k-maxitive fuzzy measures: a scalable approach to model interactions

Murillo Javier[1] **Serge Guillaume**[2] **Bulacio Pilar**[1,3]

[1]CIFASIS-CONICET, Universidad Nacional de Rosario, Argentina.
[2]Irstea, UMR ITAP, 34196 Montpellier, France.
[3]Facultad Regional San Nicolás, Universidad Tecnológica Nacional, Argentina.

Abstract

Fuzzy measures are powerful at modeling interactions between elements. Unfortunately, they use a number of coefficients that exponentially grows with the number of elements. Beyond the computational complexity, assigning a value to any coalition of a large set of elements does not make sense. k-order measures model interactions involving at most k elements. The number of coefficients to identify is reduced and their modeling capacity is preserved in real problems where the number of interacting elements is limited. In extreme situations of full redundancy or complementariness, it is mathematically proven that the complete fuzzy measure is both k-additive and k-maxitive. A learning algorithm to identify k-maxitive measures from labeled data is designed on the basis of $HLMS$ (Heuristic Least Mean Squares). In a classification context, the study of synthetic data with partial redundancy or complementariness supports the idea that the difference between full and partial interaction is a matter of degree, not of kind. Dealing with two real world datasets, a comparison of the complete fuzzy measure and a k-maxitive one shows the number of interacting elements is limited and the k-maxitive measures yield the same characterization of interactions and a comparable classification accuracy.

Keywords: Choquet, fuzzy measure, HLMS, Shapley, Möbius, k-order measures.

1. Introduction

In order to design interpretable and robust classification models, discrete fuzzy measures [23] may contribute to characterize set behavior in a complex data domain, i.e. data with high dimension, correlation, or noise. The interpretability goal points to the comprehension of the relationships between inputs (feature subsets) and outputs (class labels), to achieve more compact and computationally simpler models. Hence, let $N = \{1, \cdots, n\}$ be a set of elements, features in the case of classifier design. A fuzzy measure, μ, weighs all subsets $A \subset N$, $0 \leq \mu(A) \leq 1$, to state the coalition importance for the classification process. To go further in the expressiveness of set behavior, e.g. characterizing redundancy or complementariness, other fuzzy measure representations are considered: the Möbius representation, m of μ, to characterize the type and strength of interactions among the elements of N; and the interaction index [6, 21], I, to characterize the average contribution of a coalition considering all subsets it is part of. Hence, μ, m, and I representations provide different viewpoints of a set characterization.

Despite the descriptive power of fuzzy measures, their practical implementation is limited by the coefficient identification complexity: n elements require the evaluation of 2^n-2 coefficients. This exponential growth is their *Achilles's heel*, restricting their use to problems with a handy number of elements. Trying to overcome the identification scalability, simplified fuzzy measures have been proposed based on the inclusion of new restrictions. The λ-measures [24] reduce the number of coefficients to be identified to n+1, the singletons and λ, but lose modeling capability. To model the interaction between k elements specific fuzzy measures were proposed: k-additive [6] and k-maxitive [15, 16] ones.

Beyond computational complexity, semantics also argues for simplified fuzzy measures. As the number of interacting elements in real decision making problems is limited, one should wonder if the complete fuzzy measure identification makes sense. The answer should take into account the problem data cardinality: all coefficients may be needed for a reduced number of elements, e.g. n=3 elements, but when this number gets average or high, e.g. n=30, the complete fuzzy measure becomes meaningless. Is it really useful to assign a specific weight to each of (n-1)-size coalitions? Modeling k-order interaction meets the needs of both complexity and semantic.

The goal of this paper is to study the potential of k-order fuzzy measures and their use in a supervised learning process for classification. First, the case of full interaction is analyzed. In such extreme situation, the fuzzy measure is both k-additive and k-maxitive. The second objective of this work is the proposal of a k-maxitive measure learning algorithm based on $HLMS$ [4]. It is first used on synthetic data, to assess the k-maxitive measure ability for modeling partial, and more realistic, interactions, either redundancy or complementariness. Then, the learning algorithm is included within a pipeline that starts the learning process from raw data. This allows for managing real world data. Two well known datasets are used for illustrating the characterization the fuzzy measure is likely to provide in the process of feature selection (semantics) and for comparing complete and k-maxitive measures (complexity).

The outline of the paper is as follows: Section 2 introduces basic concepts related to fuzzy measures. In Section 3 specific measures to model k-order interactions, presented in the literature, are analyzed. The relationship between the complete fuzzy measure and the k-order ones is formalized in the case of extreme situation of full interaction. The learning algorithm is described in Section 4. The numerical experiments are carried out in Section 5 with synthetic and real world datasets. Finally, Section 6 summarizes the main conclusions and perspectives.

2. Preliminaries

This section introduces basic concepts related to fuzzy measures, discrete Choquet integral and the generalized interaction index [6]. Let us consider a finite set $N=\{1,\ldots,n\}$ and let $\mathcal{P}(N)$ denotes its power set. In this paper, a set is noted by a letter in uppercase and its cardinality with the same letter in lowercase, $a = |A|$.

2.1. Fuzzy measures and the discrete Choquet integral

A fuzzy measure (FM) is a set function $\mu\colon \mathcal{P}(N) \to [0,1]$ fulfilling the following two axioms [9]:

1. Normalization: $\mu(\emptyset) = 0$, $\mu(N) = 1$
2. Monotonicity: $A \subseteq B \subseteq N \Rightarrow \mu(A) \leq \mu(B)$

While the former allows for fuzzy measure comparisons, the latter ensures that adding any element to a given subset does not make it less informative.

Fuzzy measures are used in the definition of the discrete Choquet integral aggregation operator. For a given $f\colon N \to \Re^+$, its discrete Choquet integral $\mathcal{C}$ with respect to a fuzzy measure $\mu\colon \mathcal{P}(N) \to [0,1]$ is defined as follows:

$$\mathcal{C}_\mu(f) \triangleq \sum_{i=1}^{n} \big(f_{(i)} - f_{(i-1)}\big)\mu(\{i,\ldots,n\}) \tag{1}$$

where $f_{(\cdot)}$ is the rearrangement induced by f_i, $i = 1,\ldots,n$, sorted in ascending order, i.e., $f_{(1)} < \cdots < f_{(n)}$, by convention $f_{(0)} = 0$.

2.2. Semantic interpretation of fuzzy measure coefficients

Three kinds of interaction between two elements were defined in [5] according to the relationship between coefficients of singletons and pair of elements:

Redundancy The coefficient value associated with $\{i,j\}$ is almost the same as the individual value for each element, i.e., $\mu(\{i,j\}) < \mu(\{i\}) + \mu(\{j\})$. This kind of interaction is also called negative synergy.

Complementariness The coefficient value associated with $\{i,j\}$ is large, although these elements have small values if they are considered separately, i.e., $\mu(\{i,j\}) > \mu(\{i\}) + \mu(\{j\})$. This kind of interaction is also called positive synergy.

