k-maxitive Sugeno integrals as aggregation models for ordinal preferences

Quentin Brabant, Miguel Couceiro

k-maxitive Sugeno integrals as aggregation models for ordinal preferences (preprint version)

Quentin Brabant, Miguel Couceiro
LORIA (CNRS, Inira Nancy Grand Est - Université de Lorraine)
Équipe Orpailleur
54506 Vandœuvre-lès-Nancy, France
{quentin.brabant, miguel.couceiro}@inria.fr

Abstract

We consider an order variant of k-additivity, so-called k-maxitivity, and present an axiomatization of the class of k-maxitive Sugeno integrals over distributive lattices. To this goal, we characterize the class of lattice polynomial functions with degree at most k and show that k-maxitive Sugeno integrals coincide exactly with idempotent lattice polynomial functions whose degree is at most k. We also discuss the use of this parametrized notion in preference aggregation and learning. In particular, we address the question of determining optimal values of k through a case study on empirical data.
Keywords: Sugeno integral, k-maxitivity, lattice polynomial, degree, preference aggregation, preference learning.

1 Introduction

The Sugeno integral was introduced in [27] and it became a widely used aggregation function in the qualitative approach to decision making since it provides a meaningful procedure to fuse values within universes where no structure (other than an order) is considered [5, 10, 16]. Originally, the Sugeno integral was defined over real intervals but it can be extended to wider domains, namely, distributive lattices, via the notion of lattice polynomial function (i.e., a combination of variables and constants using the lattice operations $\wedge$ and $\vee$). In fact, idempotent lattice polynomial functions coincide exactly with (discrete) Sugeno integrals (see e.g. [6, 22]). The latter observation is particularly interesting in the context of multicriteria decision making as it provides a way of aggregating preferences that are not total orders. In fact, preference aggregation in the qualitative approach to preference modeling is the problem that motivates the current paper, and that we now briefly discuss.

We consider a multicriteria framework, where alternatives are described according to a set of criteria. We denote the set of alternatives by X and we denote the set of criteria indices by $[n] = \{1, \ldots, n\}$. Here, a criterion is a pair consisting of an attribute X_i together with a utility function $\varphi_i : X_i \to L$ modeling our preference on X_i. The set L is thought of as an *evaluation space*. In the qualitative approach that we consider, L is an ordered set, not necessarily numerical. In fact, throughout this paper we will assume that L is a bounded distributive lattice. Note that this does not constitute a serious restriction since any partial order can be embeded into a distributive lattice, such that the original order is preserved, by Dedekind-MacNeille's Completion (see, e.g, [9]).

Let $\varphi : X \to L^n$ be a mapping from alternatives to their evaluations on criteria. For $x \in X$ we write

$$\varphi(x) = (x_1, \ldots, x_n),$$

where x_i is the evaluation of x in the i^{th} criterion. By a *preference relation* we simply mean a binary relation $\preccurlyeq$ over X that is reflexive and transitive. Furthermore we represent such a preference relation $\preccurlyeq$ by a *utility function* $U : X \to L$ (for a suitable L) through the rule:

$$x \preccurlyeq y \Leftrightarrow U(x) \leq U(y)[3].$$

Note that the relation $\preccurlyeq$ thus defined is transitive, reflexive, but not necessarily anti-symmetric since $x \preccurlyeq y$ and $y \preccurlyeq x$ implies $U(x) = U(y)$ but not necessarily $x = y$. In this paper we focus on utility functions that can be defined as

$$U(x) = A(\varphi(x)) = A(x_1, \ldots, x_n),$$

where $A : L^n \to L$ is an aggregation function that we call the *preference aggregation model*. As L is a distributive lattice, the Sugeno integral is the most relevant class of functions for A [12]. Note that in the case when L is a numerical interval, the aggregation model of choice is the Choquet integral [16]. In some sense, the Sugeno integral can be viewed as the ordinal analogue of the Choquet integral.

