} = ...
\]

To introduce hierarchically S-decomposable fuzzy measures we need to define the extension of a node \( h \) in the hierarchy (i.e., the set of elements of \( X \) that are embedded in \( h \)) and the so-called labelled hierarchy (i.e., a value for each leaf and a t-conorm for each node).

**Definition 2.4** [11]. The extension EXT of a node \( h \) in a hierarchy \( H \) is defined as:

(i) \( \text{EXT}(h) = h \)

(ii) \( \text{EXT}(h) = \cup_{h_i \in h} \text{EXT}(h_i) \)

if \(|h| = 1\)

if \(|h| > 1\)

**Definition 2.5** [11]. A labelled hierarchy \( L \) of elements \( X \) is a t-uple \( L = (H, S, m) \) where \( H \) is a hierarchy of the elements in \( X \), \( S \) is a function that maps each \( h \in H \) that is not a leaf into a t-conorm and \( m \) is a function that maps each singleton into a value in \( K \). For the sake of simplicity we express \( S(h) \) as \( S_h \).

Labelled hierarchies define fuzzy measures. A measure of this kind is defined so that for all the singletons \( \{a\} \), \( \mu(\{a\}) \) corresponds to the value defined according to the function \( m \) in the tuple \( L = (H, S, m) \) (i.e., \( \mu(\{a\}) = m(\{a\}) \)). When the measure is applied to a non singleton, it is
calculated combining the measures of the singletons using the t-conorms in S.

**Definition 2.6** [11]. Given a labelled hierarchy L=⟨H, S, m⟩ of elements X, we define its corresponding hierarchically S-decomposable fuzzy measure (HDFM for short) as \( \mu(B) = \mu_{\text{root}}(B) \) where \( \mu_A \) is defined as:

\[
\begin{align*}
\mu_A(B) &= 0 & \text{if } |B| &= 0 \\
\mu_A(B) &= m(B) & \text{if } |B| &= 1 \\
\mu_A(B) &= S_A(\mu_{a_1}(B_1), \ldots, \mu_{a_n}(B_1)) & \text{if } |B| &= 1 \\
& & \text{where } A = \{a_1, \ldots, a_n\} \text{ and } B_i = B \cap \text{EXT}(a_i) \text{ for all } a_i \in A
\end{align*}
\]

Now, we consider the Choquet integral. This is an aggregation operator that uses fuzzy measures to express the importance of sets of sources. The use of fuzzy measures is the main difference in relation to other operators as the weighted mean. Note that in this latter case we can only consider the importance of individual elements instead of sets of them.

**Definition 2.7** [4]. Given a fuzzy measure \( \mu \), the Choquet integral with respect to \( \mu \) is defined by:

\[
C_\mu(f, x) = C_\mu(f(x_1), \ldots, f(x_N)) = \sum_{i=1}^N (f(x_{s(i)}) - f(x_{s(i-1)})) \mu(A_{s(i)})
\]

where \( f(x_{s(i)}) \) indicates that the indices have been permuted so that \( 0 = f(x_{s(1)}) = \ldots = f(x_{s(N)}) = 1 \), \( A_{s(i)} = \{x_{s(i)}, \ldots, x_{s(N)}\} \) and \( f(x_{s(0)}) = 0 \).

### 3. Types of information for aggregation operators

Intelligent systems that use aggregation operators for knowledge intergation face two different kind of data [5]: redundant and complementary. Their difference is extremely rellevant as the existence of both types in a single operator can lead to biased (and sometimes meaningless) results. Before considering the main problems of considering such data we review their informal definitions:

- **Redundant information.** Different information sources supply information that correspond to (1) similar characteristics or properties or (2) (different) samples of the same characteristics of the environment. In this case, aggregation is performed to increase the reliability of the system because, as usual, it is assumed that information sources are subject to errors. In this environment, information on the reliability of the sources or of the sample is added to aggregation operators. A typical example is the weighted mean where the reliability is given by the weights.

- **Complementary information.** Information sources supply information that correspond to (1) different characteristics or properties, or (2) to the same but obtained by independent sources. Therefore each source stands for a different point of view. Aggregation is performed to consider in a single output these characteristics. In this case, if we consider the weighted mean, weights correspond to the participation of the different points of view in the final output.

