# Approximating fuzzy measures by hierarchically decomposable ones

Josep Domingo-Ferrer
Dept. Comput. Eng. and Maths - ETSE
Universitat Rovira i Virgili
Av Països Catalans 26
43007 Tarragona, Catalonia, Spain
jdomingo@etse.urv.es

Vicenç Torra
Inst. Inv. en Intel·ligència Artificial - CSIC
Campus Universitat Autònoma de Barcelona
08193 Bellaterra, Catalonia, Spain
vtorra@iiia.csic.es

**Abstract** – *Choquet and Sugeno integrals are powerful data fusion operators for numerical data as they generalize some well known aggregation operators as arithmetic means, weighted means, order statistics, OWA operators, medians, and so on. However, real applications of these operators require the definition of the so-called fuzzy measure. Difficulties for defining these measures arise because the number of values to be determined in a fuzzy measure is $2^N$, being $N$ the number of values to be aggregated. On the one hand human experts are not usually able to supply the large amount of required values and on the other hand it is difficult to interpret fuzzy measures when learned from examples. To solve this problem, fuzzy measures of reduced complexity have been proposed in the literature. In this work we propose a method to approximate a general fuzzy measure by a Hierarchically Decomposable one (one type of fuzzy measure of reduced complexity). Two applications of the method can be underlined: (i) understanding general fuzzy measures learned from examples; (ii) complete fuzzy measures from non-complete ones (This is to find all $2^N$ values from a subset of them and, thus, helping experts on their definition).*

**Keywords:** Fuzzy measures, Hierarchically Decomposable Fuzzy Measures, Choquet integral, Sugeno integral, Fuzzy integral

## 1 Introduction

At present there exist a large number of aggregation operators for numerical data. These methods differ on the properties they satisfy and on the parameters they require. In this work we focus on a particular type of parameter that are used in several aggregation operators: fuzzy measures.

Fuzzy integrals (see [8] for details) are aggregation operators that integrate a function with respect to a fuzzy measure. When the data to be fused is seen as an application from the set of sensors to the set of numbers $\mathbb{R}$, fuzzy integrals can be interpreted as numerical aggregation operators and as such applied to fuse numerical information.

Fuzzy integrals besides of the function to be integrated require a parameter that is used to give information about the sources (their importance, their reliability, and so on). This parameter is the so-called fuzzy measure. See [2] for an up to date description of the state of the art of fuzzy measures and integrals.

From the aggregation point of view, and looking fuzzy integrals and fuzzy measures from the perspective of the well-known weighted mean, fuzzy measures are a generalization of the weighting vector used in the weighted mean. While in a weighted mean, importance can only be given to particular information sources (this is, each information source has a weight attached to it that evaluates its importance when computing the outcome or its reliability), in the case of a fuzzy measure, importance is assigned to groups of information sources.

Due to the fact that importance is assigned to sets of sources, when $N$ different sources are considered in an aggregation process, a fuzzy measure is a mapping from all the subsets of the set of $N$ sources into $[0, 1]$. This means that general fuzzy measures requires the definition of $2^N$ parameters.

In real applications, the number of parameters is typically large (the number of sensors that supply values or the number of criteria in a multi-criteria decision making application). This makes the approach of having an expert defining the fuzzy measure infeasible. To avoid the definition of such a large set of parameters, several families of fuzzy measures have been defined with reduced complexity. These families require a smaller number of parameters than the general case. Among these families, we underline Hierarchically decomposable fuzzy measures – HDFM for short – (introduced in [6] and analysed in [7]) that require $N$ real numbers and the definition of at most $N$ t-conorms (a well-known operation in fuzzy logic). Moreover, given a fuzzy measure of this family, a graphical representation can be given that helps on its interpretation when the number of sources $N$ is large.

In this work we present an algorithm for approximating general fuzzy measures by Hierarchically decomposable ones. The interest of having such approximation is twofold. On the one hand, it allows the definition of a complete measure (a measure that is defined for all subsets of the set of sources – I.e., defined for all $2^N$ values) from a non-complete one. Therefore, when an expert can supply some of the $2^N$ values, the algorithm can build from these values the complete measure. On the other hand, the approximated measure requires less parameters and, thus, it offers a more compact and better undertandable representation (also, the corresponding graphical representation can be given). Therefore, the approximate HDFM can be used to grasp the meaning of the original fuzzy measures. This is particularly suitable when the original fuzzy measure has been automatically determined from the examples. In such a case, it is rather difficult to interpret a complete fuzzy measure because the number of parameters is very large.

To present the algorithm for approximating fuzzy measures, the structure of the paper is as follows: Section 2 reviews most relevant concepts for fuzzy measures; Section 3 introduces the algorithm; Section 4 presents an example and Section 5 is devoted to the conclusions and future work.

