HAL
open science

Revised HLMS: a useful algorithm for fuzzy measure identification

J. Murillo, Serge Guillaume, E. Tapia, P. Bulacio

To cite this version:

J. Murillo, Serge Guillaume, E. Tapia, P. Bulacio. Revised HLMS: a useful algorithm for fuzzy measure identification. Information Fusion, 2013, 14 (4), p. 532 - p. 540. 10.1016/j.inffus.2013.01.002 . hal-00862610

HAL Id: hal-00862610

https://hal.science/hal-00862610

Submitted on 17 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

Revised HLMS: A useful algorithm for fuzzy measure identification

Javier Murillo[a], Serge Guillaume[b], Elizabeth Tapia[a], Pilar Bulacio[a]

[a] *CIFASIS-CONICET, Universidad Nacional de Rosario, Argentina.*
[b] *Irstea, UMR ITAP, 34196 Montpellier, France.*

Abstract

An important limitation of fuzzy integrals for information fusion is the exponential growth of coefficients for an increasing number of information sources. To overcome this problem a variety of fuzzy measure identification algorithms has been proposed. HLMS is a simple gradient-based algorithm for fuzzy measure identification which suffers from some convergence problems. In this paper, two proposals for HLMS convergence improvement are presented, a modified formula for coefficients update and new policy for monotonicity check. A comprehensive experimental work shows that these proposals indeed contribute to HLMS convergence, accuracy and robustness.

Keywords: Choquet integral, gradient descent, convergence, multicriteria

1. Introduction

In different fields of human knowledge information comes from several sources so that a fusion process is required to summarize information. Typical applications include sensor fusion [1], cooperative classifiers [2] and selection of alternatives in multicriteria decision making.

Information fusion can be accomplished by means of univariate or multivariate approaches. Univariate approaches model information sources individually and thus, relationships between them are dismissed. According to the context, univariate approaches may use different fusion operators, the *mean*, *minimum* or *maximum* are typical choices. Multivariate approaches model relationships between information sources so that redundancy, complementarity and independence between sources are taken into account. In this case, fuzzy integral operators, such as Choquet [3] and Sugeno [4] inte-

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

grals could be used. We note, however, that multivariate approaches require much more coefficients than their univariate counterparts.

Different approaches have been proposed in literature, including classical optimization techniques [5, 6], genetic algorithms [7, 8] and artificial neural networks [9, 10]. These approaches either require large datasets (classical optimization techniques), they depend on random steps (genetic algorithms) which make its traceability difficult or depend on particular network structure (artificial neural network). To overcome these limitations, we consider the alternative use of gradient-based descendant algorithms. In particular, we consider the Heuristic Least Mean Squares (HLMS) algorithm originally introduced by Grabisch in [11] and later modified in [12]. Briefly, HLMS implements a gradient approach which minimizes a squared error criterion between the expected and estimated values of targets. Remarkably, HLMS is likely to yield acceptable solutions avoiding unsatisfactory and rather counter-intuitive alternatives [13]. In recent work, a preliminary study of HLMS algorithm was performed [14] and convergence problems were observed.

In this paper, a revised HLMS implementation, which includes a new formula for the iterative estimation of fuzzy measures coefficients and a new policy for monotonicity check, is presented.

This paper is organized as follows. In Section 2, basic concepts and definitions about fuzzy measures and integrals are reviewed. In Section 3, improvements of the HLMS algorithm are motivated and two modifications are proposed: a new formula for iterative fuzzy measure estimation and a more efficient policy for monotonicity check. In Section 4, main features of the revised HLMS algorithm are analyzed in detail; synthetic and benchmark datasets are used to evaluate accuracy, convergence behavior according to the HLMS learning rate parameter and sensitivity to target variations. Finally, in Section 5, main conclusions and perspectives of future work are discussed.

2. Basic definitions

This section introduces the basic concepts related to fuzzy measures and the discrete Choquet integral. The reader may refer to [15, 16, 17, 18, 19] for details. Let us consider a finite set $X = \{x_1, \ldots, x_i, \ldots, x_n\}$ and let $\mathcal{P}(X)$ denote the power set (set of all subsets) of X.

Definition 1 A fuzzy measure is a set function $\mu\colon \mathcal{P}(X) \to [0, 1]$ fulfilling the following two axioms:

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

1. $\mu(\emptyset) = 0$, $\mu(X) = 1$
2. $A \subseteq B \subseteq X \Rightarrow \mu(A) \leq \mu(B)$

The first axiom, also called the normalization axiom, allows a meaningful comparisons of fuzzy measures. The second axiom establishes a monotonicity condition. The coefficients of a fuzzy measure μ are the weights given to elements of $\mathcal{P}(\mathcal{X})$.

Definition 2 A discrete Choquet integral of a function $f\colon X \to \Re^+$ with respect to some fuzzy measure μ over X:

$$\mathcal{C}_\mu(f(x_1), .., f(x_n)) \triangleq \sum_{i=1}^{n} (f(x_{(i)}) - f(x_{(i-1)}))\mu(A_{(i)}) \tag{1}$$

where $A_{(i)} = \{x_{(i)}, x_{(i+1)}, \ldots, x_{(n)}\}$ and $x_{(\cdot)}$ is the rearrangement induced by $f(x_i)$, $i = 1, \ldots, n$, sorted in ascending order, i.e., $f(x_{(1)}) < \cdots < f(x_{(n)})$, with the convention that $f(x_{(0)}) = 0$.

A suitable way of representing fuzzy measures in the finite case is to use a lattice representation. Each element in $\mathcal{P}(\mathcal{X})$ is associated with a vertex, ordered by inclusion. Elements with the same cardinality are mapped to vertices in the same lattice level and cardinalities can be used to identify such levels. Hence, while the top lattice level labeled with 0 will contain only the empty set ϕ, the bottom lattice level labeled with n will contain the whole set X (see Figure 1).

Considering the definition of the discrete Choquet integral, just one coefficient per lattice level is required for $\mathcal{C}_\mu(X)$ computation. The corresponding set of lattice vertices defines a complete path across the lattice (see dashed path in Figure 1).