In this paper we also consider a supervised learning problem, namely that of learning a Sugeno integral that models a set of examples $\mathscr{D} \subseteq L^n \times L$, where each element is a couple that associates an utility value to a tuple of n criteria values :

$$\mathscr{D} = \left\{ \left(\mathbf{x}^1, y^1\right), \ldots, (\mathbf{x}^m, y^m) \right\}.$$

Here, each y^j is the utility value of an alternative x^j for which we have $\mathbf{x}^j = (x_1^j, \ldots, x_n^j)$. Ideally, the result of the learning process would be a model consistent with $\mathscr{D}$, that is $A(x_1^j, \ldots, x_n^j) = y^j$ for each example in $\mathscr{D}$. However the learning set $\mathscr{D}$ is typically inconsistent, and thus the task is to learn a Sugeno integral that gives the least prediction error with respect to $\mathscr{D}$. This is a difficult optimization task, because of the intrinsic complexity of the Sugeno integral, that grows exponentially with the number of criteria.

To overcome this prohibitive complexity we consider k-maxitive Sugeno integrals, and study their properties for learning purposes, in particular, the gain in terms of complexity. Moreover we aim at finding the best value of k. To this end we provide a comparative study of k-maxitive models for different values of k on empirical data. Note that similar studies were presented for k-additive Choquet integrals; see, e.g,

Figure 1: The two forbidden substructures in a distributive lattice.

[20, 28]. We also provide a theoretical study of these k-maxitive Sugeno integrals, treated as sub-classes of lattice polynomials.

The paper is organized as follows. In Section 2, we recall basic background on lattice theory and lattice polynomials. The notion of k-maxitivity is investigated in Section 3, where we show how it is related to the notion of lattice polynomials with degree at most k. In fact, we give necessary and sufficient conditions for a lattice polynomial to have degree at most k and we show that k-maxitive Sugeno integrals coincide exactly with idempotent lattice polynomials with degree at most k (this result can be found in [3] presented at LFA 2015). In Section 4 we discuss the problem of learning a Sugeno integral and describe a method for solving it. In Section 5 we present a case study where we analyse how the accuracy of k-maxitive Sugeno integrals as predictive models changes according to different values of k. We discuss issues that are still to be resolved and indicate directions for future research in Section 6.

2 Preliminaries: Lattices, lattice polynomials and Sugeno integrals

In this section we recall some basic notions and results on lattice polynomials and lattice theory. For further background, see e.g. [2, 18, 26].

A *lattice* is an algebraic structure $\langle L, \wedge, \vee \rangle$ where L is a nonempty set, called *universe*, and where the two binary operations $\wedge$ and $\vee$, called *infimum* and *supremum* resp., satisfy the commutative, associative, absorption, and idempotent laws. We will denote a lattice by its universe L.

A lattice L is said to be *distributive* if, for every $a, b, c \in L$,

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \quad \text{or, equivalently,} \quad a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c).$$

Alternatively, a lattice L is distributive if and only if it does not embed any of the two sublattices shown in Figure 1 ; see, e.g., [9, 19].

Throughout this paper, L will denote an arbitrary bounded distributive lattice with least and greatest elements 0 and 1, respectively. For $a, b \in L$, $a \leqslant b$ simply means that $a \wedge b = a$ or, equivalently, $a \vee b = b$.

For $c \in L$ and $n > 1$, we shall denote tuples of the form $(c \ldots, c) \in L^n$ by $\mathbf{c}$. For

$\mathbf{x}, \mathbf{y} \in L^n$, we also make use of the following short-hand notations

$$\mathbf{x} \vee \mathbf{y} = (x_1 \vee y_1, \ldots, x_n \vee y_n) \quad \text{and} \quad \mathbf{x} \wedge \mathbf{y} = (x_1 \wedge y_1, \ldots, x_n \wedge y_n).$$

Similarly, for $\mathbf{x}^1, \ldots, \mathbf{x}^n \in L^n$, $m \in \mathbb{N}$, $m > 1$, we write

$$\bigvee_{i=1}^{m} \mathbf{x}^i = (x_1^1 \vee \ldots \vee x_1^m, \ldots, x_n^1 \vee \ldots \vee x_n^m),$$
$$\bigwedge_{i=1}^{m} \mathbf{x}^i = \left(x_1^1 \wedge \ldots \wedge x_1^m, \ldots, x_n^1 \wedge \ldots \wedge x_n^m\right).$$