## 2 On fuzzy measures: a review

In this section we review definitions that are used latter on in this work. We begin with the definition of a t-conorm and give the expression of a family of t-conorms that are used in the rest of the paper. Then, we review fuzzy measures and hierarchically decomposable ones. The section finishes with the definition of the Choquet integral to illustrate how measures can be combined with the data from sensors for aggregation. Sugeno integrals are not reviewed but their use is similar to Choquet integral (Sugeno integrals were introduced in [5]). Expressions and properties differ.

From now on, we assume that the set of sources is denoted by $X = \{x_1, \cdots, x_N\}$. This set of sources corresponds to either criteria, sensors, alternatives, ... Information to be fused is attached to these sources by means of a mapping $f$ (i.e., $f(x_i)$ is the value attached to source $x_i$). Therefore, $f(x_1), \cdots, f(x_N)$ are the values to be fused (e.g. using either an arithmetic mean - $\sum_i f(x_i)/N$ - or a Choquet integral).

When $X$ is a set of criteria (as in applications for multicriteria decision making), then $f(x_i)$ is the evaluation of an object according to criteria $x_i$ (e.g. price). When $X$ corresponds to a set of sensors, then $f(x_i)$ would correspond to the reading of the sensor $x_i$.

Fuzzy measures will be defined on subsets of $X$ to measure the importance or reliability of the subset. As hierarchically decomposable fuzzy measures are based on t-conorms (a well-known operation in fuzzy logic that corresponds to logical disjunction in classical logic – see [4] for fuzzy sets, fuzzy logic and related operators and justification of these operators) we start with their definition.

**Definition 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)** $S(a, 0) = a$ *(neutrum element)*

**(2)** $b \leq d$ *implies* $S(a, b) \leq S(a, d)$ *(monotonicity)*

**(3)** $S(a, b) = S(b, a)$ *(commutativity)*

**(4)** $S(a, S(b, d)) = S(S(a, b), d)$ *(associativity)*

Two examples of t-conorms are the following ones:

1. The maximum t-conorm: $S(a, b) = max(a, b)$
2. The Yager family of t-conorms. This is a parametric definition (the parameter is $\alpha$) defined as follows:

$$S_\alpha(a, b) = \min(1, (a^\alpha + b^\alpha)^{1/\alpha})$$

Let us turn now into fuzzy measures. We start with the general definition of fuzzy measures.

**Definition 2** *A function $\mu : \wp(X) \to [0, 1]$ is a fuzzy measure if and only if it satisfies the following axioms:*

**(i)** $\mu(\emptyset) = 0, \mu(X) = 1$ *(boundary conditions)*

**(ii)** $B_1 \subseteq B_2 \subseteq X$ *implies* $\mu(B_1) \leq \mu(B_2)$ *(monotonicity)*

Fuzzy measures replace the axiom of additivity in probability measures ($\mu(A \cup B) = \mu(A) + \mu(B)$ when $A \cap B = \emptyset$) by a more general one: monotonicity ($\mu(A \cup B) \geq \mu(A)$). Thus, probability measures are also fuzzy measures. Fuzzy measures are used in Choquet integrals to express the importance of a set of information sources. In this setting, when additivity is not satisfied it means that the importance of a set is not the addition of the importance of the elements by themselves. In this framework, weighting vectors for weighted means are seen as equivalent to additive fuzzy measures.

When $\mu$ satisfies that for all $A, B \in \wp(X)$ with $A \subseteq B$ we have that $\mu(A \cup B) = S(\mu(A), \mu(B))$ for a given t-conorm $S$ we say that $\mu$ is decomposable.

A detailed description of fuzzy measures and their properties is given in [8].

The requirement that $\mu(X) = 1$ is not always included in the definition of fuzzy measures. In fact, the measure of the whole set is arbitrary. Measures satisfying this boundary condition are called normalized fuzzy measures. In fact, when fuzzy measures are used for aggregation this constraint is appropriate and it is equivalent to the one of a weighting vector (for the weighted mean) that weights adds to one. Decomposable fuzzy measures also require this constraint (note that t-conorm are operations in the unit interval).

Another class 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 by means of a dendogram with n-ary nodes. This is, they follow a structure similar to the one in Figure 1.

**Definition 3** *[6] H is a hierarchy of elements X if and only if the following conditions are fulfilled:*

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

**(ii)** $|\{r|$ *does not exists any*

$$h \in H \text{ such that } r \in h\}| = 1$$

**(iii)** *for all* $n \neq root$,

$$|\{h_i | n \in h_i \text{ and } h_i \in H\}| = 1$$

**(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* $|h| \neq 1$,
*it is satisfied that for all* $h_i \in h$, $h_i \in H$.

The conditions stated below are understood as follows: (i) all the elements in $X$ belong to the hierarchy (the elements $a \in X$ are the leaves, this is $\{a\} \in H$); (ii) there is only one element in the hierarchy that is not included in another one (this is the *root*); (iii) all nodes in 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.