Definition 3 The neighbors of a lattice vertex are the set of connected vertices. The neighbors of a lattice vertex at level (l) are the connected vertices at levels $(l-1)$ and $(l+1)$.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

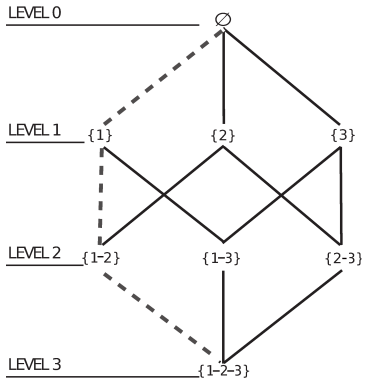

Figure 1: Lattice representation of the inclusion relation for a set of three attributes

3. The revised HLMS algorithm

HLMS [11] is a supervised algorithm for fuzzy measure identification. The training data for HLMS is a set S of m examples made up of n input attributes and the values of some target concept T:

$$S = \begin{pmatrix} x_1^1 & \dots & x_i^1 & \dots & x_n^1 & T^1 \\ \vdots & & \ddots & & \vdots & \\ x_1^j & \dots & x_i^j & \dots & x_n^j & T^j \\ \vdots & & \ddots & & \vdots & \\ x_1^m & \dots & x_i^m & \dots & x_n^m & T^m \end{pmatrix}$$

where x_i^j is the i-th attribute value on the j-th sample characterized by a target value T^j, $i = 1, \dots, n$ and $j = 1, \cdots, m$.

In multicriteria decision making, the i-th attribute denotes the i-th criterion and the j-th row denotes the j-th alternative, $x^j = x_1^j, \dots, x_n^j$. Hence, each x_i^j denotes the satisfaction degree of the i-th criterion for the j-th alternative and T^j denotes the decision value to be inferred from the n satisfaction degrees. Note that this aggregation process makes sense only if all satisfaction degrees are in the same scale.

Fuzzy measures are defined by a set of coefficients. The HLMS algorithm updates a set of initial coefficients according to the gradient of the square difference between the values of a known target T and its estimation obtained from Choquet integral result using current coefficients on training data S.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

In recent work [14], we observed that unpredictable HLMS convergence behavior could be explained by the fact that HLMS formula for coefficients update was not a true gradient. We later realized that although a true gradient update formula indeed improved HLMS convergence, there were some datasets for which the convergence problem remained unsolved. Further experimental work revealed that the problem could be in the naive HLMS implementation of the monotonicity check which was performed over the complete set of lattice paths. We noted that if we constrained monotonicity checks to those lattice paths defined by training samples in S, HLMS convergence remarkably improved. Therefore, the following Revised HLMS algorithm (1) was derived:

Algorithm 1 Revised HLMS

1: **Input:** training dataset S formed by samples $(\langle x_1^j, x_2^j, \ldots, x_n^j \rangle, T^j)$ where $j \in \{1, \ldots, m\}$
2: **Output:** 2^n fuzzy measure coefficients
3: **for** $i \in \mathcal{P}(X)$ **do** {Initialization}
4: $\quad u_i = \frac{|i|}{|X|}$
5: **end for**
6: Identify_untouched_coefficients()
7: **repeat**
8: $\quad$ examples $\leftarrow$ random$(1:m)$ {Sensitivity to data presentation order}
9: $\quad$ **for** $j \in$ examples **do**
10: $\quad\quad e^j = \mathcal{C}_\mu(x^j) - T^j$ {Individual error calculation}
11: $\quad\quad$ **for** $l \in (1 : n-1)$ **do**
12: $\quad\quad\quad u_l = u_l - \alpha \times \frac{e^j}{e_{max}} \times (x^j_{(n-l+1)} - x^j_{(n-l)})$ {Coefficient update}
13: $\quad\quad\quad$ Check_monotonicity_with_neighbors(u_l) {Monotonicity check }
14: $\quad\quad$ **end for**
15: $\quad$ **end for**
16: $\quad E \leftarrow \sqrt{\frac{1}{m} \sum_{s=1}^{m} (\mathcal{C}_\mu(x^s) - T^s)^2}$ {Error calculation}
17: **until** Stop_criterion_met()
18: Set_untouched_coefficients() {Final monotonicity correction}

1. *Initialization:* At first, fuzzy measure coefficients u_i are initialized to their equilibrium states [11] (L. 3). Then, those coefficients which do not participate in any integral computation for the given S are identified (L. 6) and left outside the main process. These coefficients are only modified in a final adjustment step (L. 18).
2. *Sensitivity to data presentation order:* Gradient algorithms are known to be sensitive to data presentation order. To prevent any bias, samples are randomly rearranged at each iteration step (L. 8).
3. *Individual error calculation:* For each sample, the error is calculated based on current coefficients (L. 10).

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

4. *Coefficient update:* After each sample presentation, the coefficients involved in the corresponding integral are updated (L. 12). This important step is discussed below.
5. *Monotonicity check:* The monotonicity condition is checked after each coefficient update (L. 13) but only with those coefficients in the neighborhood triggered by training samples. This important step is discussed below.
6. *Error calculation* The Root Mean of Squared Errors (RMSE) is computed over the whole training set (L. 16).
7. *Stop criterion:* The algorithm may stop if either the RMSE or the coefficients do not change during a given number of iterations. It is also possible to fix the number of iterations (L. 17).

In what follows, the two main modifications introduced by Revised HLMS, namely a true gradient formula for coefficients update and a revised policy for monotonicity check, are explained in detail. Their positive effects on HLMS convergence are then analyzed.

3.1. A true gradient formula for coefficients update

Let us consider a simple fuzzy measure estimation problem involving three attributes and let $\alpha > 0$ denote the learning rate parameter of the gradient-based HLMS algorithm for fuzzy measure estimation. Regarding the update formula (L. 12), let u_l denote the coefficient at the l-th lattice level triggered by the j-th sample. Hence, the Choquet integral becomes:

$$\mathcal{C}_\mu(x) = (x_{(1)} - x_{(0)})u_3 + (x_{(2)} - x_{(1)})u_2 + (x_{(3)} - x_{(2)})u_1 \tag{2}$$

so that, the partial derivative of the integral with respect to u_l becomes:

$$\frac{\partial \mathcal{C}_\mu(x)}{\partial u_l} = x_{(n-l+1)} - x_{(n-l)} \tag{3}$$

and thus, the following gradient-based formula for coefficients update rule follows:

$$u_l = u_l - \alpha \times \frac{e^j}{e_{max}} \times (x_{(n-l+1)} - x_{(n-l)}) \tag{4}$$

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

Note that Eq.4 differs from the one proposed in the original HLMS version [11] showed in Eq.5:

$$u_l = u_l - \alpha \times \frac{e^j}{e_{max}} \times (x^j_{(n-l)} - x^j_{(n-l-1)}) \tag{5}$$

It can be observed that the difference $x_{(n-l)} - x_{(n-l-1)}$ term in Eq.5 is not truly related to the gradient of the squared error. This problem is fixed by Revised HLMS.