As reported in [10], aggregation operators are used in artificial intelligence for two main purposes: when a system has to make a decision or when it needs a comprehensive representation of the application domain. In both cases, both types of information appear.

- **Systems for decision making.** Redundant information corresponds to the case of having dependent or correlated criteria in multi-criteria decision making, or having people with similar background in multi-person decision making. Complementary information corresponds to independent criteria or to people with different knowledge.

- **Systems for having a better representation of a certain domain.** Redundant information corresponds to have several sensors based on the same technology measuring the same characteristics of the environment (e.g., distance to a near object). Complementary information is when the data suppliers are based on different technologies (e.g., millimeter-wave radars, infrared sensors, vision cameras or a human expert).

The combination of both types of information is always problematic, and the definition of aggregation operators that allow mixed
information is an open problem. This is due to the fact that the addition of a new information source should not always bias the output. In fact, the output has to be more affected when the "complementarity" of the new information source is large. This is not the case when the "complementariness" of the new information source is small (i.e., the source is redundant in relation to the existing ones). In this latter case, what is expected is that the robustness of the system increases but, in general, the output is not radically affected by the new input.

Therefore, a good aggregation operator that uses both kind of information should give robustness when data is redundant and bias only when data is complementary.

4. Hierarchically S-decomposable fuzzy measures

The existence of the two kind of information sources described in the previous section can be modelled by hierarchically decomposable fuzzy measures. The model is built grouping in a single node information sources that are all homogeneous in relation to the redundancy or complementarity of the data they supply. This is, we have in a node only redundant information sources or complementary ones and we prevent at all the coexistence of both types of sources in the same node. Assuming this homogeneity on the nodes, the t-conorm used in a node reveals the type of the sources in that node.

- **Redundancy**: This is the case when the measure applied to the set of all sources is not much greater than the measure applied to each element alone. This is so because when extra sources are considered there is an extra gain, due to the inclusion of new points of view. In this case, extra sources bias the output as expected. Therefore, for hierarchically decomposable fuzzy measures the t-conorm of nodes defined from complementary sources should be similar to drastic t-conorm. Therefore, we call t-conorms used for complementary information drastic-type t-conorms.

To measure maxiness or drasticness of t-conorms (needed, respectively, to measure redundancy and complementarity of a node) we define a distance between two t-conorms. The distance is defined pointwise as follows:

\[
d(S_a, S_b) = \sqrt{\frac{1}{(n+1)^2} \sum_{x=0}^{n} \sum_{y=0}^{n} (S_a(\frac{x}{n}, \frac{y}{n}) - S_b(\frac{x}{n}, \frac{y}{n}))^2}
\]

This definition is parametric in relation to a precision parameter n. Note that, in fact, a similar definition could be given without such parameter by means of an integral. However, as we are only interested in computing this distance for particular examples, the definition with the integral is not needed. The definition given above gives meaningful comparisons for small values of n.

Using this distance, the maxiness and the drasticness of a t-conorm is defined as:

\[
\text{maxiness}_n(S) = d(S_{\text{max}}, S)
\]

\[
\text{drasticness}(S) = d(S_{\text{drastic}}, S)
\]

The classical example given in the literature about the evaluation of students on relation to the marks on three subjects (Mathematical, Physics, Literature) illustrates redundancy and complementarity of different criteria.

**Example 4.1.** [4] Let \(\mu\) be the measure defined on the set \(X=\{M, P, L\}\), where M, P and L stands for Mathematics, Physics and Literature, in the following way:

1) Standard conditions on the boundaries

\[\mu(\emptyset) = 0, \ \mu(\{M, P, L\}) = 1\]

(boundary conditions)

2) Relative importance of scientific versus literary subjects
3) Avoiding redundancy between scientific subjects
\[ \mu(\{M,P\}) = 0.5 < \mu(\{M\}) + \mu(\{P\}) \]
(measure on the pairs)

4) Support between literature and scientific subjects
\[ \mu(\{M,L\}) = \mu(\{P,L\}) = 0.9 > 0.45 + 0.3 \]
(measure on the pairs)