When the hierarchy is used for information fusion, the leaves correspond to the information sources and the hierarchy makes explicit the similarities / dissimilarities between sources or shows the relations between independent / non-independent sources. *Related* sources are put together in a single node. The rationale and properties of the approach are given in [7].

**Example 1** *The hierarchy of Figure 1 will be represented as shown below if we follow the approach in Definition 3.*
$H = \{\{ML\}, \{M\}, \cdots, \{LL\}, \{\{ML\}, \{M\}\}, ScS = \{\{\{ML\}, \{M\}\}, \{P\}\}, LL = \{\{L\}, \{LL\}\}, \{ScS, LL\} = \{\{\{\{ML\}, \{M\}\}, \{P\}\}, \{\{L\}, \{LL\}\}\}\}$

To proceed with the definition of hierarchically decomposable fuzzy measures we need the extension operator ($EXT$). This operator when applied to a node returns all those sources that are encompassed in the node. This is, the extension of a node $h \in H$ returns all the elements of $X$ that are embedded in $h$. Then we give the definition of the so-called labeled hierarchy. A labeled hierarchy attaches a value to each leaf and a t-conorm for each node (that is not a leaf).

**Definition 4** *[6] The extension EXT of a node h in a hierarchy H is defined as:*

**(i)** $EXT(h) = \{h\}$, *if* $|h| = 1$

**(ii)** $EXT(h) = \cup_{h_i \in h} EXT(h_i)$, *if* $|h| \neq 1$

**Definition 5** *[6] A labeled 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 the set of real numbers. For the sake of simplicity we express* $S(h)$ *as* $S_h$.

Hierarchically decomposable fuzzy measures are defined from labeled hierarchies. The measure of the singletons is defined as the value given by the function $m$, and the values for all non-singletons $B$ are computed recursively using some of the t-conorms that are associated with the nodes $n$ such that their extension $EXT(n)$ intersects with $B$. This is formalized below:

**Definition 6** *[6] Given a labeled hierarchy* $L =< H, S, m >$ *of elements X, we define its corresponding hierarchically S-decomposable fuzzy measure (HDFM for short) as* $\mu(B) = \mu_{root}(B)$ *where* $\mu_A$ *is defined as:*

$\mu_A(B) = 0$, *if* $|B| = 0$

$\mu_A(B) = m(B)$, *if* $|B| = 1$

$\mu_A(B) = S_A(\mu_{a_1}(B_1), \cdots, \mu_{a_n}(B_n))$, *if* $|B| > 1$

*where* $A = \{a_1, \cdots, a_n\}$ *and* $B_i = B \cap EXT(a_i)$ *for all* $a_i$ *in* $A$

According to this definition, a fuzzy measure of this family computes the measure for a subset B in X in terms of the measure of disjoint subsets of B. These latter measures are combined using the t-conorm S of the smallest node in the tree that encompasses all the elements in B.

Now, for the sake of completeness we give the definition of the Choquet integral (see [Grabisch, 1998] for details and some of its properties). This integral can be used as 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. Sugeno integral [5] is an alternative integral that also combines values with respect to fuzzy measures.

**Definition 7** *Given a fuzzy measure* $\mu$*, the Choquet integral of a function* $f$ *with respect to* $\mu$ *is defined by:*

$$C_\mu(f, X) = C_\mu(f(x_1), \cdots, f(x_N)) =$$
$$= \sum_{i=1,N} (f(x_{s(i)}) - f(x_{s(i-1)}))\mu(A_{s(i)})$$

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

# 3 Approximating fuzzy measures by HDFM

In this section we introduce the algorithm for approximating arbitrary fuzzy measures by Hierarchically Decomposable ones. The algorithm we present is an iterative greedy algorithm that builds a hierarchical structure for all the information sources and attaches to each node in the hiearchy a t-conorm. The algorithm starts considering the set of information sources and then, in each iteration step, puts together some of the sources (or already defined sets of nodes) into a new node.

To simplify the process, the algorithm only considers binary trees. In fact this restriction on the type of trees built is not really a restriction on the fuzzy measures that can be built as for all *n-ary* HDFM there exists an equivalent *binary* HDFM. This equivalency is due to the fact that t-conorms are associative. Note that here *n-ary* and *binary* corresponds to the hierarchical structure of the measure. This can be formalized in the following way:

**Proposition 1** *Let* $\mu$ *be a HDFM induced from the labeled hierarchy* $L =< H, S, m >$ *such that some of the nodes n in the hierarchy H have cardinality greater than 2 (i.e., there exist* $n \in H$ *such that* $|n| > 2$*), then there exists a labeled hierarchy* $L' =< H', S', m >$ *such that the induced fuzzy measure* $\mu_{L'}$ *is equal to* $\mu$ *and for all* $n \in H$ *such that n is not a leaf,* $|n| = 2$*. This is,* $\mu(A) = \mu_{L'}(A)$ *for all* $A \subseteq X$.