3.2. A revised policy for monotonicity check

As highlighted in Algorithm (1), the monotonicity check step is achieved in two different ways depending on the fuzzy measure coefficient status (L. 13 and L. 18). This new proposition, another important difference with respect to the original HLMS formulation, favors the algorithm convergence.

Depending on data, there might be coefficients which do not belong to any path defined by the samples when performing the Choquet integral and then, their values are not modified by the update formula (L. 12). Those coefficients which do not participate in any integral computation strongly limit the evolution of neighbors coefficients, since their values do never change. Hence, it makes sense to identify them and relegate them (L. 6), i.e., when coefficient monotonicity of an updated coefficient is analyzed, they are not taken into consideration. Relegated coefficients will only be modified at the end of the algorithm using their new neighbors values (L. 18). Let $N_{(l-1)}$ and $N_{(l+1)}$ respectively denote the sets of neighbors in the upper and lower lattice levels of an untouched coefficient u. After completion of HLMS iterations, the u value fulfilling the monotonicity condition verifies Eq.6:

$$max\{u_k | k \in N_{(l-1)}\} \leq u \leq min\{u_k | k \in N_{(l+1)}\} \tag{6}$$

and thus, we set u to the mean of the above bounds.

Regarding those coefficients which participate in the main process, monotonicity is checked in only one direction depending on the error sign. When the error (L. 10) is negative, the coefficient u at l-th lattice level is increased according to the update formula (L. 12). In this case, we only need to check the monotonicity with coefficients in the $(l+1)$-th lattice level. Similarly,

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

when the error (L. 10) is positive, the coefficient u at l-th lattice level is decreased according to the update formula (L. 12). In this case, we only need to check monotonicity with coefficients in the $(l-1)$-th lattice level.

3.3. Convergence behavior of Revised HLMS

HLMS is an stochastic algorithm [20] involving gradient computations followed by weights updates after each sample presentation. It is well-known [21] that, in this case, convergence is ensured when, for each sample in the training set, the error decreases at each iteration step, k :

$$|\mathcal{C}_\mu(x)^{k+1} - T(x)| < |\mathcal{C}_\mu(x)^{k} - T(x)| \tag{7}$$

Since convergence analysis involves only one sample at the time, the j index will be dropped to alleviate the notation. Two cases must be considered for the convergence analysis, i) both $\mathcal{C}_\mu(x)^{k+1}$ and $\mathcal{C}_\mu(x)$ are less (or higher) than $T(x)$ and ii) one of them is greater than $T(x)$ while the other is not.

Case 1: Both values are on the same side of the target (above or below). This case is illustrated in Figure 2

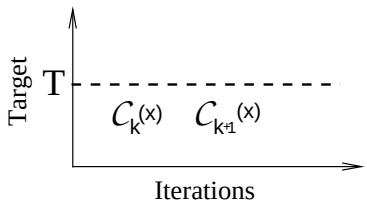

Figure 2: Both values $\mathcal{C}_k(x)$ and $\mathcal{C}_{k+1}(x)$ are below T (Case 1)

and thus, Eq.7 becomes:

$$T(x) - \mathcal{C}_\mu(x)^{k+1} < T(x) - \mathcal{C}_\mu(x)^{k} \Leftrightarrow \mathcal{C}_\mu(x)^{k+1} - \mathcal{C}_\mu(x)^{k} > 0 \tag{8}$$

As the integral is a sum of positive quantities, previous inequality holds if:

$$\exists l \mid u(l)^{k+1} > u(l)^k \ \wedge \ \forall m \neq l \; u(m)^{k+1} \geq u(m)^k \tag{9}$$

Only one coefficient is updated at a given step, so the second condition is obviously fulfilled. The first one means: $u(l)^{k+1} - u(l)^k > 0$. Substituting

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

$u(l)^{k+1}$ by its value according to the update formula (L. 12) in Algorithm (1) this leads to:

$$\alpha(\mathcal{C}_\mu^k(x) - T(x))(x_{n-l+1} - x_{n-l}) < 0 \tag{10}$$

$\alpha > 0$ and $(x_{n-l+1} - x_{n-l}) \geq 0$, $\forall l \in [1, n-1]$. $\mathcal{C}_\mu^k(x) - T(x) < 0$ in this case so the above product is negative. The demonstration is analogous when both values are greater than the target.

Case 2: The two values are on two different sides of the target. This case is illustrated in Figure 3

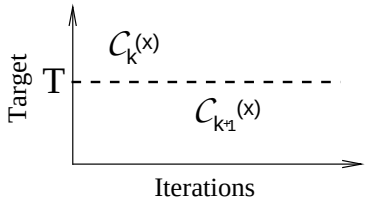

Figure 3: Both values $\mathcal{C}_k(x)$ and $\mathcal{C}_{k+1}(x)$ are on different sides of the target (Case 2)

and thus, Eq.7 becomes:

$$T(x) - \mathcal{C}_\mu(x)^{k+1} < \mathcal{C}_u(x)^k - T(x) \tag{11}$$

$$2T(x) - \mathcal{C}_\mu(x)^{k+1} - \mathcal{C}_\mu(x)^k < 0 \tag{12}$$

$$2T(x) - \sum_{i=1}^{n}(x_{(i)} - x_{(i-1)})\{u^{k+1}(A(i)) + u^k(A(i))\} < 0 \tag{13}$$

Remark that $u(A(i)) = u(i, \ldots, n)$ which is noted as $u_{(n-i+1)}$ because $A(i)$ contains $(n-i+1)$ elements. According to the update formula,

$$\begin{aligned} & u_{(n-i+1)}^{k+1} + u_{(n-i+1)}^{k} \\ &= 2u_{(n-i+1)}^{k} - \alpha(\mathcal{C}_\mu^k(x) - T(x))(x_{n-(n-i+1)+1} - x_{n-(n-i+1)}) \\ &= 2u_{(n-i+1)}^{k} - \alpha(\mathcal{C}_\mu^k(x) - T(x))(x_{(i)} - x_{(i-1)}) \end{aligned}$$

As $\sum_{i=1}^{n}(x_{(i)} - x_{(i-1)})u_{(n-i+1)}^{k} = \mathcal{C}_\mu^k(x)$, Eq.7 implies:

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

$$\alpha < \frac{2(\mathcal{C}_\mu^k(x) - T(x))}{\sum_{i=1}^{n}(x_{(i)} - x_{(i-1)})^2(\mathcal{C}_\mu^k(x) - T(x))} \tag{14}$$

And, finally:

$$\alpha < \frac{2}{\sum\limits_{i=1}^{n}(x_{(i)} - x_{(i-1)})^2} \tag{15}$$

This fraction allows a wide range of positive values for the learning rate. As previously said, the learning rate α controls how fast the algorithm identifies the fuzzy measure coefficients (converge). The denominator of Eq.15 is a sum of positive values, since data was rearranged. The upper bound becomes smaller when the differences $(x_{(i)} - x_{(i-1)})$ increase, meaning that the information from sources differ. In this case, a small value of α must be chosen in order to converge. Otherwise, when the sources give similar information, the learning process could be faster. As stated for case 1, the proof is analogous when the locations of the two values are inverted.