The class of *lattice polynomials functions* (or simply *lattice polynomials*) from L^n to L, $n \geq 1$, can be defined recursively by finitely many applications of the following rules:

1. For any $k \in [n] = \{1, \ldots, n\}$, the projection $(x_1, \ldots, x_n) \mapsto x_k$ is a lattice polynomial from L^n to L.
2. If $p, q : L^n \to L$ are lattice polynomials $p \vee q$ and $p \wedge q$ are lattice polynomials from L^n to L.

It is well-known [14] that a function $f : L^n \to L$ is a lattice polynomial if and only if it can be represented in *disjunctive normal form* (*DNF*), i.e., there is $\alpha : 2^{[n]} \to L$ such that

$$f(\mathbf{x}) = \bigvee_{I \subseteq [n]} \left(\alpha(I) \wedge \bigwedge_{i \in I} x_i\right). \tag{1}$$

Note that for a lattice polynomial $f : L^n \to L$ there may be several set functions α that provide different *DNF* representations of f. For each $I \subseteq [n]$, let $\mathbf{e}_I$ be the element of L^n whose ith component is 1, if $i \in I$, and 0, otherwise. Let $\alpha_f \colon 2^{[n]} \to L$ be the function given by $\alpha_f(I) = f(\mathbf{e}_I)$ and consider the function $\alpha_f^* \colon 2^{[n]} \to L$ defined by

$$\alpha_f^*(I) = \begin{cases} \alpha_f(I), & \text{if } \bigvee_{J \subsetneq I} \alpha_f(J) < \alpha_f(I), \\ 0, & \text{otherwise.} \end{cases}$$

As shown in [6], a function $\alpha : 2^{[n]} \to L$ can be used in the *DNF* representation of f if and only if $\alpha \in [\alpha_f^*, \alpha_f]$. It is thus easy to see that α_f^* and α_f are unique and that they give raise to the minimal and maximal, resp., *DNF* representations of f.

Remark 1 *The DNF of a lattice polynomial f is the supremum of weighted terms :*

$$f(\boldsymbol{x}) = \bigvee_{I \subseteq [n]} g_I(\boldsymbol{x}) \quad \textit{where} \quad g_I(\boldsymbol{x}) = \alpha_f(I) \wedge \bigwedge_{i \in I} x_i.$$

Note that if we have

$$g_I\left(\boldsymbol{x}^1\right) \geq g_I\left(\boldsymbol{x}^2\right) \vee \ldots \vee g_I\left(\boldsymbol{x}^{k+1}\right),$$

for all $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{k+1} \in L^n$ and all $I \subseteq [n]$, then

$$f\left(\boldsymbol{x}^1\right) \geq f\left(\boldsymbol{x}^2\right) \vee \ldots \vee f\left(\boldsymbol{x}^{k+1}\right).$$

A function $f : L^n \to L$ is said to be *idempotent* if, for every $a \in L$,

$$f(a, \ldots, a) = a.$$

As observed in [22], (discrete) Sugeno integrals can be thought of as lattice polynomials that are idempotent. or, equivalently, as lattice polynomials $f : L^n \to L$ such that α_f is a *capacity* on $[n]$, i.e., $\alpha_f(J) \leq \alpha_f(J')$ whenever $J \subseteq J'$, and $\alpha_f(\mathbf{c}) = c$ for $c \in \{0, 1\} \subseteq L$. For the sake of clarity, we shall denote capacities on $[n]$ by μ and the corresponding Sugeno integrals by $\mathcal{S}_\mu : L^n \to L$, i.e.,

$$\mathcal{S}_\mu(\mathbf{x}) = \bigvee_{I \subseteq [n]} \left(\mu(I) \wedge \bigwedge_{i \in I} x_i \right).$$

Remark 2 *For a Sugeno integral S_μ, the function $\alpha^*_{S_\mu}$ is called* the ordinal Möbius transform *of μ [17].*

3 k-maxitive Sugeno integrals and lattice polynomials with degree at most k

We now focus on k-maxitive capacities, also called *k-order possibility measures* in [4, 23]. A capcity is said to be *k-maxitive* if we have

$$\mu(I) = \bigvee_{J \subseteq I, |J| \leq k} \mu(J), \quad \text{for all } I \subseteq [n] \text{ with } |I| > k. \tag{2}$$

Note that k-maxitive capacities are thus completely determined by their values on sets with at most k elements.