**proof:** It is based on the associativity of t-conorms. Note that given a node $A$ defined by, e.g., three elements $\{a, b, c\}$ the measure $\mu(EXT(A) \cup \{d, e, f\})$ can be either computed by $S_A(S_A(m(a), m(b)), m(c))$ or by $S_A(m(a), m(b), m(c))$ combined with the measure of $\{d, e, g\}$.

The algorithm for building the approximation is given in Algorithm 1. This is a generic greedy algorithm. The algorithm receives the measure to be approximated ($\mu_o$ corresponds to the Original measure) and the set of sources $X$ (in fact, the set is implicit in the measure) and returns the HDFM measure (we call $\mu_n$ the New measure). Then it defines the set $Y$ as the set of nodes in the hierarchy that have to be gathered together (this somehow corresponds to a set of *pending* nodes) and defines the measure for the singletons of the new measure ($\mu_n(\{x_i\})$) being equal to the measure for the singletons in the original measure ($\mu_o(\{x_i\})$). While the number of *pending* nodes is not one, some of them have to be put together – aggregated – to build a new node. At each step some nodes in $Y$ are selected (this defines the set $Y_i$) and then the t-conorm for this new node is computed (in the algorithm, we use $S(Y_i)$ to denote the t-conorm for $Y_i$). The structure $Y$ is then updated accordingly. This is, the individual nodes that are put together in $Y_i$ are removed from $Y$ and the new node $Y_i$ is added to $Y$.

Note that this is a greedy algorithm because at each step, an additional node is added to the hierarchy and when a node is added its position is never modified. However, it is a generic one because no information is given on how to select the nodes. In the rest of this section, our approach for node selection and measure determination is explained (we start with the latter and then we go into the details of node selection). Node selection is based on the assumption that $Y_i$ is always a set of two nodes. Thus defining a binary hierarchy. For convenience in definitions and explanation, instead of a set we consider $Y_i$ as being a pair (order is considered). In the general structure, $Y_i$ is considered as a set as required in the definition of HDFM.

## 3.1 t-conorm selection

To simplify the process of selection of a t-conorm we have considered only t-conorms of a single family. In

**Algorithm 1** Greedy algorithm for approximating fuzzy measures by Hierarchically Decomposable Fuzzy Measures

```
Algorithm greedy-HDFM (X = {x_1, ..., x_N}: set
  of sources; μ_o: original fuzzy measure) is
  begin
    Y = X;
    μ_n({x_i}) = μ_o({x_i});
    i = 1;
    while |Y| ≠ 1 do
      Y_i := subset of Y(Y, μ_o);
      S(Y_i) := t-conorm for Y_i(Y, μ_o, Y_i);
      Y := Y − Y_i;
      Y := Y ∪ {Y_i};
      i := i + 1;
    end while
    return μ_n =< Y, S, m >;
  end
```

this case, it is appropriate to have a parameterized family of t-conorms. This reduces the complexity of the algorithm to avoid selection among arbitrary t-conorms and on the other hand it allows some customization when selecting the parameter that approximates better the original fuzzy measure.

In our case, the selected family of fuzzy measures is the Yager family reviewed in Section 2. Therefore, for each node $Y_i$, we need to determine the real value to settle the corresponding Yager t-conorm. Being $S(Y_i)$ a Yager t-conorm determined by its parameter, we will denote the parameter of $S(Y_i)$ by $S(Y_i)$ itself. So, $S(Y_i)$ will denote either a t-conorm or its parameter (a real number).

Our approach to find the t-conorm for $S(Y_i)$ is to minimize the difference between the HDFM $\mu_n$ and the original one $\mu_o$ for all the subsets of $X$ that have a measure computed using $S(Y_i)$. This is so because the minimum difference between both values would correspond to the best approximation. In fact this statement has to be made more precise, because we only use those subsets $A =< B, C >$ such that $\mu_n(A)$ is of the form $\mu_n(A) = S_{Y_i}(\mu_n(B), \mu_n(C))$ for some $B$ and $C$. So, we do not consider $A$ where the use of $S_{Y_i}$ is internal to the computation of $\mu(B)$ and $\mu(C)$. As explained latter this is due to technical difficulties.