In both cases, as no assumption is made about the input values, this result holds for all the samples in the training set and ensures the algorithm convergence provided α is positive and fulfills Eq.15.

4. Experimental design

HLMS requires input datasets with attributes expressing their contribution to a target concept in a common scale (defined up to the same positive linear transformation). These datasets are hard to obtain [22] and thus, it is a common practice to use synthetic ones. In this work, synthetic datasets involving random attributes and predefined fuzzy measures for targets computation [11, 23] are used to evaluate the *convergence* and *accuracy* of the Revised HLMS algorithm. Moreover, studies are performed on coefficient variations of Revised HLMS to different targets obtained from partial orders over samples [15].

Finally, preprocessed real benchmark datasets described in [22] are used to assess Revised HLMS practical performance. Datasets are described in Table 1.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

Table 1: Datasets characteristics

Synthetic Datasets		
	# attributes	*# samples*
Students' grade (*students*)	4	17
Computer (*computers*)	3	6
Benchmark Datasets		
Employee Selection (ES)	4	488
Employee Rejection/Acceptance (ERA)	4	1000
Lecture Evaluation (LE)	4	1000
CPU (CPU)	6	209
Breast Cancer (CB)	7	278
Den Bosch (DB)	8	119

4.1. Convergence and Accuracy on synthetic datasets

Regarding the evaluation of Revised HLMS convergence and accuracy, two synthetic datasets are considered. The first dataset, hereafter named *computers*, is used to evaluate Revised HLMS convergence in detail. The *computers* dataset involves three attributes and three targets T1, T2 and T3 (see Table 2). Attributes follow from clients' satisfaction degrees over three computer features, processor speed (S), ram memory (R) and hard disk capacity (D). Targets T1, T2 and T3 are respectively computed using the *mean*, the *minimum* (*min*) and the *maximum* (*max*) operators on samples' attributes.

Table 2: Computer satisfaction degrees

Computer	S	R	D	T1	T2	T3
1	0.96	0.78	0.42	0.72	0.42	0.96
2	0.96	0.66	0.54	0.72	0.54	0.96
3	0.36	0.78	0.42	0.52	0.36	0.78
4	0.36	0.66	0.54	0.52	0.36	0.66
5	0.60	0.36	0.74	0.56	0.36	0.74
6	0.65	0.26	0.84	0.56	0.26	0.84

The second dataset, hereafter named *students*, is used to evaluate Revised HLMS accuracy. The *students* dataset involves four attributes and two targets T1 and T2 (see Table 3). Attributes follow from students' grades in four subjects, literature (L), geography (G), physics (P) and mathematics (M). Attributes are normalized and share a common scale so that the higher the score the better the grade. Target T1 is computed using the fuzzy measure

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

shown in Table 4 while target T2 is obtained from T1 after the addition of uniformly distributed noise in [-0.1,0.1].

Table 3: Normalized students' grade data

Student	L	G	P	M	T1	T2
1	0.9	1.0	0.9	0.8	0.9	0.939
2	0.8	0.9	0.1	0.8	0.87	0.864
3	0.8	0.9	0.9	0.8	0.84	0.885
4	0.7	0.8	0.8	1.0	0.82	0.821
5	0.7	0.9	0.8	0.7	0.78	0.790
6	0.7	0.7	0.8	0.8	0.74	0.718
7	0.7	0.8	0.8	0.6	0.70	0.620
8	0.6	0.7	0.7	0.7	0.68	0.734
9	0.7	0.6	0.7	0.6	0.65	0.604
10	0.7	0.6	0.6	0.6	0.63	0.635
11	0.4	0.6	0.6	0.7	0.58	0.679
12	0.5	0.4	0.5	0.8	0.52	0.578
13	0.5	0.6	0.6	0.4	0.50	0.410
14	0.4	0.6	0.4	0.6	0.50	0.511
15	0.4	0.6	0.2	0.6	0.44	0.500
16	0.5	0.4	0.4	0.4	0.43	0.499
17	0.4	0.4	0.2	0.3	0.32	0.273

Table 4: Coefficients to aggregate students' grades. For each Coalition, the corresponding Coefficient is given

Coalition	Coefficient	Coalition	Coefficient	Coalition	Coefficient
∅	0	{1-2}	0.5	{1-2-3}	0.6
{1}	0.3	{1-3}	0.45	{1-2-4}	0.7
{2}	0.4	{1-4}	0.45	{1-3-4}	0.6
{3}	0.3	{2-3}	0.4	{2-3-4}	0.8
{4}	0.2	{2-4}	0.5	{1-2-3-4}	1
		{3-4}	0.4		

4.1.1. Convergence

Convergence depends on the choice of the learning rate parameter, α. A high value of α is likely to increase convergence speed but also to make the algorithm unstable. Figure 4 shows the evolution of coefficients value for coalition {2-3} using Revised HLMS on the *students* dataset. Four α parameterizations are considered, $\alpha = 0.1$, 1, 10 and 100. As shown in the three upper plots of Figure 4, convergence speed increases with respect to α.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