Mathematics      Physics      Literature

\[ A \]

\[ S \]

Figure 2. Hierarchy of elements corresponding to a fuzzy measure on maths, physics and literature

Although this measure is not decomposable, it is a hierarchically S-decomposable one as shown in [11]. The measure can be defined using the hierarchy in Figure 2 and two t-conorms \( S_S \) and \( S_A \) corresponding to the subset of scientific subjects (subset S in Figure 2) and to the set of all subjects (subset A in Figure 2). We define \( S_S \) and \( S_A \) as follows (note that other t-conorms are also adequate):

\[ S_S(x, y) = (x^w + y^w)^{1/w} \]
with \( w = (\ln 2)/\ln 0.5 - \ln 0.45) = 6.5788 \)
This is a t-conorm from Yager's family.
Once selecting this family of t-conorms, the solution has been obtained from the equation: \( (0.45^w + 0.45^w)^{1/w} = 0.5 \).

\[ S_A(x, y) = f^{-1}(f(x) + f(y)) \]
where \( f(x) \) is defined as:

\[ \begin{align*}
20 & \quad x \in [0, 1/2] \\
3 + 14x & \quad x \in [1/2, 3/4] \\
6 + 10x & \quad x \in [3/4, 1]
\end{align*} \]

This t-conorm has the general form of continuous archimedean t-conorms and is a solution of the equations:
\[ S_A(0.5, 0.3) = 1 \text{ and } S_A(0.45, 0.3) = 0.9. \]

Using the information in the conditions of the example we can see that both redundancy and complementarity is given. Redundancy appears in the scientific subjects (mathematics and physics are redundant to each other). Complementarity appears between the elements in S and the other element (literature).

Maxiness and drasticness can be measured for both t-conorms \( S_S \) and \( S_A \) and the corresponding results (see below) are given below. On one hand, \( S_A \) is more similar to the max t-conorm than to the drastic one. Therefore the set A is defined with redundant information sources. On the other hand \( S_S \) is more similar to the drastic t-conorm than to the maximum one. Thus, the set S is defined with complementary information.

<table>
<thead>
<tr>
<th></th>
<th>Maxiness ( 10 )</th>
<th>Drasticness ( 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_S )</td>
<td>0.0247</td>
<td>0.3289</td>
</tr>
<tr>
<td>( S_A )</td>
<td>0.2827</td>
<td>0.1539</td>
</tr>
</tbody>
</table>

Due to the fact that hierarchical structures are modular because their nodes can be defined independently, they also can be modified easily. In particular, they allow to refine an existing structure adding new information sources or removing existing ones. We give below an example of a refinement of the fuzzy measure defined in Example 4.1. We consider the addition of an extra criteria consisting on the marks on the subject Mathematical Logic.

Example 4.2. Let us consider a refinement \( \mu' \) of the fuzzy measure \( \mu \) on the set \( X = \{M, P, L\} \) adding an extra element ML that stands for Mathematical Logic.

1) We need to define the measure for the singleton \{ML\}. As this is a scientific subject, we settle it with the same value as M and P. This is: \( \mu(\{M\}) = \mu(\{M\}) = \mu(\{P\}) = 0.45 \).

2) As previously stated, we want to avoid redundancy between scientific subjects. Besides of that, we pretend that the total importance of scientific subjects does not

\[ \mu(\{M\}) = \mu(\{P\}) = 0.45, \mu(\{L\}) = 0.3 \]
(measure on the singletons)
increase with the new element. This forces that: 
\[ \mu(\{M,P, ML\}) = 0.5 \]

The revision will affect the value of \( \mu(\{M,P\}) \) so that \( \mu(\{M\}) < \mu(\{M,P\}) < \mu(\{M,P, ML\}) \).

3) All the other considerations in the previous examples are kept. They are the boundary conditions, the measure on the singletons and the support between literature and scientific subjects (this latter condition is now also applied to ML).