Let $Y_i$ be $Y_i =< y_1^i, y_2^i >$, then a subset of $X$ such that their measure is computed using the t-conorm $S(Y_i)$ is the union of one subset of $EXT(y_1^i)$ and one subset of $EXT(y_2^i)$. Therefore, we consider the two sets $\beta_1$ and $\beta_2$ that are all subsets of $EXT(y_1^i)$ and $EXT(y_2^i)$, and the product of all $\beta_1$ and $\beta_2$. This is formally expressed by:

$$\beta_{Y_i} = \beta_1 \times \beta_2$$

where $\beta_1 = \wp(EXT(y_1^i)) - \{\emptyset\}$ and $\beta_2 = \wp(EXT(y_2^i)) - \{\emptyset\}$ and where, $A \times B$ stands for the set $\{c | c = a \cup b, a \in A, b \in B\}$.

Note that we need to use the function $EXT$ to obtain the elements of $X$ because $y_1^i$ and $y_2^i$ are pieces of the hierarchical structure.

Therefore, for a given $Y_i$ of the form $Y_i =< y_1^i, y_2^i >$, the best t-cornom $S$ for $Y_i$ is the one that minimizes the addition of:

$$\left(S(\mu_n(EXT(b_1)), \mu_n(EXT(b_2))) - \mu_o(b_1 \cup b_2)\right)^2$$

for all $b_1 \subseteq EXT(y_1^i)$ and $b_2 \subseteq EXT(y_2^i)$

Or, equivalently, and assuming a t-conorm of the Yager family ($S_\alpha$), we need to find $\alpha$ that minimizes the following expression:

$$\sum_{\beta \in \beta_{Y_i}} \left(S_\alpha(\mu_n(EXT(y_1^i) \cap \beta), \mu_n(EXT(y_2^i) \cap \beta)) - \mu_o(\beta)\right)^2$$

where $\beta_{Y_i} = \beta_1 \times \beta_2$, $\beta_1 = \wp(y_1^i) - \emptyset$ and $\beta_2 = \wp(y_2^i) - \emptyset$

Replacing the t-conorm $S$ by its definition, we get:

$$\textstyle\sum_{\beta \in \beta_{Y_i}} \Big(\min \quad (((\mu_n(EXT(y_1^i) \cap \beta))^\alpha + (\mu_n(EXT(y_2^i) \cap \beta))^\alpha)^{1/\alpha}, 1) - \mu_o(\beta)\Big)^2$$

It has been argued in the introduction that for building the HDFM approximation we do not require that the original fuzzy measure $\mu_o$ is complete. This is, it is possible for some sets $\beta$, the measure $\mu_o(\beta)$ is not known. When this value is not available, we use, instead, an approximation of the measure. To do so, first, we compute the interval where the measure lies. As the measure is monotonic, we know that the value $\mu_o(\beta)$ should belong to the interval $[max_{\alpha \subseteq \beta} \mu_o(\alpha), min_{\alpha \supseteq \beta} \mu_o(\alpha)]$. Then, we approximate the value by the central point of this interval. Thus, we approximate the measure by:

$$\mu_o'(\beta) = \frac{(max_{\alpha \subseteq \beta} \mu_o(\alpha) + min_{\alpha \supseteq \beta} \mu_o(\alpha)}{2}$$

Note that the algorithm described in this section is correct because when determining the value $\alpha$ for a node $Y_i$, the values required for such computation are already known. In particular, for the determination of the expression to minimize:

- Values $\mu_o(\beta)$ are known because they correspond to the original fuzzy measure
- Values $\mu_n(EXT(y_1^i) \cap \beta)$ and $\mu_n(EXT(y_2^i) \cap \beta)$ are known because the sets $EXT(y_1^i) \cap \beta$ and

$EXT(y_2^i) \cap \beta$ are, respectively, subsets of $EXT(y_1^i)$ and $EXT(y_2^i)$ and the t-conorms for $y_1^i$ and $y_2^i$ (the same is also true for all the nodes in the hierarchy that are *below* these ones) are already determined.

Note also that in the case of a set $\beta$ with an unknown measure $\mu_o(\beta)$, the values needed to compute the approximation given above are available because they use values of the original measure.

Finally it is important to underline that a HDFM defined according to the algorithm described in the previous section and using the t-conorm selection as described here builds a fuzzy measure. This is so because the measure build satisfy boundary conditions (the t-conorm forces that $\mu_n(X) = 1$ and monotonicity (this is so because HDFM are fuzzy measures as proven in [6]).

We have said above that when determining the t-conorm for a node $Y_i$, we do not really use all values $\mu_n(A)$ that require $S(Y_i)$ for their calculation but only some of them. In particular, we only use the sets $A$ such that their measure $\mu(A)$ follows the expression: $\mu(A) = S_{Y_i} = (\mu(B), \mu(C))$. Note that sets that do not follow this form also use the t-conorm $S(Y_i)$ if they can be expressed as: $B \cup C$ for $B \subseteq EXT(y_1^i) \cup EXT(y_2^i)$ and $C \subseteq X - EXT(y_1^i) - EXT(y_2^i)$ ($B$ and $C$ not equal to $\emptyset$). In such case, the computation of $\mu(A)$ would use the t-conorm $S(Y_i)$ but this computation should be combined by a t-conorm that is still unknwon.