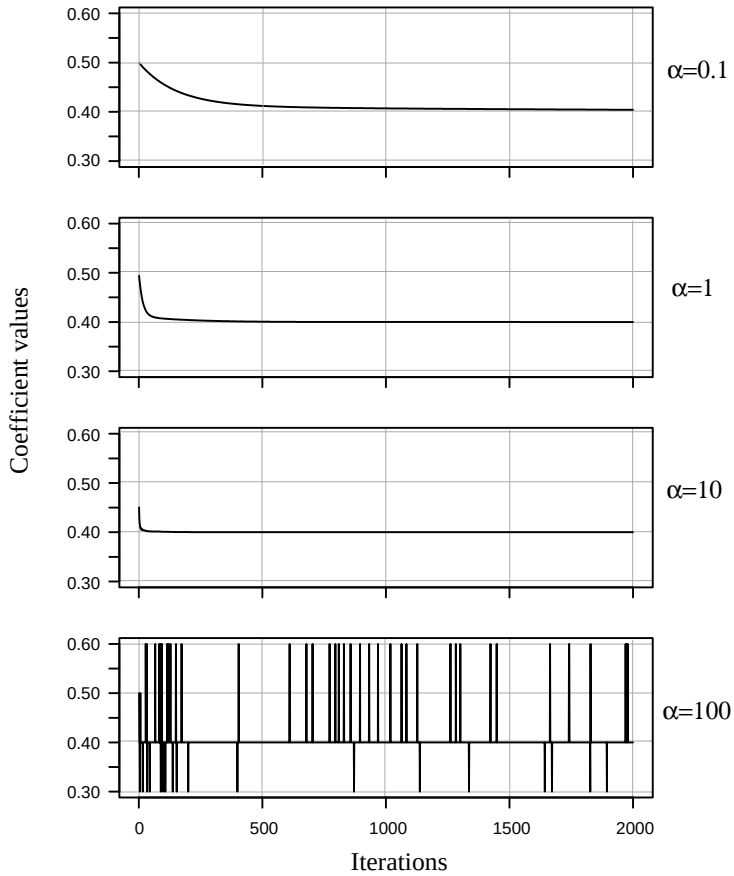

Figure 4: Revised HLMS convergence behavior on *students* data. The evolution of the coefficient $\mu\{2\text{-}3\}$ for an increasing number of iterations is shown for $\alpha = 0.1$, 1, 10 and 100

Bottom plot of Figure 4 shows that when α value is high some coefficients start to oscillate.

4.1.2. Accuracy

Firstly, to gain insight into the accuracy of Revised HLMS, the *computer* dataset is used. The *computer* dataset is small enough to track Revised HLMS steps in detail. Moreover, all lattice paths are covered allowing all the coefficients to be updated. Targets T1, T2 and T3 in the *computer* dataset are computed using the *mean*, the *min* and the *max* aggregation operators. The coefficients of the Choquet integral for these type of targets are well-known [15]. For the *mean* operator, Choquet integral coefficients values are

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

exactly those used in the initialization step (L. 4), then, a convergence without error is obtained in one iteration without any coefficient update. For the *min* operator, Choquet integral coefficients values must be all zeros except $\mu(X) = 1$. Finally, for the *max* operator, Choquet integral coefficients values must be all ones except $\mu(\emptyset) = 0$. These expected results are used to check the accuracy of Revised HLMS algorithm. Later, predefined coefficients are used to calculate the target values of students' grades and they are compared with the ones returned by Revised HLMS.

Aggregation operators *min* and *max* have an analogous behavior. The evolution of RMSE and coefficients according to the number of iterations are shown in Figure 5 and Figure 6.

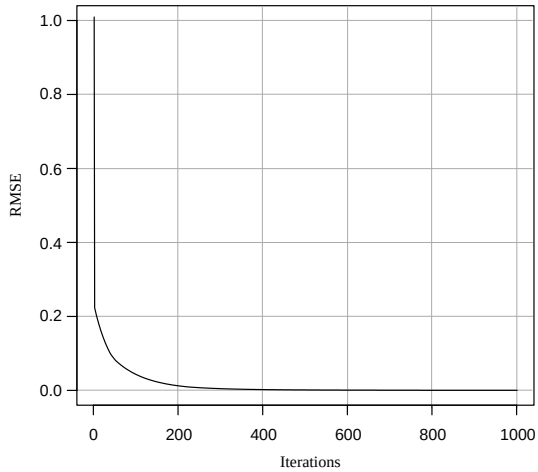

Figure 5: Root mean square error evolution when Choquet integral generalize aggregation operator *minimun*. The horizontal axis shows the iteration number. Vertical axis shows the values of RMSE

Figure 5 clearly shows the convergence of the error to the expected value, zero. In Figure 6, several convergence speeds can be observed.

The singleton coefficients converge faster than the ones of coalitions. This is due to the initialization: singleton coefficients are set to 1/3, closer to their final value, 0, than the others, which are initialized to 2/3. One of the coefficients, coalition {2-3}, converges slower than the others. This effect is data dependent. Only two samples, 3 and 4, update this coefficient with a small difference in the attributes, $(0.42 - 0.36)$, in both cases. For example, for coalition {1-2}, this difference is $(0.78 - 0.42)$. The previous remark also explains the difference between error and coefficient convergence. The error

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

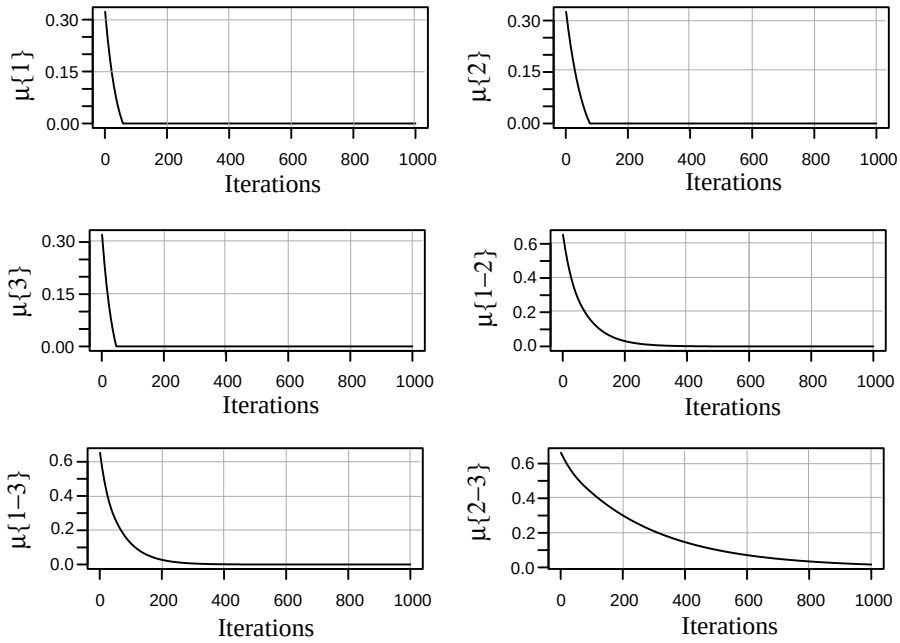

Figure 6: Convergence evolution of singletons {1},{2} and {3} and coalitions {1-2}, {1-3} and {2-3} when Choquet integral generalizes the aggregation operator *minimun*. The horizontal axis shows the iteration number. The vertical axis shows coefficient value

is kept low even if this coefficient has not reached its final value because its influence in the RMSE is quite low.