A Sugeno integral that is defined with respect to a k-maxitive capacity is also said to be *k-maxitive*. It is easy to see from (2) that for a k-maxitive Sugeno integral $\mathcal{S}_\mu$, $\alpha^*_{\mathcal{S}_\mu} = 0$ for all $|I| > k$, and thus $\mathcal{S}_\mu$ can be expressed as a supremum of terms with at most k variables. As we will now see, in the case of lattice polynomials, k-maxitivity translates into the the notion of "degree at most k". The *degree* of a lattice polynomial f, denoted $deg(f)$, is defined by

$$deg(f) = max\{|I| : \alpha^*_f(I) \neq 0\}.$$

In other words, the degree of f is the size of the longest term in minimal DNF representation of f. We denote by $\mathscr{C}^k$ the class of lattice polynomials with degree at most k, that is,

$$\mathscr{C}^k = \{f \text{ is a lattice polynomial } : deg(f) \leq k\}. \tag{3}$$

It is now easy to see that k-maxitive Sugeno integrals coincide exactly with those idempotent lattice polynomials in $\mathscr{C}^k$.

Remark 3 *The notion of k-maxitivity is somewhat the ordinal variant of "k-additivity" for real-valued capacities. Let $\mathbb{I}$ be a real interval and $\mu : 2^{[n]} \to \mathbb{I}$ a capacity. The*

Möbius transform m_μ of μ is defined by

$$m_\mu(I) = \sum_{J \subseteq I} (-1)^{|I|-|J|} \mu(J) \quad \textit{[21].}$$

It is well known [17] that the Choquet integral w.r.t. μ can then be defined in terms of the Möbius transform m_μ by

$$\mathcal{C}_\mu(\boldsymbol{x}) = \sum_{I \subseteq [n]} m_\mu(I) \cdot \bigwedge_{i \in I} x_i. \tag{4}$$

*A capacity μ is said to be k-*additive *if, for every $|I| > k$, $m_\mu(I) = 0$. Similarly, a Choquet integral is said to be k-*aditive *if it is defined with respect to a k-additive capacity.*

The appealing feature of k-additivity becomes apparent from (4) *since it implies that k-additive Choquet integrals can be expressed as sums of weighted terms of size at most k. In the context of multicriteria decision making, k-additive Choquet integrals correspond then to aggregation models where interaction indices among more than k criteria are equal to 0. This fact confirms the analogy between k-maxitive Sugeno integrals and k-additive Choquet integrals: when $|i| > k$, we have that $\alpha^*_{\mathcal{S}_\mu}(I) = 0$ in the former case, whereas $m_\mu(I) = 0$ in the latter case.*

Now it is natural to ask for a criterion to check the whether data can be modeled by a k-maxitive Sugeno integral. For this purpose, we provide an axiomatization of the class $\mathscr{C}^k$ based on existing data.

Theorem 1 *Let $f : L^n \to L$ be a lattice polynomial. The three following statements are equivalent :*

(i) $f \in \mathscr{C}^k$,

(ii) for all $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{k+1} \in L^n$ *:*

$$f\left(\bigvee_{i=1}^{k+1} \bigwedge_{j \neq i} \boldsymbol{x}^j\right) \leq f\left(\boldsymbol{x}^1\right) \vee \ldots \vee f\left(\boldsymbol{x}^{k+1}\right),$$

(iii) for all $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{k+1} \in \{0,1\}^n$ *:*

$$f\left(\bigvee_{i=1}^{k+1} \bigwedge_{j \neq i} \boldsymbol{x}^j\right) \leq f\left(\boldsymbol{x}^1\right) \vee \ldots \vee f\left(\boldsymbol{x}^{k+1}\right).$$

Proof 1 *This proof was corrected after reviewing, and is avalaible in the published version of the paper.*

Note that in the particular case when $L = \{0, 1\}$, we get the characterization of the class of nondecreasing Boolean functions with degree at most k in [8].