## 3.2 Set selection

Following the hypothesis of considering only binary hierarchies, $Y_i$ is defined by two sets. Therefore, at each step two of the sets in $Y$ have to be selected on the basis of some properties.

The general approach is to select sets on the basis of their behavior with respect to the other sets. In particular, we put together first those sources that have a similar behavior. As the only information at our disposal is the original fuzzy measure, we use this fuzzy measure to determine the common behavior.

In principle, the selection is done considering the two sets that are more correlated in relation to the measures of all non common supersets. The rationale of the approach is as follows: If $\alpha_1$ and $\alpha_2$ are correlated, it means that their behaviour in relation to the other sets (e.g. set $\tau$) is similar. Therefore, if we would need a t-conorm $S_{\alpha_1}$ to combine the measure of the set $\alpha_1$ and the set $\tau$, and a t-conorm $S_{\alpha_2}$ to combine the measure of the set $\alpha_2$ and the set $\tau$, these t-conorms are *similar* and thus if sets $\alpha_1$ and $\alpha_2$ are put together, the t-conorm needed to combine the new set $\alpha_1 - \alpha_2$ (say, $S_{\alpha_1 - \alpha_2}$) with $\tau$ would be similar to $S_{\alpha_1}$ and $S_{\alpha_2}$ ($S_{\alpha_1-\alpha_2}$ *similar* to $S_{\alpha_1}$ and *similar* to $S_{\alpha_2}$). According to this, the aggregation of the nodes would not introduce large errors.

We formalize below the definition of correlation for a set.

Let us consider the correlation for the pair $(\alpha_1, \alpha_2)$. Let $T = X - EXT(\alpha_1) - EXT(\alpha_2)$ be the set of sources not included neither in $\alpha_1$ nor in $\alpha_2$. Then, subsets with either $\alpha_1$ or $\alpha_2$ and some elements in $T$ are built (i.e., $\alpha_1 \cup \tau$ and $\alpha_2 \cup \tau$ for some $\tau \subseteq T$). Correlation between the fuzzy measures for these two sets are calculated. This is:

$$c(\alpha_1, \alpha_2) = correlation\{\mu_o(\alpha_1 \cup \tau)\mu_o(\alpha_2 \cup \tau)\}_{\tau \subseteq T}$$

where $T = X - EXT(\alpha_1) - EXT(\alpha_2)$

It is important that $T$ excludes the elements in $\alpha_1$ and $\alpha_2$ to avoid including in the correlation the same sets.

To select a pair $Y_i = < y_1^i, y_2^i >$ we have to compute the correlations for all pairs $(\alpha_1, \alpha_2)$ with $\alpha_1, \alpha_2$ in $Y_i$. The pair with maximum correlation will be selected.

In this definition of correlation all supersets with non common elements in $EXT(\alpha_1) \cup EXP(\alpha_2)$ are considered, however, in general, when non-complete fuzzy measures are considered, we need to remove in the correlation process all sets without a defined measure.

Although, in general, this value can be computed for a pair $(\alpha_1, \alpha_2)$, there are some situations in which some difficulties arise. In fact, this is so when there are not enough values to compute the correlation (less or equal than 2). This can be caused because the cardinality of $Y$ is small or because there are several non-defined values that reduce the number of available measures.

In the case that the correlation cannot be applied, we approximate it by an average of correlations. If we need to compute the correlation between $\alpha_1$ and $\alpha_2$, we approximate the correlation between $\alpha_1$ and $\alpha_2$ by the average of the correlation of all pairs of elements $x_i, x_j$ with $x_i \in \alpha_1$ and $x_j \in \alpha_2$.

# 4 Example

We have applied the approach described in this work to the non-complete fuzzy measure described in Table 1. This measure is based on the one in [1] for the evaluation of a set of students with respect to their marks in several subjects. Undefined values are marked in this table with *UD*.

**Example 2** *A high school needs to prepare a ranking of a set of students according to their level in different subjects. The subjects under considerations are: mathematics (M), physics (P), mathematical logic (ML), literature (L) and latin literature (LL). To determine the importance of the subjects, a fuzzy measure is to* *be defined. The definition of the measure follows the following guidelines (here only some of the values are given, the complete listing of the values are given in Table 1 column $\mu_o$):*

1. *Boundary conditions:*

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

2. *Relative importance of scientific versus literay subjects:*

   $\mu(\{M\}) = \mu(\{P\}) = 0.45$, $\mu(\{L\}) = 0.3$, $\mu(\{L\}) = 0.25$

3. *Support between literature and scientific subjects:*

   $\mu(\{M, L\}) = \mu(\{P, L\}) = 0.8 > \mu(\{P\}) + \mu(\{L\})$

4. Redudancy *between scientific subjects:*

   $\mu(\{ML, M, P\}) = 0.5 < \mu(\{ML\}) + \mu(\{M\}) + \mu(\{P\}) = 1.35$

For building the approximate measure, the first step is to find the sets to be aggregated. To do so we need to compute the correlation for all pairs of elements in $Y$. As initially $Y = X$ this corresponds to compute the correlation for all pairs of individual sources (the correlation between pairs of subjects). These correlations are displayed in Table 2.