Now, the convergence of Revised HLMS for *students* dataset using T1 and T2 is performed (see Table 5). The algorithm is run with α=0.1. Each row corresponds to a sample with its target value and the solution for different numbers of iterations. The RMSE evolution is shown in the last row. With the T1 target, a quite low error value is reached after 10 iterations (0.008) and the final value is close to 0, as expected. With noisy data T2, the solution is no longer exact, the RMSE is 0.03 instead of 0 (0.045 after 10 iterations).

The coefficients of Choquet integral that best fit *students* dataset are the ones in Table 4 (unique solution). This allows a comparison between them and the coefficients returned by Revised HLMS. The structure of Table 6 is similar to the previous one.

Each row corresponds to a given coefficient. The value in column (C), used to compute the T1 target, and the coefficients estimated by Revised HLMS for different numbers of iterations are shown. They all tend to their

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

Table 5: Students' grade convergence of sample targets

Sample	T1	it:10	it:100	it:1000	T2	it:1000
1	0.90	0.899	0.896	0.897	0.939	0.879
2	0.87	0.874	0.870	0.867	0.864	0.864
3	0.84	0.849	0.845	0.840	0.885	0.839
4	0.82	0.823	0.816	0.816	0.821	0.859
5	0.78	0.774	0.771	0.776	0.790	0.779
6	0.74	0.749	0.749	0.743	0.718	0.735
7	0.70	0.723	0.716	0.702	0.620	0.679
8	0.68	0.675	0.676	0.681	0.734	0.700
9	0.65	0.649	0.648	0.645	0.604	0.638
10	0.63	0.625	0.625	0.629	0.635	0.638
11	0.58	0.574	0.573	0.580	0.679	0.629
12	0.52	0.547	0.532	0.521	0.578	0.570
13	0.50	0.523	0.516	0.502	0.410	0.479
14	0.50	0.499	0.497	0.499	0.511	0.529
15	0.44	0.449	0.444	0.440	0.500	0.481
16	0.43	0.425	0.425	0.429	0.499	0.438
17	0.32	0.324	0.322	0.319	0.274	0.315
RMSE	0	0.008	0.006	0.002	0	0.03

Table 6: *Student*'s grade convergence of coefficients. The coefficients coalitions are shown in the first column. The 'C' column represents the values were the coefficients must converge. Columns 3 to 6 represent the coefficient value in 10, 100, 1000 and 10000 iterations

Coefficients	C	it:10	it:100	it:1000	it:10000
∅	0	0	0	0	0
{1 }	0.3	0.250	0.259	0.293	0.3
{2 }	0.4	0.250	0.258	0.354	0.399
{3 }	0.3	0.249	0.247	0.270	0.299
{4 }	0.2	0.241	0.198	0.174	0.194
{1-2 }	0.5	0.499	0.496	0.491	0.499
{1-3 }	0.45	0.498	0.484	0.457	0.449
{1-4 }	0.45	0.499	0.494	0.494	0.460
{2-3 }	0.40	0.494	0.456	0.406	0.400
{2-4 }	0.5	0.498	0.489	0.496	0.499
{3-4 }	0.4	0.499	0.490	0.436	0.400
{1-2-3 }	0.6	0.745	0.710	0.622	0.599
{1-2-4 }	0.7	0.747	0.732	0.706	0.700
{1-3-4 }	0.6	0.747	0.730	0.695	0.620
{2-3-4 }	0.8	0.751	0.767	0.816	0.803
{1-2-3-4 }	1	1	1	1	1

expected value. As previously discussed the convergence speed is not the same for all coefficients as it is data dependent.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

4.2. Coefficients sensitivity to target variations

Previous experiments assume well-known targets but in real applications, targets are rarely known and must be expert-defined. Experts may assign different target to a given sample. For similar targets, Revised HLMS is expected to give similar coefficients. In this section, the targets are induced by a partial order over samples. Briefly, let us assume all the attributes have the same scale and meaning, for instance the higher the score the better the grade for attributes in the *students* dataset. The alternative i is preferred to alternative j if all the attributes values for i are higher or equal to their equivalent for j and there is at least one of them strictly higher. This order relationship is formalized in Eq.16.

$$\forall k \quad x_k^i \geq x_k^j \quad and \quad \exists\, p \quad x_p^i > x_p^j \Rightarrow i \succ j \tag{16}$$

with $k, p \in \{1..n\}$. If neither $i \succ j$ nor $j \succ i$, then i and j are not comparable.

For the *students* dataset, Eq.16 means that student #1 should be preferred to student #3 and to all students in the range $[5, 17]$ and that student #1 cannot be compared to students #2 and #4. In addition, all students should be preferred to student #17. These type of preference could be used as a constraint to generate a ranking of the students; those students which cannot be compared can be randomly ranked. As a result a large set of rankings can be induced from the *students* datasets. Among all such possible rankings, two of them are randomly selected as shown in Table 7:

Table 7: R1 and R2 are two rankings over the *students* dataset. For each student number in the first row, its corresponding ranking in R1 (R2) is given the second (third) row

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
R1	2	3	4	1	6	5	8	10	11	9	13	7	15	12	16	14	17
R2	3	1	4	2	5	6	7	8	11	10	9	14	12	15	16	13	17

For each of the two rankings R1 and R2, 100 targets are randomly generated using two constraints, i) the rank constraint: if sample i ranks before sample j, then $T(i) \geq T(j)$ and ii)the range constraint: for each sample i, $min(x_k^i) \leq T(i) \leq max(x_k^i)$ holds. For each group of 100 targets derived either from R1 or R2, a set of 100 fuzzy measures are obtained using Revised HLMS. Fuzzy measures within each group of targets are expected to be similar but different from those derived from the other group. Hence, a

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

group G1 (G2) of 4950 pairwise Euclidean distances between fuzzy measures [24] for targets induced from ranking R1 (R2) is computed. In addition, a group G3 of 10000 pairwise Euclidean distances between fuzzy measures for targets induced from ranking R1 and R2 is computed.

Three groups are considered:

- Group 1: d(i, j) —$\forall$i, j = i; i, j $\in$ R1 (4950 distances)
- Group 2: d(i, j) —$\forall$i, j = i; i, j $\in$ R2 (4950 distances))
- Group 3: d(i, j) —$\forall$i $\in$ R1, j$\in$ R2 (10000 distances)

An ANOVA test is carried out to test for differences among the three groups, G1, G2 and G3 (see Figure 7). The factor under study is the distance between fuzzy measures in each group. Results are shown in Table 8. The test highlights the influence of the factor and shows that the null hypothesis must be rejected, i.e., at least one of the groups G1, G2 and G3 is significantly different from the others.