Independence The coefficient value associated with $\{i,j\}$ is equal to the sum of their individual values, i.e., $\mu(\{i,j\}) = \mu(\{i\}) + \mu(\{j\})$.

2.3. Interaction index

In the field of cooperative game theory, the Shapley index can be used to characterize the importance of individual features [22]:

$$\phi_i = \sum_{K \subseteq N \setminus i} \frac{(n-k-1)!k!}{n!} \big(\mu(K \cup \{i\}) - \mu(K)\big) \tag{2}$$

where 0!=1 as usual. The Shapley value of μ is the vector $\phi = [\phi_1 \cdots \phi_n]$ which has the property to be linear with respect to μ, and to satisfy:

$$\sum_{i=1}^{n} \phi_i = \mu(N) = 1 \tag{3}$$

This index has been generalized, first to characterize the importance of pairs [21] and finally for subsets A of arbitrary cardinality [6]:

$$I(A) = \sum_{K \subseteq N \setminus A} \frac{(n-k-a)!k!}{(n-a+1)!} \sum_{B \subseteq A} (-1)^{a-b} \mu(K \cup B) \tag{4}$$

$I(A)$ reduces to Shapley index when A is a singleton. The Shapley index ranges in $[0, 1]$ and the interaction index for pairs in $[-1, 1]$.

When there is a subset $R \subseteq N$, of r fully redundant elements, while the others only bring noise, then for any $A \subseteq R$, of size a, it was mathematically proven [19] that the interaction index becomes:

$$I(A) = \frac{(-1)^{a+1}}{r-a+1} \tag{5}$$

Similarly, when there is a subset $C \subseteq N$ of c fully complementary elements, for any $A \subseteq C$, it is:

$$I(A) = \frac{1}{c-a+1} \tag{6}$$

2.4. Möbius transform

A Möbius transform of a fuzzy measure μ is a set function m on N defined by [25]:

$$m(T) = \sum_{K \subseteq T} (-1)^{t-k} \mu(K), \ \ \forall T \subseteq N \tag{7}$$

The interaction index and the Möbius transform provide alternative representations of a fuzzy measure. There is a one to one correspondence between these three spaces [6].

The fuzzy measure coefficients are computed from the Möbius representation using the the Zeta-transform:

$$\mu(T) = \sum_{S \subseteq T} m(S) \ \ \forall T \subseteq N \tag{8}$$

3. Modeling k-order interaction

In real world data the number of interacting elements is limited. This section recalls the ways of modeling such interactions and propose new results in the particular case of full interaction, either redundancy or complementariness.

3.1. k-order measures

Grabisch [6] first introduced the k-order additivity concept for discrete spaces, then it was generalized to arbitrary measurable spaces by Mesiar [15].

Definition 1 A fuzzy measure μ is said to be k-additive if its Möbius transform satisfies $m(S) = 0$ for any S such that $s > k$ and there exists at least one subset $S \subset N$ of exactly k elements such that $m(S) \neq 0$.

The particular case of 2-additive FM has been especially studied. The necessary and sufficient conditions for a fuzzy measure to be 2-additive are [14]:

$$\begin{gathered} \mu_i \geq 0, \forall i \in N \\ \sum_{\{i,j\} \subseteq N} \mu_{i,j} - (n-2) \sum_{i \in N} \mu_i = 1 \\ \sum_{i \in A \setminus \{k\}} (\mu_{ik} - \mu_i) \geq (a-2)\mu_k, \forall A \subseteq N, a \geq 2, \forall k \in A. \end{gathered} \tag{9}$$

where $\mu_i = \mu(\{i\})$ and $\mu_{ij} = \mu(\{i,j\})$.

That means that a 2-additive fuzzy measure is entirely determined by the coefficients of singletons and pairs of elements.

The Choquet integral of a 2-additive fuzzy measure can be easily computed from the Shapley indices of singletons and the interaction indices of pairs:

$$C_\mu(x) = \sum_{i=1}^{n} I_i x_i - \sum_{\{i,j\} \subseteq N} I_{ij} |x_i - x_j|$$

The first idea to reach k-additivity is to truncate the fuzzy measure in the Möbius space, by setting to zero values associated with higher than k cardinality sets. Unfortunately, this truncation does not yield a fuzzy measure in the general case: the above mentioned constraints are usually unsatisfied, monotonicity and normality may be lost [3].

In [18], these constraints are included in a fuzzy measure learning algorithm for singletons and pairs to deal with feature selection in a classification context. A linear programming optimization, also for a 2-additive measures, is proposed in [13]. Both works show the extension to higher values of k is not straightforward.

In [14], a 2-additive Choquet integral is used for cardinal information representation by the means of *cyclones*. According to the authors, "in practice cyclones are not easy to detect". They then consider MOPI (Monotonicity of Preferential Information) conditions, but unfortunately "the final number of necessary and sufficient conditions could be very large".

Despite its interest and stimulating studies, the practical application of k-additive measure is still limited by the lack of learning algorithms.

Inspired from the k-additive measure, the k-maxitive fuzzy measure was proposed [2, 17]. It is based on an alternative Möbius transform called possibilistic Möbius transform. It can be seen as another framework for modeling k-order interaction in the coefficient space.

Definition 2 The possibilistic Möbius transform of a fuzzy measure μ on N is a $\mathcal{P}(N) \to [0,1]$ mapping m_p defined by:

$$m_p(A) = \begin{cases} \mu(A) & \text{if } \mu(A) > \max_{B \subset A} \mu(B) \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

The possibilistic Zeta transform of $\mathcal{P}(N) \to [0,1]$ mapping m is the $\mathcal{P}(N) \to [0,1]$ mapping $\mathcal{Z}_m$ defined by:

$$\mathcal{Z}_{m_p}(A) = \max_{B \subseteq A} m_p(B) \tag{11}$$

Under some conditions [2], $\mathcal{Z}_{m_p} = \mu$.

The k-maxitive fuzzy measure is defined analogously to the k-additive one but using the possibilistic Möbius transform.

Definition 3 A fuzzy measure μ is called k-maxitive if its possibilistic Möbius transform satisfies $m_p(S) = 0$ for any S such that $s > k$ and there exists at least one subset S of N of exactly k elements such that $m_p(S) \neq 0$.

The constraints the fuzzy measure coefficients must satisfy for being k-maxitive are easier to check than in the case of k-additivity.

A fuzzy measure where the coefficients of coalitions with more than k elements are computed as the maximum of the ones of included subsets with size up to k, is a k-maxitive fuzzy measure. The underlying semantics is not to give a specific weight to subsets of size higher than k. This is a way of modeling a k-order interaction in the coefficient space.

3.2. The particular case of full interaction

In some extreme situations, the fuzzy measure is k-additive and k-maxitive.

Theorem 1 A fuzzy measure that characterizes a set N of n elements among which k of them are fully complementary or redundant and the others only bring noise is k-additive.

Proof The proof considers separately the cases of complementariness and redundancy.