Remark 4 *Theorem 1 cannot be strengthened by replacing the inequality by an equality in (ii) and in (iii). Indeed, consider the lattice polynomial*

$$f(x_1, x_2, x_3) = x_1 \vee (x_2 \wedge x_3).$$

We see that f is a lattice polynomial with degree 2. Consider $\boldsymbol{x}^1 = (0,0,0)$, $\boldsymbol{x}^2 = (0,0,0)$ and $\boldsymbol{x}^3 = (1,1,1)$. We have

$$f\left(\bigvee_{i=1}^{3} \bigwedge_{j \neq i} \boldsymbol{x}^j\right) < f\left(\boldsymbol{x}^1\right) \vee f\left(\boldsymbol{x}^2\right) \vee f\left(\boldsymbol{x}^3\right).$$

4 Learning Sugeno integrals from empirical data

In this section we consider the problem of learning a Sugeno integral from a dataset $\mathscr{D}$ containing the feature tuple together with the global utility of each alternative, that is

$$\mathscr{D} = \left\{ \left(\mathbf{x}^1, y^1\right), \ldots, \left(\mathbf{x}^m, y^m\right) \right\}.$$

where $\mathbf{x}^j = (x_1^j, \ldots, x_n^j)$ and y^j are respectively the criteria values and the utility of the j^{th} alternative.

We make the assumption that a Sugeno integral is the aggregation model underlying the preferences expressed $\mathscr{D}$. We want to learn a Sugeno integral $\mathcal{S}_\mu$ able to predict, for $j \in \{1, \ldots, m\}$, the value of y^j from the value of $\mathbf{x}^j$. As typically the data that $\mathscr{D}$ contains are noisy, it is impossible to find any Sugeno integral such that $\mathcal{S}_\mu(\mathbf{x}^j) = y^j$ for all $j \in \{1, \ldots, m\}$ ($\mathscr{D}$ is said to be inconsistent [25]). Learning the aggregation model consists in searching the Sugeno integral which can predict global utility values with as few errors as possible.

Remark 5 *Note that we do not assume any distance between the elements of L, but that utility values define a partial ordering of the alternatives. This is why a suitable error measure in this context is the pairwise error, which corresponds to the ratio of pairs of elements that are wrongly ordered by the aggregation model. Thus, what we consider to be the best Sugeno integral is the one that gives the most faithful ordering of the alternatives, that can differ from the one giving the closest prediction for the utility value, although accuracy in those two tasks are strongly correlated.*

Because of the intrinsic complexity of the Sugeno integral, learning an integral with the least prediction error is a difficult optimization problem. The ordinal nature of the treated values forbids using classical methods such as gradient decent (which can be efficiently applied for learning Choquet integrals [15]), and as a capacity is defined by its value for each subset of $[n]$, 2^n variables have to be considered ; thus, for a high number of attributes, running the optimization process can be costful.

A method that can be applied to this problem is a meta-heuristics such as simulated annealing (as proposed in [24]). This algorithm considers the space of every solution, and associates each solution with a cost (a value to minimize). Then it searches for the better solution by iteratively modifying the current solution. Modifying a solution corresponds to traveling to a close element in the solution space. When the modification of the current solution leads to an increase of the cost, there is a probability to refuse the change ; the higher is the increase, the higher is the probability. The probability to refuse a disadvantageous change also depends on the *temperature*, a value decreasing through the iterations ; at the beginning of the process, the probability of accepting a disadvantageous change is higher and decreases through the iterations, until it becomes close to 0. The aim of this variation is to avoid being "trapped" in local minima during the first iterations. For a more detailed description of simulated annealing, see for example [13].