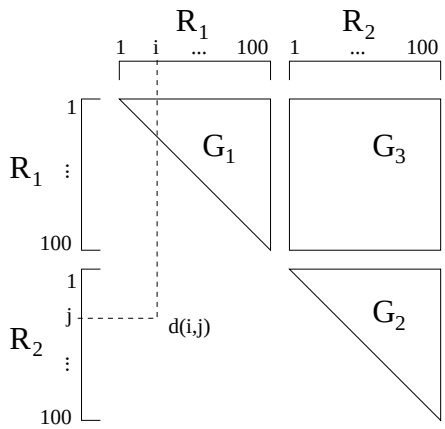

Figure 7: Group configuration

A Tukey test is then performed to quantify differences between the means of groups G1, G2 and G3. Results are shown in Table 9:

These results, taken together, suggest that Revised HLMS coefficients for targets induced from different ranks are indeed different.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

Table 8: ANOVA test on groups G1, G2 and G3 related to fuzzy measures for targets from rankings R1, R2 and R1-R2

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Factor	2	14175	7.0876	14932	<2.2e-16
Residuals	19897	9444	0.0005		

Table 9: Tukey multiple comparisons of means for groups G1, G2 and G3. 95% family-wise confidence intervals $[lwr, upr]$ for the differences *diff* between means are given

	diff	lwr	upr
2-1	0.0028	0.0018	0.0038
3-1	0.0547	0.0538	0.0556
3-2	0.0519	0.0510	0.0527

4.3. Revised HLMS on benchmark data

In this section, we evaluate Revised HLMS performance on six benchmark datasets from the UCI repository[1] and the WEKA machine learning toolbox[2]. The datasets are preprocessed, following the methodology described in [22], in order to convert each attribute value into a satisfaction degree. As usually occurs with real data, the generated target are tainted with noise. Therefore, convergence with non-null root mean square error (RMSE) should be expected.

For each benchmark dataset, differences between target values and Choquet integral calculations based on Revised HLMS fuzzy measure estimations are evaluated and the corresponding RMSEs across samples are computed. Revised HLMS is run to 1000 iterations and $\alpha = 0.05$. Results are shown in Figure 8.

RMSE decreases with the number of iterations without any oscillation, and, converges, as expected, to a non null value. In order to check the algorithm accuracy, new targets are generated for the ESL, ERA and LE datasets using the fuzzy measure given in Table 10. In this case, the algorithm is expected to identify the correct coefficients and to converge with a zero error. For all the datasets, the algorithm behaves as expected. These results are il-

[1]http://archive.ics.uci.edu/ml/
[2]http://www.cs.waikato.ac.nz/ml/weka/

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

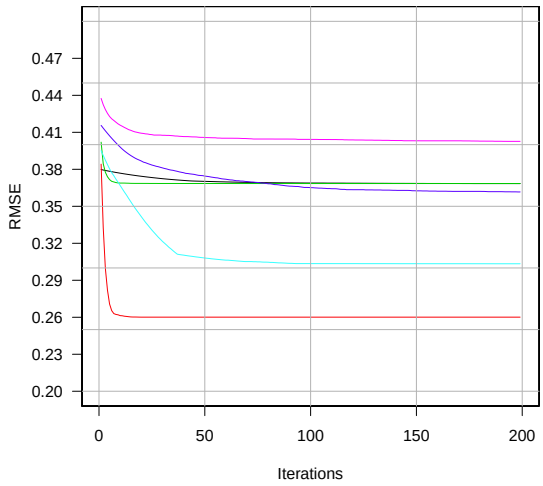

Figure 8: The evolution of the RMSE on the six benchmark datasets

lustrated in Figure 9, RMSE congence to zero, and in Figure 10, convergence to expected value for some of the coefficients.

Table 10: Fuzzy measure coefficients defining new targets for the ESL, ERA and LE datasets. For each Coalition, the corresponding Coefficient is given

Coalition	Coefficient	Coalition	Coefficient	Coalition	Coefficient
∅	0	{1-2}	0.35	{1-2-3}	0.7
{1}	0.2	{1-3}	0.2	{1-2-4}	0.8
{2}	0.3	{1-4}	0.5	{1-3-4}	0.9
{3}	0.1	{2-3}	0.45	{2-3-4}	0.75
{4}	0.4	{2-4}	0.65	{1-2-3-4}	1
		{3-4}	0.6		

5. Conclusions and future work

HLMS is a gradient descent algorithm for identifying the Choquet integral parameters that best fit training data. This paper describes a revised version where the update formula and the monotonicity check have been improved in order to guarantee the algorithm convergence.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

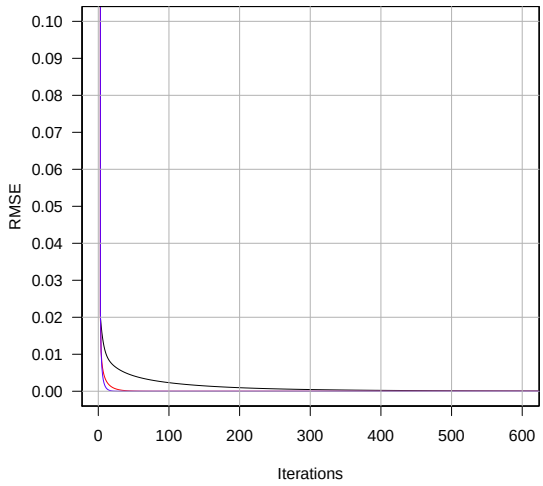

Figure 9: Convergence of the RMSE for the ESL, ERA and LE datasets. Target values were generated using fuzzy measure coefficients in Table 10

The convergence property is deeply analyzed. It highly depends on the learning rate: this work shows that the range of possible values is wide. To manage this trade-off between speed and convergence, an adaptive learning rate could be implemented.

The algorithm accuracy, i.e. its ability to identify the proper coefficients, is tested comparing its output with well known aggregation operator parameters. When the target is not defined, it can be generated from a sample ranking. Note that different rankings may be compatible with a dataset, and various target values are allowed for a given ranking. Anyway, HLMS proves to be sensitive to the ranking and robust to the targets associated with the ranking.