Complementary elements Let us consider the case of a set $C \subseteq N$ of c fully complementary informative elements while the other elements only bring noise. As we know from previous results [19], $\mu(S) = 1 \iff C \subseteq S$ and $\mu(S) = 0$ otherwise.

The Möbius values which may be different from zero are those of sets which include at least one subset with a non null fuzzy measure coefficient, meaning the sets $S \supseteq C$.

For the set C, the transform reduces to: $m(C) = (-1)^0 = 1$, as the fuzzy measure coefficient is zero for all $A \subset C$.

$\forall T \supset C,\ m(T) = \sum_{i=0}^{t-c} \binom{t-c}{i}(-1)^{t-c-i}$

This value is zero, thanks to the binomial formula:

$$m(T) = \sum_{i=0}^{t-c} \binom{t-c}{i} (-1)^{t-c-i} (+1)^i = (-1+1)^{t-c} = 0$$

.

The two conditions for a measure to be k-additive are fulfilled. So, in this extreme situation of c fully complementary elements, the fuzzy measure is k-additive.

Redundant elements The set of elements includes now a subset $R \subseteq N$ of r redundant informative elements while the rest of the elements do not bring any useful information. We know from previous results [19] that $\mu(S) = 1 \iff R \cap S \neq \emptyset$ and zero otherwise.

Let us check the two conditions to be met for the fuzzy measure to be k-additive, with k=r. First at least one subset of cardinality r should have a non null Möbius value. This is true for the set R:

$$m(R) = \sum_{i=1}^{r} \binom{r}{i} (-1)^{r-i}$$

The sum starts now at $r = 1$, as for $r = 0$ the corresponding set is the empty set, for which fuzzy measure coefficient is zero. $m(R) = \pm 1$ according to r value, negative when r is even.

The second condition states: $\forall T,\ t > r,\ m(T) = 0$.

If $t > r$, that means T includes at least one element which does not belong to R. Any T can be defined as an union: $T = \{L \subseteq R\} \cup \{M \subseteq (N \setminus R)\}$, with the cardinalities: $t = l + m$. For such a given T, the expression of its Möbius transform becomes:

$$m(T) = \sum_{p=1}^{l} \left[\binom{l}{p} \sum_{q=0}^{m} \binom{m}{q} (-1)^{t-p-q} \right]$$

Thanks to the binomial theorem, this value is zero for all sets T.

This proves that, in this extreme situation of r fully redundant elements, the fuzzy measure is k-additive.

These extreme situations, where the fuzzy measure coefficients take values in $\{0, 1\}$, show how meaningful the concept of k-additivity is: when the number of interacting elements is limited to k, the fuzzy measure is k-additive under the condition the k value is carefully chosen. Unfortunately k-additive fuzzy measures are not so easy to generate and only restricted attempts for k=2 have been reported.

k-maxitive measures are considered as a potential alternative for k-interaction modeling.

Theorem 2 A fuzzy measure that characterizes a set N of n elements among which k of them are fully complementary or redundant and the others only bring noise is k-maxitive.

Proof The proof is trivial. The maximum value for a coefficient, 1, is reached, in both cases of full redundancy and complementariness, at the k level.

When there are k fully interacting elements, the fuzzy measure is both k-additive and k-maxitive. As the k-maxitivity can be managed in the coefficient space, this kind of measure is easier to design, whatever the k value, than the k-additive one.

4. k-$HLMS$ algorithm

In order to learn k-maxitive measures from data, a new supervised algorithm based on $HLMS$ [4, 20], called k-$HLMS$ (Algorithm 1), is presented.

The algorithm input is a training dataset D composed of m samples described by n features and a reference target. The dataset is organized as follows:

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

In this matrix, a column represents a feature, and a row a sample, $x^j = x_1^j, \dots, x_n^j$. Each element x_i^j represents the satisfaction degree of the feature i for sample j. T^j is the output value (target) to infer from the satisfaction degrees. In this way, the partial information provided by each feature is integrated to get a global result. The data, both the description and the target, must be commensurable, i.e. ranging in the same scale and having the same meaning. In the case of classification, the data can be degrees of evidence for the sample to belong to a given class.

The aggregation operator used to integrate the information provided by each feature is the Choquet integral. It enables the consideration of underlying interactions among features. The goal of the algorithm is to learn the fuzzy measure coefficients that best reproduce the target from the description.

In k-$HLMS$, only the coefficients for coalitions that includes at most k elements are learned. For all sets $A \subset N$, $k < a < n$, the coefficients are computed as:

$$\mu(A) = \max_{\substack{L \in \mathcal{P}(A) \\ l=k}} \mu(L) \tag{12}$$

Finally, to satisfy the normalization axiom, $\mu(N)$ is set to 1.

A given sample always uses the same coefficients to compute the Choquet integral, one for each subset size between 1 and n-1. The coefficients which are used by only a few samples are identified as *untouched coefficients* and are not used during the training process (Line 3). The threshold is set to $max(3, m/100)$. It ensures that the coefficient values are supported by a significant number of samples.

Initially, all the fuzzy measure coefficients to be learned are initialized to the equilibrium state, $|i|/n$ for a i-size coalition (Line 4-6), and stored in the $\mathbf{u^k}$ vector.

At each iteration, samples are randomly sorted to prevent any bias related to presentation order (Line 8). To compute the Choquet integral (Line 11), only the coefficients required by the current sample are defined according to Eq. (12). This is done by the *UncutFm* function (Lines 10 and 19). Thus, there is no need to store the whole coefficient set.

The coefficients associated with sets of cardinality up to k involved in the corresponding integral are updated (Line 13) according to the learning rate, $\alpha > 0$, and the difference, for the current sample, between the Choquet integral and the target. In this formula, u_l stands for the coefficient of the l-size set triggered by the current sample.

Monotonicity check, is done with neighbors up to level k (Line 14).

The stop criterion can be based upon the root mean square of errors (E) convergence (Lines 17-22), or, plainly, on a predefined number of iterations. In the latter, there is no need to compute E.