We apply a simulated annealing to the problem of learning a Sugeno integral by associating each possible value of μ to a solution, whose cost is equal to the prediction error of $\mathcal{S}_\mu$ on $\mathscr{D}$. For creating a new solution μ_{new} from the current one, a subset of $[n]$ and a value from L (respectively I and y) are randomly picked, both with respect to an uniform probability distribution. The generated solution is then equal to the old one, yet with $\mu_{new}(I) = y$ (the values of μ_{new} for other sets are subsequently modified so that the monotonicity property of the capacity is respected). Note that this general method can be applied for learning k-maxitive Sugeno integrals simply by restricting the subsets of $[n]$ that can be picked to those of at most k criteria.

Evaluating the new solution produced at each iteration is in general the most time-consuming part of the simmulated annealing process. The complexity of computing the pairwise error is quadratic with respect to the number of elements in the learning set ; on the contrary the compexity of computing a mean absolute error (MAE) is linear with respect to $|\mathscr{D}|$. The MAE is fundamentally a measure that makes sense in a numerical context, since it computes the average absolute difference between predicted and actual values. The absolute difference can be generalized as a distance ; in our setting, the value of the MAE is then

$$\frac{1}{|\mathscr{D}|} \sum_{(\mathbf{x},y)\in\mathscr{D}} distance(\mathcal{S}_\mu(\mathbf{x}), y).$$

Nonetheless, this expression requires a notion of distance between two elements of L. For this we state that the distance between two neighbours elements of L is 1, and we define the distance between any two elements as the the length of the shortest path between them. By Birkhoff's Representation Theorem (see [1]), we know that any finite distributive lattice can be embeded in a Boolean (powerset) lattice. Computing the length of the shortest path between two elements of L is then equivalent to computing the cardinality of the symmetric difference of two sets. This operation can be done in a linear time (with respect to the number of criteria, which is usually very low compared to the cardinality of $\mathscr{D}$). As this error measure is less time-consuming and is in practice strongly correlated to the pairwise error, we used it for defining the cost of each solution in the simulated annealing process.

An interesting point of k-maxitivity is precisely to reduce the inner complexity of capacities. A k-maxitive capacity is indeed defined by its value on sets of size lower than or equal to k, that is to say $\sum_i^k \binom{i}{n}$. Therefore, restricting possible solutions to k-maxitive ones reduces the number of variables of the optimization problem, although

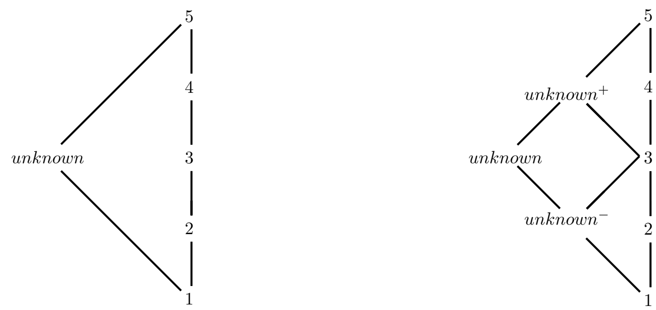

Figure 2: The evaluation space built from the data, in a non-distributive and distributive version (resp. left and right).

it can affect the precision of the resulting aggregation model. This will be discussed in the next section.

5 An case study

The aim of the following application is to illustrate at which extent k-maxitive Sugeno integrals could be used as preference aggregation models and which precision can be obtained for each k-values.

Our data set are users' hotel ratings from Trip Advisor [1]. Each rating is made by one user for one hotel, according to 7 criteria and associated to an overall evaluation. The rating scale is $\{1, 2, 3, 4, 5\}$ and, as some values can be missing, the evaluation space that we use contains an $unknown$ value. Here we suppose that $1 < unknown < 5$ without making any further assumption. We obtain a partial order that can be embeded in a distributive lattice, as shown in Figure 5.