Simulation experiments were performed on synthetic and preprocessed real benchmark datasets which are usually used for the evaluation of classification algorithms. This was motivated by the lack of publicly available datasets in the format required by the Choquet Integral aggregation scheme, i.e., both the attributes and the target values must represent a preference degree. Such datasets might be hard to obtain in practice. This suggests the need to develop automatic preprocessing strategies for the conversion of real datasets to the Choquet integral format. This is work in progress.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

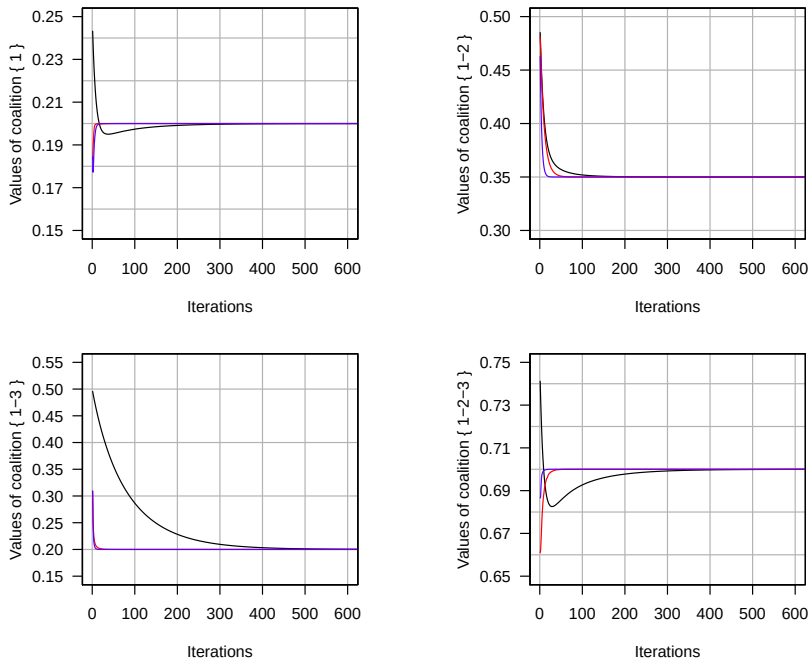

Figure 10: Convergence of attribute {1} and coalitions {1-2}, {1-3}, {1-2-3} to predefined fuzzy measure coefficients in Table 10 for the ESL, ERA and LE datasets

Altogether, our results contribute to the robustness of HLMS, an algorithm with desirable characteristics for fuzzy measure identification.

References

[1] B. Khaleghi, A. Khamis, F. O. Karray, S. N. Razavi, Multisensor data fusion: A review of the state-of-the-art, Information Fusion (0) (2011) –. doi:10.1016/j.inffus.2011.08.001.
URL http://www.sciencedirect.com/science/article/pii/S1566253511000558

[2] P. Bulacio, S. Guillaume, E. Tapia, L. Magdalena, A selection approach for scalable fuzzy integral combination, Information Fusion 11 (2) (2010) 208 – 213. doi:10.1016/j.inffus.2009.06.003.
URL http://www.sciencedirect.com/science/article/pii/S1566253509000621

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

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

[4] M. Sugeno, Theory of fuzzy integrals and its applications, Ph.D. thesis, Tokio Institute of Technology (1974).

[5] J.-L. Marichal, M. Roubens, Determination of weights of interacting criteria from a reference set, European Journal of Operational Research 124 (3) (2000) 641–650.

[6] M. Suo, Y. Li, G. Huang, Multicriteria decision making under uncertainty: An advanced ordered weighted averaging operator for planning electric power systems, Engineering Applications of Artificial Intelligence 25 (1) (2012) 72 – 81. doi:10.1016/j.engappai.2011.08.007. URL http://www.sciencedirect.com/science/article/pii/S0952197611001503

[7] Z. Wang, K. Leung, J. Wang, A genetic algorithm for determining nonadditive set functions in information fusion, Fuzzy Sets and Systems 102 (1999) 436–469.

[8] Z. Wang, K. L. Wong, J. Wang, A new type of nonlinear and the computational algorithm, Fuzzy Sets and Systems 112 (2000) 223–231.

[9] J. Wang, Z. Wang, Using neural networks to determine sugeno measures by statistics, Neural Networks 10 (1997) 183–195.

[10] J. Wang, Z. Wang, Detecting constructions of nonlinear integral systems from input-output data: An application of neural network, in: Fuzzy Information Processing Society, North America, 1996.

[11] M. Grabisch, A new algorithm for identifying fuzzy measures and its application to pattern recognition, in: Fourth IEEE international conference on fuzzy systems, Yokohama, Japan, 1995, pp. 145–150.

[12] M. Grabisch, E. Raufaste, An empirical study of statistical properties of the choquet and sugeno integrals, IEEE Transactions on Fuzzy Systems 16 (4) (2008) 839 –850. doi:10.1109/TFUZZ.2008.917295.

[13] P. Miranda, M. Grabisch, Optimization issues for fuzzy measures, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.

[14] J. Murillo, S. Guillaume, E. Tapia, P. Bulacio, On the convergence of hlms algorithm (2011) 817–821.

Author-produced version of the article published in Information Fusion, 2013, 14(4), 532-540.
The original publication is available at http://www.sciencedirect.com
DOI : 10.1016/j.inffus.2013.01.002

[15] V. Torra, Y. Narukawa, Modeling Decisions, Springer, 2007.

[16] M. Grabisch, H. Nguyen, E. Walker, Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference, Kluwer Academic, 1995.

[17] Z. Wang, G. J. Klir, Fuzzy measure theory, Plenum, 1992.

[18] D. Denneberg, Non-Additive Measure and Integral, Kluwer Academic, 1994.

[19] E. Pap, Null-Additive Set Functions, Kluwer Academic, 1995.

[20] T. Mitchell, Machine Learning, McGraw-Hill, 1997.

[21] L. Bottou, Une approche théorique de l'apprentissage connexionniste: Applications à la reconnaissance de la parole, Ph.D. thesis, Université de Paris XI, Orsay, France (1991).
URL `http://leon.bottou.org/papers/bottou-91a`

[22] A. F. Tehrani, W. Cheng, E. Hüllermeier, Choquistic regression: generalizing logistic regression using the Choquet integral, in: S. Galichet, J. Montero, G. Mauris (Eds.), Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology, Atlantis Press, Aix-Les-Bains, France, 2011, pp. 868–875.

[23] S. H. Alavi, J. Jassbi, P. J. Serra, R. Ribeiro, Comparison of genetic and gradient descent algorithms for determining fuzzy measures, in: IEEE 10th International Conference on Intelligent Engineering Systems (INES 06), London (UK), 2006.

[24] K. I, Quadratic distances for capacity and bi-capacity approximation and identification, 4OR: A Quarterly Journal of Operations Research 5:2 (2007) 117–142.