On the basis of the correlations displayed in Table 2, the subjects selected to be put together are Mathematical Logic and Mathematics. Therefore, the new node will contain the elements $\{ML, M\}$ and, thus, $Y_1 =< \{ML\}, \{M\} >$.

At this point, the t-conorm $S(Y_1)$ has to be fixed. As we have assumed Yager t-conorms, we need to determine its parameter $\alpha$. As detailed above, this corresponds to a minimization problem. However in this case, being $\beta_1 = \wp(EXT(\{ML\})) - \emptyset$ and $\beta_2 = \wp(EXT(\{M\})) - \emptyset$ it results that there is only one set in $\beta_{Y_i}$. This is, $\beta_{Y_i} = \{\{ML, M\}\}$. Therefore, the minimization problem is equivalent to the following equation (minimum value will be achieved when difference is zero):

$$\left(((\mu_n(\{ML\}))^\alpha + (\mu_n(\{M\}))^\alpha)^{1/\alpha} - \mu_o(\{ML, M\})\right) = 0$$

However, in this case, $\mu_o(\{ML, M\})$ is not known. Therefore, we approximate this value with the central point of the interval $[0.45, 0.5]$ and then, the equation leads to $\alpha = 12.82$.

Now, we need to select the next nodes to be fused. This requires the computation of the correlation between the set $\{ML, M\}$ and the other nodes. At this point, there are some values that cannot be calculated but approximated using the mean. For example, the correlation between $\{ML, M\}$ and $\{P\}$ is approximated by the mean of the value for $\{ML, P\}$ and the value for $\{ML, P\}$ in Table 2. Table 3 displays correlations and approximated values (the latter are marked with an asterisk).

Then, according to Table 3, the new node will be the one with $\{ML, M, P\}$ and $Y_2 =< \{\{ML\}, \{M\}\}, \{P\} >$. The corresponding $\alpha$ equals to 11.40. Then, to select the next node, the correlation between new pairs should be computed. The new correlation values are displayed in Table 4. As before, the values with an asterisk correspond to approximated values. Here, the correlation between $\{ML, M, P\}$ and $\{L\}$ is the mean of the values for the pairs $< \{ML\}, \{L\} >$, $< \{M\}, \{L\} >$ and $< \{P\}, \{L\} >$. and the correlation for $\{ML, M, P\}$

| | $\mu_o$ | $\mu_n$ |
|---|---|---|
| {} | 0.0 | 0.0 |
| {LL, } | 0.25 | 0.25 |
| {L, } | 0.3 | 0.3 |
| {L, LL, } | 0.4 | 0.40000156279769838 |
| {P, } | 0.45 | 0.45 |
| {P, LL, } | 0.75 | 0.79877913532159395 |
| {P, L, } | 0.8 | 0.85890823555748663 |
| {P, L, LL, } | 0.9 | 0.97643805635298986 |
| {M, } | 0.45 | 0.45 |
| {M, LL, } | 0.75 | 0.79877913532159395 |
| {M, L, } | 0.8 | 0.85890823555748663 |
| {M, L, LL, } | 0.9 | 0.97643805635298986 |
| {M, P, } | UD | 0.47821280611055256 |
| {M, P, LL, } | 0.95 | 0.82976957911936433 |
| {M, P, L, } | 0.95 | 0.89027075642642362 |
| {M, P, L, LL, } | UD | 1.0 |
| {ML, } | 0.45 | 0.45 |
| {ML, LL, } | 0.75 | 0.79877913532159395 |
| {ML, L, } | 0.8 | 0.85890823555748663 |
| {ML, L, LL, } | 0.9 | 0.97643805635298986 |
| {ML, P, } | UD | 0.47821280611055256 |
| {ML, P, LL, } | 0.95 | 0.82976957911936433 |
| {ML, P, L, } | 0.95 | 0.89027075642642362 |
| {ML, P, L, LL, } | UD | 1.0 |
| {ML, M, } | UD | 0.475 |
| {ML, M, LL, } | 0.95 | 0.82624632724255465 |
| {ML, M, L, } | 0.95 | 0.88670577911997461 |
| {ML, M, L, LL, } | UD | 1.0 |
| {ML, M, P, } | 0.5 | 0.4933359570423006 |
| {ML, M, P, LL, } | UD | 0.84633475354796484 |
| {ML, M, P, L, } | UD | 0.90703019877629909 |
| {ML, M, P, L, LL, } | 1.0 | 1.0 |