Figure 3. A refinement of the hierarchy of elements in example 4.1

Taking all this into account, we have that the hierarchy of elements corresponding to the new fuzzy measure follows the structure given in Figure 3, and that as in the previous example we need only to define \( S_A \) and \( S_S \). However, as the addition of a criteria is on the set \( S \), only \( S_A \) is required to be modified. Therefore, \( S_A \) is left as it was in the previous definition. Now, considering \( \mu(S) \) constant and equal to 0.5, and keeping the t-conorm \( S_P \) being Yager's t-conorm, we have that the following equation has to hold:

\[
(0.45^w + 0.45^w + 0.45^w)^{1/w} = 0.5
\]

Therefore, this leads to have \( S_S(x,y) = (x^w + y^w)^{1/w} \)

with \( w = (\ln 3)/(\ln 0.5 - \ln 0.45) = 10.427 \)

This example has shown the easiness of refining a hierarchically decomposable fuzzy measure. A similar approach would be followed if we refine the mark on the subject mathematics introducing the result of several exams. Let ME1, ME2, ME3 be three exams on mathematical subjects. In this case, to adapt the measure we need to change the node \{M\} by a set of nodes corresponding to the exams and the node of all three exams. This is, if \( H \) is the hierarchy of elements, then

\[ H' = H - \{M\} \cup \{\{ME1\}, \{ME2\}, \{ME3\}, \{ME1, ME2, ME3\}\} \]

Also, we need to define the corresponding measure on the singletons, and the t-conorm \( S_\eta \) attached to the node \( \eta \). The single restriction is that \( S_\eta(\mu(\{ME1\}), \mu(\{ME2\}), \mu(\{ME3\})) = 0.45 \)

(the previous value for \( \mu(\{M\}) \)).

With this example, it has been shown that adding new elements in hierarchically decomposable fuzzy measures is easy. This fact contrasts with the case of modifying general fuzzy measures where the addition of a new element to a measure defined for \( n \) elements requires the assignment of \( 2^n \) values. Note that introducing the element \{ML\} requires \( 2^3 = 8 \) values and introducing \( \{ME1, ME2, ME3\} \) (on the same measure \( \mu \)) requires \( 2^6 - 2^3 = 56 \) new values.

5. Conclusions and future work

As it has been shown in this work, the use of hierarchically decomposable fuzzy measures presents some advantages in relation to general fuzzy measures. These advantages can be summarized as follows:

1. Reduced complexity. While general fuzzy measures require \( 2^n \) values (being \( n \) the number of elements) in order to be defined, hierarchically decomposable fuzzy measures require a hierarchical structure of the elements, \( n \) values (one for each element) and a t-conorm for each node in the tree (except for the nodes). This structure is flexible enough to be adapted for different complexities.

2. Naturality. Hierarchical structures are a natural way to structure knowledge. See for example the use of hierarchies to structure knowledge in frame systems [8] or to define programs in object oriented programming [7]. Also, the definition of information source
importance using hierarchies is natural. Moreover, interactions among sources can be easily seen in a pictorial representation.

3. **Modularity.** Information sources are gathered together in nodes. As nodes in the hierarchy are defined independently, modularity is achieved. This reduces the complexity of the refinements to be done later in the model. Changing a node (introducing new information sources or removing irrelevant ones) does not imply the modification of the whole structure but only the affected node. In fact, only the t-conorm attached to the node is required to be modified so that the value of the measure applied to the node is not affected.

4. **Simplicity on refining.** Modularity also helps when a refinement of a particular criteria is needed. Splitting of a particular criteria into a set of them is straightforward. Also, collapsing a set of criteria into a single one is an easy task.

5. **Simplicity on dealing with redundancy and complementarity.** As it has been argued before, hierarchically decomposable fuzzy measures are adequate for expressing redundancy and complementarity of the information sources. Besides of that, these measures can encompass both aspects and therefore an aggregation operator can be defined that use both kind of information.

As a future work, we plan to study the learning of hierarchically decomposable fuzzy measures. The learning of this measures, besides of the intrinsic relevance of building models for decision making or data fusion, are of interest because they can make explicit redundancy and/or complementarity of different information sources.

**Acknowledgements**
The author acknowledges partial support of the CYCIT project SMASH (TIC96-1138-C04-04).

**References**