Note that adding the values $unknown^-$ and $unknown^+$ to the original set of values preserves the order relations (and incomparabilities) between the elements of $\{1, 2, 3, 4, 5, unknown\}$. Nonetheless, it does not preserve the infimum and supremum operations, since we have, for example, $2 \wedge unknown = unknown^-$. The distributive lattice thus constructed is then used as the evaluation space.

Remark 6 *When L is a non-distributive lattice, we could still define aggregation functions on L^n that consist of expressions built with $\wedge$, $\vee$ and constants. However, without the distributivity assomption, it may be impossible to represent such functions by a DNF.*

The results are obtained from 30 random samples from the original data set, each of which were divided in two parts: a learning set containing 250 elements, and a testing set containing 1000 elements. For each sample, a k-maxitive Sugeno integral was trained on the learning set, for $k = 1, 2, 3, 4, 5, 6$. The resulting integrals were evaluated with respect to the testing set. Figure 5 shows the average pairwise error among samples, for each value of k.

[1]Tripadvisor Dataset : `http://sifaka.cs.uiuc.edu/~wang296/Data/index.html`.

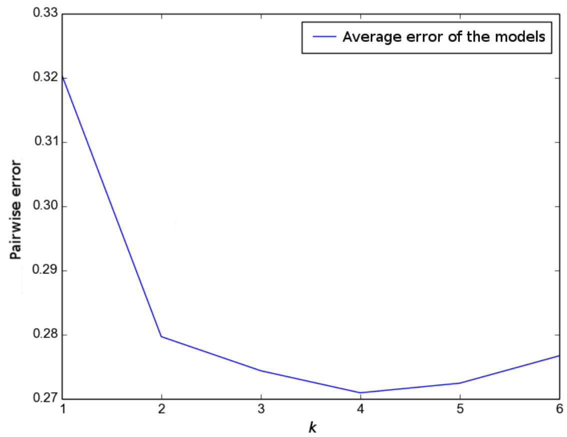

Figure 3: Average pairwise error of a k-maxitive Sugeno integral on the testing set.

6 Conclusion and further work

The results presented in Section 5 show that most of the accuracy is gained while increasing k from 1 (a 1-maxitive Sugeno integral that corresponds in fact to a weighted supremum operation) to 2, although it is not possible to decide whether this phenomenon is particular to our data or could be generalized. This observation could be compared to those made on k-additive Choquet integrals, that often show good performances when k is set to 2 (see for example [28]). It also seems that the optimal value of k (from the accuracy point of view) is 4, certainly because higher values tend to cause over-fitting of the model. Even though more empirical study should be made in that direction for a better understanding of the behavior of k-maxitive models, our case study shows that k-maxitivity could be an interesting property for limitating the inner complexity of Sugeno integrals, but also for preventing over-fitting.

A few problems are still to be investigated. From the application point of view, it would be interesting to predict the best value of k for a given dataset (for this case study we trained Sugeno integrals for each value of k, which is of course not a suitable solution for systems required to be efficient).

The principle of k-maxitivity is to force the values of a capacity μ to depend solely on the values of μ on subsets of size at most k. Another way of expressing this constraint is to state that the sets of cardinality lower than or equal to k are the only possible focal sets (i.e., sets I for which $\alpha^*_{\mathcal{S}_\mu}(I) > 0$ [11]) of the capacity. One could imagine setting other restrictions on the focal sets, so that the the inner complexity is kept as low as with k-maxitivity, while allowing more flexibility in the model and greater precision for practical applications such as preference aggregation.

Furthermore, the results that we have obtained in the current paper should be broadened to the framework of pseudo-lattice polynomial functions [7], that is to say, functions that can be written as

$$F(x_1, \dots, x_n) = \mathcal{S}_\mu(\varphi_1(x_1), \dots, \varphi_n(x_n))$$

with $\mathcal{S}_\mu : L^n \to L$ being a Sugeno integral, and $\varphi_i : X_i \to L$. These functions of course allow a greater expressiveness but learning them from data is less straitforward, since both the φ_i's and the Sugeno integral have to be learned, while being interdependent.

References

[1] G. Birkhoff. On the combination of subalgebras. *Mathematical Proceedings of the Cambridge Philosophical Society*, 29(4):441–464, 1933.