Algorithm 1 k-$HLMS$

Input: Training data D, m samples, (x^j, T^j), $j = \{1, \dots, m\}$, of $N = \{1, \dots, n\}$ features
  Max order: $1 < k < n$
Output: Fuzzy measure coefficients up to level k, $\mathbf{u^k}$
$Z \leftarrow$ Identify_Untouched(D) {Untouched coefficients identification}
for $i \in \mathcal{P}(N) \setminus Z$ and $|i| \leq k$ **do**
  $\mathbf{u}_{\{i\}} \leftarrow |i|/n$ {Initialization of usable coefficients up to k}
end for
repeat
  examples $\leftarrow$ random$(1 : m)$ {Sensitivity to data presentation order}
  for $j \in$ examples **do**
    $\mathbf{u} \leftarrow UncutFm(\mathbf{u^k}, x^j)$ {Complete coefficients using Eq.(12)}
    $e^j \leftarrow \mathcal{C}_{\mathbf{u}}(x^j) - T^j$ {Individual error calculation}
    for $l \in (1 : k)$ **do**
      $u_l \leftarrow u_l - \alpha \times \frac{e^j}{e_{max}} \times (x^j_{(n-l+1)} - x^j_{(n-l)})$ {Coefficient update up to level k}
      $CheckMonoUptoK(u_l, k)$ {Monotonicity check, up to level k, with neighbors }
    end for
  end for
  $E \leftarrow 0$ {Global error calculation}
  for $j \in$ examples **do**
    $\mathbf{u} \leftarrow UncutFm(\mathbf{u^k}, x^j)$
    $E \leftarrow E + (\mathcal{C}_{\mathbf{u}}(x^j) - T^j)^2$
  end for
  $E \leftarrow \sqrt{\frac{1}{m}E}$
until Stop_Criterion is met
return($\mathbf{u^k}$)

5. Numerical experiments

In this section, modeling ability of k-maxitive measure is evaluated considering two data scenarios: i) synthetic data, with partially redundant and complementary set of features, and ii) real benchmark datasets. In both cases, the k-maxitive behavior is compared to the complete fuzzy measure whose coefficients are identified through (n-1)-$HLMS$ algorithm (since $\mu(N)$=1). The k-$HLMS$ algorithm parameters are set as follows: learning rate α=0.05 and stop after 3000 iterations.

5.1. k-maxitive behavior with synthetic data

In order to study the k-maxitive modeling ability, synthetic data with partially redundant or complementary set of features are analyzed. Datasets have 7 features $\{1, 2, \dots, 7\}$ and 440 samples, target value T=$\{Class0, Class1\}$, row values represent confidence degrees that the associated sample belongs to Class1. Values of k in the range $[2, 6]$ were tested, where k=6 corresponds to the complete fuzzy measure.

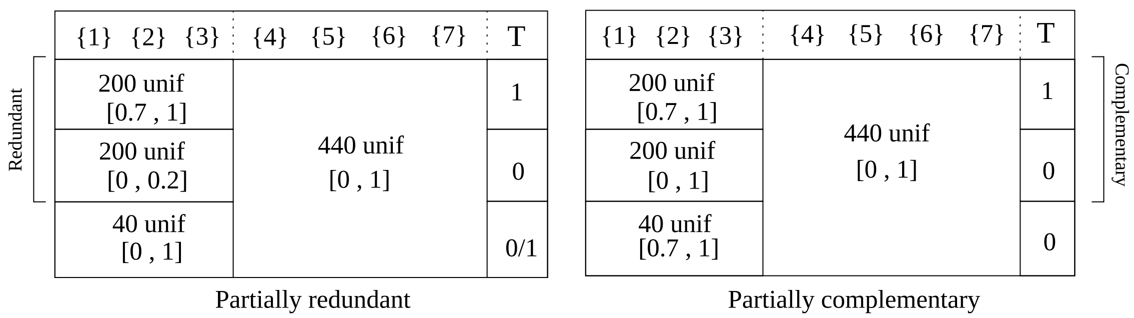

Fig. 1: Synthetic data design

Partially redundant set of features. The partial redundancy among $\{1, 2, 3\}$ features is achieved by 400 samples fully redundant, i.e., $\{1, 2, 3\}$ bringing the same information, and 40 samples bringing noise, as shown in Fig. 1. For features 1 to 3 random values were generated from uniform distributions in non-overlapping intervals to get full redundancy: 200 samples for Class1 in $[0.7, 1]$; 200 samples for Class0 in $[0, 0.2]$. The 40 noisy samples were generated in $[0, 1]$ with classes chosen at random. Features 4 to 7 bring noise, i.e., 440 samples with uniformly distributed values in $[0, 1]$.

The most relevant feature belongs to $\{1,\ 2,\ 3\}$ for most samples. Thus, coefficient values associated with singletons get high values as well as the ones associated with subsets of $\{1, 2, 3\}$ due to fuzzy measures monotonicity[19]. Consequently, we expect good results even with k=2 simplification, meaning the I values should be significant for all subsets $L \subseteq \{1, 2, 3\}$ and zero for the rest of coefficients. In addition, I signs should alternate according to Eq.(5).

Table 1 displays the interaction index values for sets $L \subseteq \{1, 2, 3\}$, computed using Eq.(4). Only significant values, $|I|>0.1$, are shown. In order to compensate the lack of contribution of high order sets in the Choquet integral (Line 11 of k-$HLMS$) coefficients up to k may be overestimated. This is likely to affect the interaction index calculation.

Coalition	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}	$RMSE_I$
k=2	0.390	0.282	0.292	-0.499	-0.314	-0.196	0.426	0.13
k=3	0.428	0.260	0.281	-0.376	-0.302	-0.285	0.417	0.13
k=4	0.410	0.265	0.297	-0.410	-0.288	-0.275	0.371	0.12
k=5	0.412	0.256	0.320	-0.411	-0.216	-0.238	0.400	0.11
k=6	0.393	0.286	0.313	-0.317	-0.213	-0.213	0.685	0

Table 1: I of coalitions $L \subseteq \{1, 2, 3\}$ for partially redundant dataset.

Table 1 shows that all k approximations model the partial redundancy among the three features: significant values are obtained for subsets in $\{1, 2, 3\}$ and their signs are negative for even cardinality sets and positive for odd ones. Finally, the last column ($RMSE_I$) is the root mean squared error between each approximation (k-row) and the complete fuzzy measure (last row). As expected, higher values of k approximates better the complete fuzzy measure. In addition, Table 2 shows the Möbius values. For any k, the only significant coefficients are the ones associated with subsets in $\{1, 2, 3\}$ confirming the three order interaction.

Coalition	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}	$RMSE_m$
k=2	0.850	0.613	0.591	-0.596	-0.441	-0.245	0.228	0.21
k=3	0.822	0.649	0.619	-0.634	-0.443	-0.435	0.422	0.11
k=4	0.815	0.651	0.627	-0.651	-0.445	-0.464	0.467	0.09
k=5	0.783	0.640	0.607	-0.611	-0.395	-0.438	0.513	0.09
k=6	0.761	0.663	0.642	-0.648	-0.542	-0.552	0.667	0

Table 2: Möbius coefficients (m) of coalitions $L \subseteq \{1, 2, 3\}$ for partially redundant dataset.

We conclude that for partially redundant sets of features, k-maxitive measures provide a good approximation to the complete fuzzy measure for any k.

Partially complementary set of features. The partial complementariness among $\{1, 2, 3\}$ is achieved by 400 samples for which the features are fully complementary, and 40 noisy samples. For features 1 to 3 random values were generated from uniform distributions in overlapping intervals to get full complementariness: 200 samples for Class1 in $[0.7, 1]$; 200 samples for Class0 in $[0, 1]$. The partialness is given by 40 samples in the range used for Class1, $[0.7, 1]$, with a Class0 label. In this way, features 1 to 3 bring simultaneous wrong support to the classification process. Features 4 to 7 bring noise, i.e., 440 samples with uniformly distributed values in $[0, 1]$.