Table 1: $\mu_o$: non-complete fuzzy measure to be approximated

|  | ML | M | P | L | LL |
|---|---|---|---|---|---|
| ML | 1 | 1 | 1 | 0.4549 | 0.4265 |
| M | 1 | 1 | 1 | 0.4549 | 0.4265 |
| P | 1 | 1 | 1 | 0.4549 | 0.4265 |
| L | 0.4549 | 0.4549 | 0.4549 | 1 | 0.9965 |
| LL | 0.4265 | 0.4265 | 0.4265 | 0.9965 | 1 |

Table 2: Correlation between pairs of subjects: step 1

|  | {ML,M} | P | L | LL |
|---|---|---|---|---|
| {ML,M} | 1 | 1* | 0.4549* | 0.4265* |
| P | 1* | 1 | 0.4549 | 0.4365 |
| L | 0.4549* | 0.4549 | 1 | 0.9965 |
| LL | 0.4265* | 0.4549 | 0.99659 | 1 |

Table 3: Correlation between pairs of subjects: step 2. Asterisks refers to approximated values

and $\{LL\}$ is the mean of the values in the table for the pairs $< \{ML\},\{LL\} >$, $< \{M\},\{LL\} >$ and $< \{P\},\{LL\} >$.

|  | {ML,M,P} | L | LL |
|---|---|---|---|
| {ML,M,P} | 1 | 0.4549* | 0.4265* |
| L | 0.4549* | 1 | 0.9965* |
| LL | 0.4265* | 0.9965* | 1 |

Table 4: Correlation between pairs of subjects: step 3. Asterisks refers to approximated values

As the largest value in Table 4 (without considering the diagonal) is the correlation between $L$ and $LL$, these are the elements to be aggregated to define the new node. Therefore, $Y_3 =< \{L\},\{LL\} >$.

Once these elements are aggregated, there are only two nodes in the hierarchy and, thus, $Y_4 =< \{\{\{ML\},\{M\}\},\{P\}\},\{\{L\},\{LL\}\} >$. This node completes the hierarchy.

The resulting hierarchy is shown in Figure 1. This figure clearly identifies the two main groups of subjects *Scientific subjects* ($ScS$) and *Literary subjects* ($L$). $ScS$ and $LS$ point out these two groups in the hierarchy. This figure includes for each node the corresponding value for $\alpha$. Table 1 includes the approximated measure for each subset of $X$ (column $\mu_n$).

## 5 Conclusions and future work

In this work we have introduced an algorithm for approximating general non-complete fuzzy measures by Hierarchically Decomposable ones. The approach can be used either for understanding a general fuzzy measure obtained from examples (e.g., using the algorithms described in [3]) and to obtain a complete fuzzy measure from a non-complete one. We have shown the application of the approach to a particular non-complete fuzzy measure.

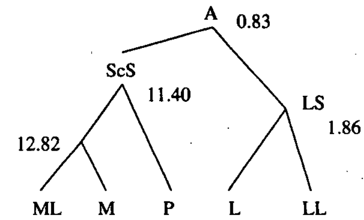


Figure 1: Hierarchy of elements in the approximated fuzzy measure $\mu_n$ and the corresponding values for $\alpha$

The approach has to be extensively tested and future work is planned in this direction. A detailed comparison between the generated fuzzy measures in relation to the original ones is needed.

## Acknowledgements

Josep Domingo-Ferrer and Vicenc Torra are partially supported by the EU project CASC: Contract: IST-2000-25069 and CICYT project STREAMOBILE (TIC2001-0633-C03-01/02).

## References

[1] Grabisch, M., (1995), Fuzzy integral in multicriteria decision making, Fuzzy Sets and Systems 69 279-298.

[2] Grabisch, M., Murofushi, T., Sugeno, M., 2000, (Eds.), Fuzzy Measures and Integrals: Theory and Applications, Physica-Verlag.

[3] Imai, H., Miyamori, M., Miyakosi, M., Sato, Y., (2000), An algorithm Based on Alternative Projections for a Fuzzy Measures Identification Problem, Proc. Int. Conf. Soft Comp., Iizuka, Japan.

[4] Klir, G., Yuan, B., (1995), Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall, U.K.

[5] Sugeno, M., (1974), Theory of Fuzzy Integrals and its Applications. (PhD Dissertation). Tokyo Institute of Technology, Tokyo, Japan.

[6] Torra, V., (1999), On Hierarchically S-Decomposable Fuzzy Measures, Int. J. of Intel. Systems, 14:9 923-934.

[7] Torra, V., (2000), On a family of measures for data fusion with reduced complexity, Proc. of the Information Fusion conference, Paris, France.

[8] Wang, Z., Klir, G., (1992), Fuzzy measure theory, Plenum Press, New York.