[2] G. Birkhoff. *Lattice Theory*, volume 25. American Mathematical Society, New York, USA, 1940.

[3] Q. Brabant and M. Couceiro. Axiomatisation des integrales de Sugeno k-maxitives. In *Proceedings of the 24th Rencontre francophones sur la Logique Floue et ses Applications*, pages 179–186, Poitiers, France, 2015.

[4] 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*, 6(6):533–550, 1998.

[5] M. Couceiro, D. Dubois, H. Prade, and T. Waldhauser. Decision-making with Sugeno integrals: bridging the gap between multicriteria evaluation and decision under uncertainty. *Order*, 33(3):517–535, 2016.

[6] M. Couceiro and J.-L. Marichal. Characterizations of discrete Sugeno integrals as polynomial functions over distributive lattices. *Fuzzy Sets and Systems*, 161(5):694–707, 2010.

[7] M. Couceiro and T. Waldhauser. Pseudo-polynomial functions over finite distributive lattices. *Fuzzy Sets and Systems*, 239:21–34, 2014.

[8] Y. Crama and P. L. Hammer. *Boolean Functions: Theory, Algorithms, and Applications*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, Cambridge, UK, 2011.

[9] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, UK, 2002.

[10] D. Dubois, J.-L. Marichal, H. Prade, M. Roubens, and R. Sabbadin. The use of the discrete Sugeno integral in decision-making: a survey. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(5):539–561, 2001.

[11] D. Dubois, H. Prade, and A. Rico. Representing qualitative capacities as families of possibility measures. *International Journal of Approximate Reasoning*, 58:3–24, 2015.

[12] D. Dubois, H. Prade, and R. Sabbadin. Qualitative decision theory with Sugeno integrals. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 121–128, San Francisco, USA, 1998. Morgan Kaufmann Publishers Inc.

[13] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, USA, 2012.

[14] R. L. Goodstein. The solution of equations in a lattice. *Proceedings of the Royal Society of Edinburgh. Section A*, 67:231–242, 1965/1967.

[15] M. Grabisch. A new algorithm for identifying fuzzy measures and its application to pattern recognition. In *Proceedings of the International Joint Conference of the 4th IEEE International Conference on Fuzzy Systems and the 2nd International Fuzzy Engineering Symposium*, volume 1, pages 145–150, Yokohama, Japan, 1995.

[16] M. Grabisch and C. Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *Annals of Operations Research*, 175(1):247–286, 2010.

[17] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation Functions*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, Cambridge, UK, 2009.

[18] G. Grätzer and B. A. Davey. *General Lattice Theory*. Springer Science & Business Media, Berlin, Germany, 2003.

[19] G. Grätzer. *Lattice Theory : Foundation*. Springer Basel, 2011.

[20] E. Hüllermeier and A. F. Tehrani. Efficient learning of classifiers based on the 2-additive Choquet integral. In *Computational Intelligence in Intelligent Data Analysis*, pages 17–29. Springer, Berlin, Germany, 2013.

[21] J.-L. Marichal. On Choquet and Sugeno integrals as aggregation functions. *Fuzzy Measures and Integrals-Theory and Applications*, pages 247–272, 2000.

[22] J.-L. Marichal. Weighted lattice polynomials. *Discrete Mathematics*, 309(4):814–820, 2009.

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

[24] 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.

[25] H. Prade, A. Rico, M. Serrurier, and E. Raufaste. Elicitating Sugeno integrals: Methodology and a case study. In *Proceedings of the 10th European Conference Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 712–723, Verona, Italy, 2009. Springer.

[26] S. Rudeanu. *Lattice Functions and Equations*. Discrete Mathematics and Theoretical Computer Science. Springer Science & Business Media, London, UK, 2001.

[27] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, 1974.

[28] A. F. Tehrani, W. Cheng, and E. Hüllermeier. Preference learning using the Choquet integral: The case of multipartite ranking. *IEEE Transactions on Fuzzy Systems*, 20(6):1102–1113, 2012.