For a set C of complementary features, only fuzzy measure coefficients of coalitions that include C are non-null (see [19]). When $c = 3$, 2-maxitive simplification should be wrong since coefficients of order higher than 2 should be set from uninformative values, according to Eq.(12). The interaction index values should be significant and positive for all subsets $L \subseteq \{1, 2, 3\}$ when $k \geq 3$, and zero for remaining coalitions.

The results summarized in Table 3 show that the I values are consistent with the expected behavior: for $k \geq 3$, I of $L \subseteq \{1, 2, 3\}$ are positive, then k-maxitive measure is able to model the partial complementariness

Coalition	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}	$RMSE_I$
k=2	0.429	**0.095**	0.292	**0.193**	0.585	**-0.093**	**-0.224**	0.49
k=3	0.330	0.325	0.330	0.485	0.496	0.484	0.965	0.03
k=4	0.328	0.328	0.329	0.483	0.490	0.485	0.946	0.02
k=5	0.325	0.319	0.322	0.473	0.479	0.473	0.925	0.01
k=6	0.314	0.313	0.314	0.456	0.462	0.458	0.894	0

Table 3: I of coalitions $L \subseteq \{1, 2, 3\}$ for partially complementary dataset.

getting better with increasing k. See $RMSE_I$ column. For k=2, the corresponding values are inconsistent. The negative value for $\{2, 3\}$ could be interpreted as redundancy between both features, but in this case the $\{1, 2, 3\}$ index value would have been positive. Consequently, the approximation for $k < c$ does not work.

The Möbius transform of the fuzzy measure was also calculated (Table 4). Values for $k \geq 3$ should be significant only for coalition $\{1, 2, 3\}$ stating that the three elements are complementary. However, for k=2 significant values are observed for other subsets included in $\{1, 2, 3\}$ due to the inconsistencies in 2-maxitive values.

Coalition	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}	$RMSE_m$
k=2	0.007	0.003	0.004	**0.291**	**0.676**	0.095	**-0.294**	0.51
k=3	0.008	0.005	0.002	0.001	0.000	0.002	0.959	0.04
k=4	0.001	0.002	0.003	0.008	0.001	0.001	0.867	0.01
k=5	0.004	0.000	0.003	0.002	0.001	0.002	0.855	0.00
k=6	0.006	0.001	0.001	0.001	0.002	0.002	0.839	0

Table 4: Möbius of coalitions $L \subseteq \{1, 2, 3\}$ for partially complementary dataset.

The analysis of both experiments suggests that if the maximum cardinality of simultaneously interacting features is known, or could be in some way estimated, the use of k-maxitive measure preserves the characterization ability of complete fuzzy measures while reducing the complexity. It was shown in section 3 that for full interaction, k-maxitive are identical to the complete fuzzy measure, providing a perfect characterization. This study shows that the difference between partial and full interaction is a matter of degree, not of kind.

5.2. Application to benchmark data

Classification, meaning assigning a class label to a sample on the basis of its description, is one the most popular task in information processing. When the number of features gets average or high, classifier design is achieved by means of learning algorithms. Many of them are available for supervised learning. They generally include a feature selection step which is usually based on individual evaluations [10] assuming a rarely met condition of feature independence [7, 12]. Taking interaction into account may improve the final accuracy. This can be done either in the feature selection step [4, 5, 18] or in classifier combination [11, 1].

The framework for classifier design, presented in [19], and illustrated in Fig. 2, can be used with k-maxitive measures. It designs and uses a fuzzy measure for each of the possible classes.

The process starts with the raw data. The first component aims to convert feature values into commensurable confidence degrees. For each sample the relevance of an individual feature value for the identification of a given class is taken from the probability density distribution of feature values within the class. The higher the transformed feature value, the higher the evidence provided by the feature that the sample belongs to the given class. The Gaussian densities are designed in the training stage and used in the testing one.

Then, for each class, the k-maxitive measure is used to integrate the evidence provided by each feature for the considered class. The training step consists in learning the fuzzy measure coefficients using the k-$HLMS$ algorithm. The resulting fuzzy measures are used in the testing phase to classify new observations.

The class label is the one for which the global evidence, computed using the Choquet integral with respect to the fuzzy measure, is the highest.

The real world data used in this section are from the UCI repository[1]. They have been chosen because their

[1] https://archive.ics.uci.edu/ml/datasets.html

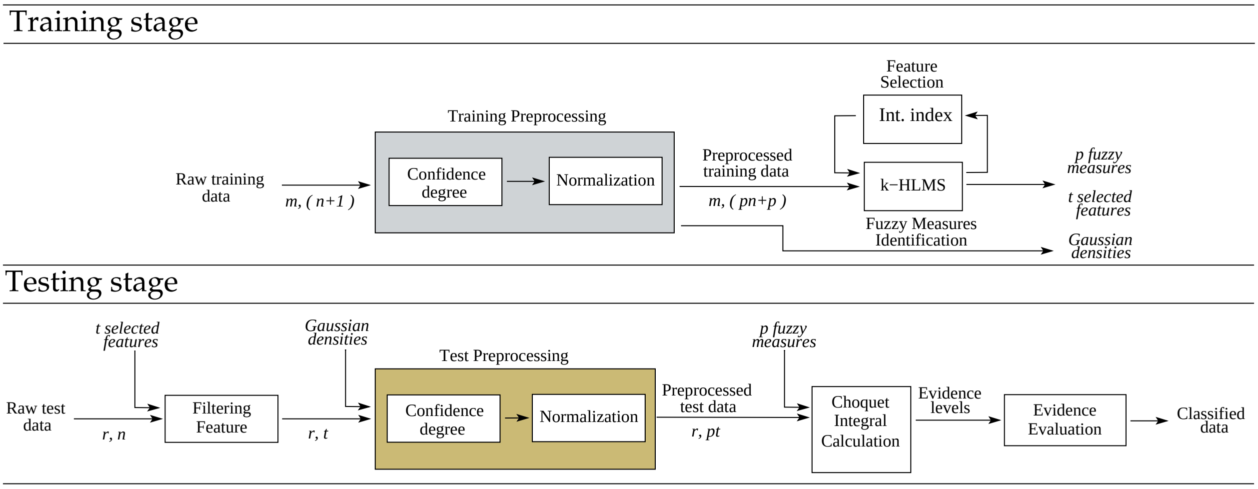

Fig. 2: Classification framework based on [19]. It presents two stages: Training and Testing. Rectangles represent components, arrows show data flow direction. The letter meaning is the following: n is the number of features; p is the number of classes and t is the number of selected features; m is the number of samples for the training stage; r is the number of samples for the testing stage

number of features is high enough to illustrate that the number of interacting elements is limited, and also small enough to allow the complete fuzzy measure evaluation. Their main characteristics are summarized in Table 5.

Datasets	#Features	#Samples	#Classes
Breast Cancer	9	683	2
Wine	13	178	3

Table 5: Benchmark data characteristics

The comparison between the k-maxitive and the complete fuzzy measure considers four aspects:

- #Coeff: Number of identified coefficients calculated as $\sum_{j=1}^{k} \binom{n}{j} - Untouched(up\ to\ k)$;
- $\sum$ Coeff: Sum over all classes of the coefficients up to level k;
- $RMSE_I$: Root mean squared error for I values between the k-maxitive and the complete fuzzy measure up to level k;
- Error: Classification error rate.

To analyze the interaction index representation, results were filtered using an absolute value threshold of a/n for a-order sets. In the Möbius space, bounds are not always symmetric [8]. The positive and negative ranges are thus considered separately. Each of them is divided into four equal intervals, corresponding to the four linguistic interaction levels: *null*, *poor*, *medium* or *high*. Only the last three are taken into account. For instance, for 3 features the range is $[-2, 1]$ and only values lower than -0.5 or higher than 0.25 are considered as relevant.

5.2.1. Breast Cancer

For 9 features, a whole fuzzy measure definition requires 2^9=512 coefficients. The #Coeff row, in Table 6, shows that most of them are not used by the dataset and are labeled as untouched coefficients (Line 3, Algorithm 1). The complete fuzzy measure, 8-$HLMS$, needs only 77 coefficients to be learned with 683 samples. This is likely to make their estimation more robust.

The four rows $\sum$ Coeff show that k-maxitive coefficients up to k are overestimated with respect to the complete fuzzy measure, k=8. The difference gets higher when k decreases. This is expected since the restrictions imposed by k-maxitive measure affect a bigger number of coefficients for smaller k values. As the interaction indices are computed from coefficients, their approximation becomes better with increasing k values. This is clearly stated by the $RMSE_I$ row. This trend of I with respect to k-maxitive approximation can be observed in Fig. 3. The plots show the $\{2, 7\}$-maxitive approximation (dashed line) against the complete fuzzy measure (solid line). The coalitions, in lexical order for increasing cardinality, are in the abscissa, and the corresponding interaction indices in the ordinate. The figure shows the general shape of the complete fuzzy measure is preserved when $\{2, 7\}$-maxitive measure is used.

The last row of the table shows that the use of $\{2, 3, 5, 7\}$-maxitive measures slightly impact the classification accuracy, which is around 98.5%. The difference between 2.91% and 2.18% in the test set represents only one sample misclassification.

	k=2	k=3	k=5	k=7	k=8
#Coeff	21	31	46	66	77
$\sum$ Coeff	3.2				1.6
$\sum$ Coeff		8.7			6.8
$\sum$ Coeff			23.8		21.9
$\sum$ Coeff				57.2	57.1
$RMSE_I$	0.16	0.10	0.09	0.08	0
Error	2.91%	2.91%	2.18%	2.18%	2.18%

Table 6: k-maxitive analysis for Breast Cancer dataset. Complete fuzzy measure is for k=8.

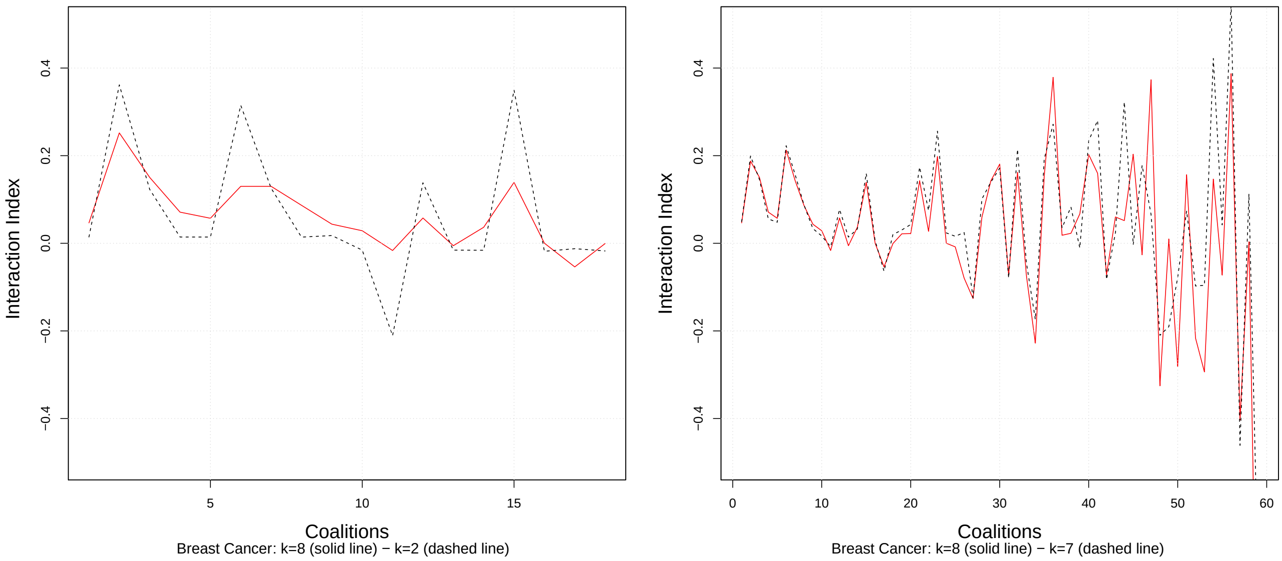

Fig. 3: Interaction index comparison for Breast Cancer. The $\{2, 7\}$-maxitive measure is used (dashed lines) against the complete fuzzy measure, k=8 (solid lines). The abscissa represents the coefficients ordered according to their cardinality and feature number, e.g., $\{1\}, \{2\}, \{3\}, \ldots, \{9\}, \{1, 2\}, \{1, 3\}, \ldots$

To characterize the interactions, Möbius and Interaction indices values are analyzed. The results for the complete fuzzy measure are given in Table 7. I results for features 2, 3, 6 and 7 suggest they are relevant to classification. Moreover, Möbius values of features 2 and 6 indicates that they are partially complementary for Class1 and the corresponding I values show that this coalition is relevant. Möbius values of feature set $\{1, 2, 3\}$ and $\{2, 3, 7\}$ show they interact; however, their corresponding I value indicates that their contribution is not significant. The overall analysis suggest that 3-maxitive measure may be a good approximation to the complete fuzzy measure for both classes. Approximation values using k=3 are shown in Table 8. These results highlight that the 3-maxitive measure yields the same conclusions about feature interactions than the complete fuzzy measure. As shown in Table 6, k=3 also provide a comparable classification accuracy. Consequently, the 3-maxitive measure is a good approximation of the complete fuzzy measure for Breast Cancer data.

	Class1		Class2	
Coalition	I	Möbius	I	Möbius
{2}	0.252		0.138	
{3}	0.152		0.288	
{6}	0.139		0.181	
{7}	0.117		0.112	
{2,3}	0.290		0.220	
{2,6}	0.220	0.374		
{3,6}				0.543
{1,2,3}		0.633		0.732
{2,3,7}		0.787		0.736

Table 7: I and Möbius relevant values for Breast Cancer with k=8.

	Class1		Class2	
Coalition	I	Möbius	I	Möbius
{2}	0.441		0.127	
{3}	0.152		0.324	
{6}	0.143		0.181	
{7}	0.137		0.120	
{2,3}	0.353		0.233	
{2,6}	0.432	0.621		
{3,6}				0.453
{1,2,3}		0.986		0.830
{2,3,7}		0.991		0.921

Table 8: I and Möbius relevant values for Breast Cancer with k=3.

5.2.2. Wine

The results for Wine dataset are shown in Table 9. The number of coefficients for a whole fuzzy measure for 13 features is 2^{13}=8192. With only 178 samples for their identification, what kind of support would have had the obtained results? The actual number of identified coefficients is 145 for the complete fuzzy measure and drops to 30 when k=2. Next rows show the same behavior described for Breast cancer dataset regarding coefficient overestimation and I values with respect to k, i.e., coefficients up to k are overestimated and I approximation improves as k increases. In Fig. 4 Class1 values of I are shown. There is only one relevant peak corresponding to the value of feature 13 while the general shape is preserved for $\{2, 9\}$-maxitive measure approximation.

The classification error is around 7%. The error rate oscillates between 5.73% and 8.53%: this variation represents only one sample in the test set.

	k=2	k=3	k=5	k=7	k=9	k=12
#Coeff	30	43	63	81	103	145
$\sum$ Coeff	16.2					14.0
$\sum$ Coeff		28.2				25.3
$\sum$ Coeff			44.3			41.6
$\sum$ Coeff				53.4		53.4
$\sum$ Coeff					70.5	70.2
$RMSE_I$	0.08	0.07	0.07	0.07	0.05	0
Error	5.73%	8.53%	5.73%	8.53%	5.73%	5,73%

Table 9: k-maxitive analysis for Wine dataset. Complete fuzzy measure is for k=12.

The I and Möbius values for the complete fuzzy measure (Table 10) show that for Class1 there is no interaction among features: only feature 13 is relevant. Class2 is more interesting to analyze: features 1, 7 and 10 are relevant according to their interaction index. Negative values for coalitions $\{1, 10\}$ and $\{7, 10\}$ indicate redundancy while the positive value for coalition $\{1, 7\}$ indicates complementariness. This suggests that there are two complementary features, 1 and 7, and a third one redundant with both of them, 10. The interaction between these three features can also be analyzed in the Möbius space values associated with coalition $\{1, 7, 10\}$. Möbius coefficient associated with $\{1, 7\}$ suggests that the complementary interaction is more relevant than the redundancy. These interaction indices can be compared to the corresponding extreme situation of a fully complementary pair of features and a third one fully redundant with the pair. The fuzzy measure coefficients are reported in the first row of Table 11 and the interaction indices in the second row. They show that the difference between the experimental and the extreme cases is a matter of degree. For Class3, an analogous analysis could be made for features 7, 11 and 12: features 11 and 12 are complementary and feature 7 is redundant with them. These results state that for Class2 and Class3 the fuzzy measure might be 3-additive.

Table 12 shows that the same analysis can be carried out from 3-maxitive Möbius coefficients and interaction indices. The 3-maxitive approximation also yields comparable results in classification as displayed in Table 9: it is a good approximation of the complete fuzzy measure.

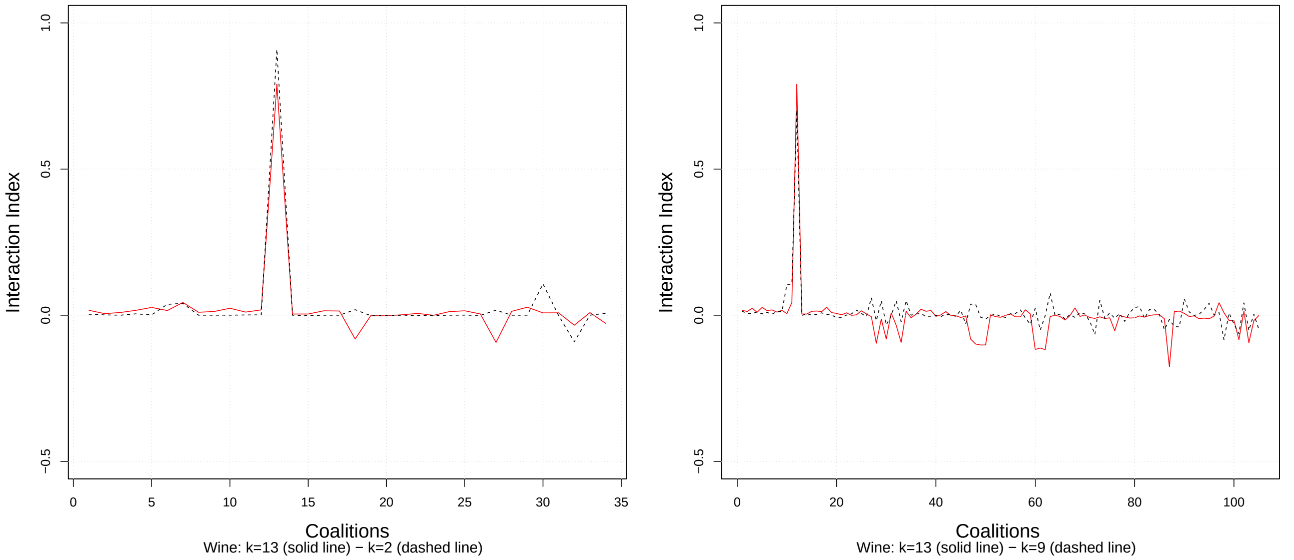

Fig. 4: Interaction index comparison for Wine (Class 1). The $\{2, 9\}$-maxitive measure is used (dashed lines) against the complete fuzzy measure, k=12 (solid lines). The abscissa represents the coefficients ordered according to their cardinality and feature number, E.g., $\{1\}, \{2\}, \{3\}, \ldots, \{13\}, \{1, 2\}, \{1, 3\}, \ldots$.

	Class1		Class2		Class3	
Coalition	I	Möbius	I	Möbius	I	Möbius
{1}			0.17			
{7}			0.14	0.10	0.63	0.95
{10}			0.52	0.74		
{11}					0.14	
{12}					0.17	
{13}	0.80	0.98		0.21		
{1,7}			0.34	0.77		
{1,10}			-0.16			
{1,13}			-0.17			
{7,10}			-0.37			
{7,11}					-0.34	
{7,12}					-0.45	
{11,12}					0.36	0.84
{1,7,10}			-0.60	-0.77		
{1,10,13}						
{1,7,11,12}					-0.77	-0.88

Table 10: I and Möbius relevant values for Wine with k=12.

	{1}	{2}	{3}	{4}	{1,2}	{1,3}	{1,4}	{2,3}	{2,4}	{3,4}	{1,2,3}	{1,2,4}	{1,3,4}	{2,3,4}	{1,2,3,4}
FM	0	0	1	0	1	1	0	1	0	1	1	1	1	1	1
I	0.17	0.17	0.66	0	0.5	-0.5	0	-0.5	0	0	-1	0	0	0	1

Table 11: Coefficients and I values for 2 fully complementary features ($\{1\}$ and $\{2\}$), a third one fully redundant with them ($\{3\}$), and some noise ($\{4\}$).

6. Conclusions

This paper aimed to study the ability of k-order fuzzy measure to characterize and model k-order interactions in a classification context. k-order measures are likely to meet the needs of semantics, as the number of interacting elements in real world data is limited, and complexity, the number of coefficients to identify is drastically

	Class1		Class2		Class3	
Coalition	I	Möbius	I	Möbius	I	Möbius
{1}			0.15			
{7}			0.18	0.10	0.41	0.81
{10}			0.54	0.94		
{11}					0.28	
{12}					0.22	
{13}	0.77	0.99		0.35		
{1,7}			0.45	0.79		
{1,10}			-0.14			
{1,13}			-0.18			
{7,10}			-0.29			
{7,11}					-0.25	
{7,12}					-0.20	
{11,12}					0.34	0.90
{1,7,10}			-0.53	-0.83		
{1,10,13}						
{1,7,11,12 }					-0.61	-0.61

Table 12: I and Möbius relevant values for Wine with k=3.

reduced.

In extreme situations, where elements are fully redundant or complementary, the fuzzy measure coefficients take binary values. In this case, and when the k value is set to the number of interacting elements, it is mathematically proven that the complete fuzzy measure is both k-additive and k-maxitive.

To assess the behavior and characterization ability of k-maxitive fuzzy measures in more realistic situations, an algorithm, based on $HLMS$, is proposed to identify the measure from labeled training data. The coefficients of up to k-size coalitions are identified using the gradient descent approach while the others are set to the maximum value of all included subsets. That means the minimum allowed value that guarantees monotonicity. This way, no groundless information is added.

The algorithm is used with synthetic datasets for which the number of interacting elements is known and the level of interaction is controlled. This study shows that partial redundancy or complementariness is properly characterized by k-maxitive measures. Both the interaction indices and Möbius coefficients exhibit the expected behavior. This supports the idea that the difference between full and partial interaction is a matter of degree, not of kind.

To deal with real world data, the learning algorithm is included within a pipeline that starts the process from raw data and converts each feature value into a class support degree. Then, the fuzzy measure can be identified and used to achieve the classification task.

The complete fuzzy measure is compared to a k-maxitive one according to several aspects. It is first highlighted that coefficients up to k-size coalitions are overestimated as expected. This bias does not impact the characterization power of the k-maxitive measure as shown by a comparison of the interaction indices and the Möbius coefficients. Thanks to these indices, a semantic analysis about feature interaction is carried out for the two considered datasets. In both cases the number of interacting elements is three. Finally, the classification accuracy of the complete and the 3-maxitive fuzzy measures prove to be comparable.

The number of coefficients to identify is significantly reduced for a k-maxitive measure, making larger datasets tractable and estimation more robust. To work properly, the k-maxitive measure learning algorithm has to be run with an adequate value of k: it must be higher or equal to the number of interacting elements, especially when they are complementary. An interesting and open perspective is the automatic estimation of k.

References

[1] P. Bulacio, S. Guillaume, E. Tapia, and L. Magdalena. A selection approach for scalable fuzzy integral combination. *Information Fusion*, 11(2):208–213, 2010.

[2] T. Calvo and B. de Baets. Aggregation operators defined by k-order additive/maxitive fuzzy measures. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 06(06):533–550, 1998.

[3] A. Chateauneuf and J-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of mobius inversion. *Mathematical Social Sciences*, 17(3):263–283, June 1989.

[4] M. Grabisch. A new algorithm for identifying fuzzy measures and its application to pattern recognition. In *Fourth IEEE international conference on fuzzy systems*, pages 145–150, Yokohama, Japan, 1995.

[5] M. Grabisch. The representation of importance and interaction of features by fuzzy measures. *Pattern Recognition Letters*, 17(6):567 – 575, 1996.

[6] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92(2):167–189, 1997.

[7] M. Grabisch, J-L. Marichal, R. Mesiar, and E. Pap. Aggregation functions: Means. *Information Sciences*, 181(1):1 – 22, 2011.

[8] M. Grabisch and P. Miranda. Exact bounds of the möbius inverse of monotone set functions. *Discrete Applied Mathematics*, 186:7–12, 2015.

[9] M. Grabisch, H.T. Nguyen, and E.A. Walker. *Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference*. Kluwer Academic, 1995.

[10] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182, March 2003.

[11] L. I. Kuncheva. "fuzzy" versus "nonfuzzy" in combining classifiers designed by boosting. *IEEE Transactions on Fuzzy Systems*, 11(6):729–741, 2003.

[12] L. I. Kuncheva and C. J. Lakhmi. Nearest neighbor classifier: Simultaneous editing and feature selection. *Pattern Recognition Letters*, 20(11-13):1149–1156, 1999.

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

[14] B. Mayag, M. Grabisch, and C. Labreuche. A characterization of the 2-additive choquet integral through cardinal information. *Fuzzy Sets and Systems*, 184(1):84 – 105, 2011.

[15] R. Mesiar. Generalizations of k-order additive discrete fuzzy measures. *Fuzzy Sets Syst.*, 102(3):423–428, March 1999.

[16] R. Mesiar. k-order additive fuzzy measures. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 07(06):561–568, 1999.

[17] R. Mesiar. k-order additivity and maxitivity. *Atti del Seminario Matematico e Fisico dell'Universita di Modena*, 23(51):179–189, 2003.

[18] L. Mikenina and H.-J. Zimmermann. Improved feature selection and classification by the 2-additive fuzzy measure. *Fuzzy Sets and Systems*, 107(2):197 – 218, 1999.

[19] J. Murillo, S. Guillaume, F. Spetale, E. Tapia, and P. Bulacio. Set characterization-selection towards classification based on interaction index. *Fuzzy Sets and Systems*, 270(0):74 – 89, 2015.

[20] J. Murillo, S. Guillaume, E. Tapia, and P. Bulacio. Revised hlms: A useful algorithm for fuzzy measure identification. *Information Fusion*, 14(4):532–540, 2013.

[21] T. Murofushi and S. Soneda. Techniques for reading fuzzy measures (iii): interaction index. In *9th Fuzzy System Symposium*, pages 693–696, Sapporo, Japan, 1993.

[22] L. S. Shapley. A value for n-person games. In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the Theory of Games, vol II*, volume 28 of *Annals of Mathematics Studies*, pages 307–317. Princeton University Press, 1953.

[23] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[24] M. Sugeno and T. Terano. A model of learning based on fuzzy information. *Kybernetes*, 6:157–166, 1977.

[25] V. Torra and Y. Narukawa. *Modeling Decisions*. Springer, 2007.